Casualty Actuarial Society
E-Forum, Spring 2017

The CAS *E-Forum*, Spring 2017

The Spring 2017 edition of the CAS *E-Forum* is a cooperative effort between the CAS *E-Forum* Committee and various CAS committees, task forces, working parties and special interest sections.

This *E-Forum* contains two research projects funded by the CAS.

"An Adaptation of the Classical CAPM to Insurance: The Weighted Insurance Pricing Model," is commissioned by the CAS Theory of Risk Committee. It is written by professors from York University and the University Western Ontario in Canada.

"Compendium of Credit Risk Resources," is commissioned by the CAS Credit Risk Special Interest Section (CRiSIS) with the support of the former CAS Committee on Valuation, Finance and Investments. The compendium was undertaken by members of the Quantact Actuarial and Financial Mathematics Laboratory, who are professors affiliated with the University of Montréal, University of Québec at Montréal, Concordia University and Laval University.

This *E-Forum* also contains three ratemaking call papers, which were created in response to a call for papers issued by the CAS Ratemaking Committee, and one independent research paper.

CAS Credit Risk Special Interest Section (CRiSIS)

Michael C. Schmitz, *President*
David L. Ruhm, *Vice President*
John E. Gaines, *Secretary*

Financial Reporting and Analysis Committee

Marc F. Oberholtzer, *Chairperson*

Bradley Andrekus
Michael Belfatti
Frank Chang
Andrew Cheng
A. Cummings
Stephen DiCenso
Jacob Fetzer
Dane Grand-Maison
Denis Guenthner
Aaron Haning
Gareth Kennedy
Alexander Kozmin
Clifford Lau
Rasa McKean
David Rosenzweig
Daniel Schlemmer
Michael Schmitz
Lisa Slotznick
David Shleifer
Lijia Tian
Glenn Tobleman
Ned Tyrrell
Kun Zhang
David Core, *Staff Liaison*
Cheri Widowski, *Staff Liaison*

Ratemaking Committee

Morgan Haire Bugbee, *Chairperson*
Sandra J. Callanan, *Vice Chairperson*

LeRoy A. Boison
Caryn C. Carmean
William M. Carpenter
James Chang
Donald L. Closter
Christopher L. Cooksey
Sean R. Devlin
Greg Frankowiak
Serhat Guven
Ray Yau Kui Ho
Duk Inn Kim
Dennis L. Lange
Ronald S. Lettofsky
Shan Lin
Gregory F. McNulty
Jane C. Taylor
Bruno Tremblay
Karen Sonnet, *Staff Liaison*

CAS*E-Forum*, Spring 2017

Table of Contents

For information on submitting a paper to the *E-Forum*, visit http://www.casact.org/pubs/forum/.

An Adaptation of the Classical CAPM to Insurance: The Weighted Insurance Pricing Model

Edward Furman

Department of Mathematics and Statistics, York University, Toronto, Ontario M3J 1P3, Canada

Ričardas Zitikis

Department of Statistical and Actuarial Sciences, University of Western Ontario, London, Ontario N6A 5B7, Canada

Abstract. We present and discuss an insurance version of the classical Capital Asset Pricing Model that offers economic pricing and risk capital allocation rules for a large class of risks, including those that are non-symmetric and heavy tailed. A number of illustrative examples are given, and convenient computational formulas suggested.

Key words and phrases: capital asset pricing model, weighted insurance pricing model, Gini-type insurance pricing model, beta, Gini correlation.

1. Introduction

The Capital Asset Pricing Model (CAPM) has profoundly influenced Finance and Insurance, with numerous articles and books written on the topic by academics and practitioners (e.g., Levy, 2011; and references therein). In this paper we aim at modifying the classical CAPM to accommodate some of the 'peculiarities' of insurance risks, in particular their positivity, skewness, and heavy tails.

We start with the obvious. Namely, the classical CAPM links the expected riskiness of portfolio constituents with the overall portfolio riskiness. Specifically, expressed in its classical form, the CAPM equation is

$$\mathbf{E}[R_i] = r_f + \beta_i\big(\mathbf{E}[R_m] - r_f\big), \tag{1.1}$$

where $\mathbf{E}[R_i]$ is the expected return on asset i, $\mathbf{E}[R_m]$ is the expected market rate of return, r_f is the risk-free rate of return, and β_i is the proportionality coefficient, widely known as 'beta' and given by the equation

$$\beta_i = \frac{\mathbf{Cov}[R_i, R_m]}{\mathbf{Cov}[R_m, R_m]}. \tag{1.2}$$

At this initial point of our discussion, it is instructive to recall the classical linear regression equation, which, under the assumption of bivariate normality on the pair (R_i, R_m), says that the conditional expectation $\mathbf{E}[R_i \mid R_m = s]$, as a function of s, is the straight line $a + bs$ with the slope $b = \beta_i$ and the intercept $a = \mathbf{E}[R_i] - \beta_i \mathbf{E}[R_m]$. This representation and its similarity to CAPM equation (1.1) explain the central role of the multivariate normal distribution in the CAPM literature.

Of course, departures from the normality assumption (e.g., Owen and Rabinovitch, 1983, for elliptical distributions) have been established and extensively discussed (Levy, 2011; and references therein). Indeed, risks generally deviate from symmetry and are often heavy tailed. In addition, insurance risks are as a rule positively valued (e.g., Klugman et al., 2008). Due to these and other reasons, when applying the CAPM equation to price insurance products or allocate capital to individual risks, we inevitably find ourselves in a position of doubt. In this paper, therefore, we propose to tweak the classical CAPM so that it would mitigate, if not resolve, the aforementioned issues. We illustrate the underlying idea in the next section using the so-called modified-variance risk measure and the corresponding risk capital allocation.

Throughout the rest of the paper, we use $X_1, X_2, \ldots, X_d$ to denote real-valued (i.e., not necessarily positive) risk random variables whose aggregate riskiness is expressed by a random variable S (e.g., $S = X_1 + X_2 + \cdots + X_d$).

2. An illuminating example and our general aim

We start with an example illustrating that departure from normality is not difficult to achieve. To make our initial arguments as simple as possible, we work with the modified-variance risk measure (Heilmann, 1989)

$$\begin{aligned} \mathrm{mv}[S] &= \mathbf{E}[S] + \frac{1}{\mathbf{E}[S]}\mathbf{Var}[S] \\ &= \frac{\mathbf{E}[S^2]}{\mathbf{E}[S]} \end{aligned} \tag{2.1}$$

and, for $i \in \{1, \ldots, d\}$, the corresponding risk capital allocation rule

$$\mathrm{MV}[X_i \mid S] = \frac{\mathbf{E}[X_i S]}{\mathbf{E}[S]}. \tag{2.2}$$

According to our view of the CAPM, we want to express $\mathrm{MV}[X_i \mid S]$, which measures the riskiness of X_i within the collection of risks, in terms of $\mathrm{mv}[S]$, which measures the aggregate riskiness. We achieve this goal with the help of simple algebra and, most importantly, without imposing any distributional constraints on the pair (X_i, S). Namely,

we have the equations

$$\begin{aligned}\mathrm{MV}[X_i \mid S] - \mathbf{E}[X_i] &= \frac{\mathbf{E}[X_i S]}{\mathbf{E}[S]} - \mathbf{E}[X_i] \\ &= \frac{\mathbf{Cov}[X_i, S]}{\mathbf{E}[S]} \\ &= \frac{\mathbf{Cov}[X_i, S]}{\mathbf{Cov}[S, S]} \frac{\mathbf{Cov}[S, S]}{\mathbf{E}[S]} \\ &= \frac{\mathbf{Cov}[X_i, S]}{\mathbf{Cov}[S, S]} \frac{\mathbf{E}[S^2] - (\mathbf{E}[S])^2}{\mathbf{E}[S]} \\ &= \beta_i\big(\mathrm{mv}[S] - \mathbf{E}[S]\big), \end{aligned} \tag{2.3}$$

where the 'beta'

$$\beta_i = \frac{\mathbf{Cov}[X_i, S]}{\mathbf{Cov}[S, S]} \tag{2.4}$$

is of the same form as that given by equation (1.2).

Hence, in summary, in the modified-variance case, the insurance analogue of CAPM equation (1.1) is the equation

$$\mathrm{MV}[X_i \mid S] = \mathbf{E}[X_i] + \beta_i\big(\mathrm{mv}[S] - \mathbf{E}[S]\big), \tag{2.5}$$

which holds for all pairs (X_i, S) for which $\mathrm{MV}[X_i \mid S]$ and $\mathrm{mv}[S]$ are well-defined and finite: no specific distribution on the risks has been imposed.

Despite the latter optimistic message, we still rely on the existence of finite second moments of the underlying random risks, but this is only due to our choice of the modified-variance risk measure and the capital allocation rule. To accommodate heavier-tailed risks, we therefore wish to depart from the above risk measure and the capital allocation rule, and for this we put forward a research program whose main idea hinges on the following modification of CAPM equation (1.1):

(1) replace the two risk-free rates of return r_f by the corresponding averages $\mathbf{E}[X_i]$ and $\mathbf{E}[S]$, frequently called net premiums in the actuarial literature;
(2) replace the expected market rate of return $\mathbf{E}[R_m]$ by a risk measure $\pi[S]$ of the aggregate risk S;
(3) replace the expected return $\mathbf{E}[R_i]$ on the asset i by a risk capital allocation rule $\Pi[X_i \mid S]$ due to the risk X_i;
(4) find, if possible, an appropriate proportionally coefficient β_i – which we keep calling 'beta' to maintain consistency with the already accepted terminology in the CAPM literature – that does not depend on any utility, weight, distortion, etc. 'subjective' function.

Hence, in various scenarios of practical interest, in what follows we aim at deriving the equation

$$\Pi[X_i \mid S] = \mathbf{E}[X_i] + \beta_i\big(\pi[S] - \mathbf{E}[S]\big), \tag{2.6}$$

which we generally call the insurance pricing model (IPM) equation. An important clarification is needed at this point in order to avoid a potential misunderstanding.

Namely, from the mathematical point of view, equation (2.6) always holds with $\beta_i = \beta_i(\Pi, \pi)$ defined by

$$\beta_i(\Pi, \pi) = \frac{\Pi[X_i \mid S] - \mathbf{E}[X_i]}{\pi[S] - \mathbf{E}[S]}, \tag{2.7}$$

whenever of course $\pi[S]$ is positively loaded, that is $\pi[S] > \mathbf{E}[S]$. The ratio of loadings $\beta_i(\Pi, \pi)$ may, in general, depend on 'subjective' functions (e.g., utility, weight, distortion, etc.) that define the risk capital allocation rule $\Pi[X_i \mid S]$ and the risk measure $\pi[S]$. But we say that equation (2.6) is the IPM equation only when $\beta_i(\Pi, \pi)$ does not depend on these functions. Hence, our proposed IPM hinges on the fact that under certain but quite general conditions, ratio (2.7) is independent of any subjective function, and it is only in this case that we call ratio (2.7) 'beta.'

In what follows, we discuss several versions of the IPM equation: the weighted insurance pricing model (WIPM) equation in Section 3, and the Gini-type weighted insurance pricing model (G-WIPM) equation in Section 4.

3. Weighted insurance pricing model

The weighted risk measure (Furman and Zitikis, 2008a, 2009), which is very general and allows us to accommodate virtually every risk irrespective of its tail-heaviness as long as we appropriately choose a weight function $w : (-\infty, \infty) \to [0, \infty)$, is defined by

$$\pi_w[S] = \frac{\mathbf{E}[Sw(S)]}{\mathbf{E}[w(S)]}. \tag{3.1}$$

The weight function w is usually assumed, or chosen, to be non-decreasing, which ensures, for example, non-negative loading of the risk measure. The corresponding weighted risk capital allocation rule is (Furman and Zitikis, 2008b)

$$\Pi_w[X_i \mid S] = \frac{\mathbf{E}[X_i w(S)]}{\mathbf{E}[w(S)]}. \tag{3.2}$$

For example, by choosing the weight functions $w : [0, \infty) \to [0, \infty)$ given by

$$w(s) = s^\lambda,$$

$$w(s) = e^{\lambda s},$$

$$w(s) = 1 - e^{-\lambda s},$$

$$w(s) = \mathbf{1}\{s > \lambda\},$$

where $\lambda > 0$ is a parameter, we reduce $\pi_w[S]$ to the size-biased, Esscher's, Kamps's, and excess-of-loss risk measures, and we in turn reduce $\Pi_w[X_i \mid S]$ to the corresponding risk capital allocation rules. A few other examples will follow later in this paper, but next we show how the IPM equation (which we call WIPM) arises in the case of the weighted risk measure and the corresponding capital allocation rule.

Using simple algebra and following the same route as in the modified-variance case, we obtain the equations

$$\begin{aligned}\Pi_w[X_i \mid S] - \mathbf{E}[X_i] &= \frac{\mathbf{E}[X_i w(S)]}{\mathbf{E}[w(S)]} - \mathbf{E}[X_i] \\ &= \frac{\mathbf{Cov}[X_i, w(S)]}{\mathbf{E}[w(S)]} \\ &= \frac{\mathbf{Cov}[X_i, w(S)]}{\mathbf{Cov}[S, w(S)]} \frac{\mathbf{Cov}[S, w(S)]}{\mathbf{E}[w(S)]} \\ &= \frac{\mathbf{Cov}[X_i, w(S)]}{\mathbf{Cov}[S, w(S)]} \frac{\mathbf{E}[Sw(S)] - \mathbf{E}[S]\mathbf{E}[w(S)]}{\mathbf{E}[w(S)]} \\ &= \beta_{i,w}\big(\pi_w[S] - \mathbf{E}[S]\big), \end{aligned} \tag{3.3}$$

where the ratio of loadings $\beta_{i,w}$ (we refrain from calling it 'beta' because it may, in general, depend on the 'subjective' weight function w) is given by the equation

$$\beta_{i,w} = \frac{\mathbf{Cov}[X_i, w(S)]}{\mathbf{Cov}[S, w(S)]}. \tag{3.4}$$

When, however, $\beta_{i,w}$ does not depend on w, that is, $\beta_{i,w} = \beta_i$ for some β_i, the above considerations give rise to the equation (cf. Furman and Zitikis, 2010)

$$\Pi_w[X_i \mid S] = \mathbf{E}[X_i] + \beta_i\big(\pi_w[S] - \mathbf{E}[S]\big), \tag{3.5}$$

which we call the WIPM equation, and which is our proposed insurance analogue of CAPM equation (1.1). Note that when $w(s) = s$, equation (3.5) reduces to equation (2.5), but this fact does not imply that β_i in equation (3.5) is the same as in equation (2.4), as we shall see in a moment. We next show the validity of WIPM equation (3.5) in two special cases.

Case 1: linear regression. Assume that the regression function

$$r_i(s) = \mathbf{E}[X_i \mid S = s] \tag{3.6}$$

is linear, that is,

$$r_i(s) = a + bs \tag{3.7}$$

for some constants a and b, called the intercept and the slope, respectively. Then

$$\begin{aligned}\beta_{i,w} &= \frac{\mathbf{Cov}[X_i, w(S)]}{\mathbf{Cov}[S, w(S)]} \\ &= \frac{\mathbf{Cov}[r_i(S), w(S)]}{\mathbf{Cov}[S, w(S)]} \\ &= \frac{\mathbf{Cov}[a + bS, w(S)]}{\mathbf{Cov}[S, w(S)]} \\ &= b\frac{\mathbf{Cov}[S, w(S)]}{\mathbf{Cov}[S, w(S)]} = b. \end{aligned} \tag{3.8}$$

Hence, the ratio of loadings $\beta_{i,w}$ is equal to the slope b, which is of course free of the weight function w and can thus be called 'beta.' In turn, WIPM equation (3.5) becomes

$$\Pi_w[X_i \mid S] = \mathbf{E}[X_i] + b\big(\pi_w[S] - \mathbf{E}[S]\big). \tag{3.9}$$

The regression function is linear in a number of popular multivariate risk models. We refer to Furman and Zitikis (2010), Su (2016), Su and Furman (2017), Furman and Zitikis (2016a,b) for examples, details, and further references. In particular, in these works we find expressions of the slope b in terms of distribution parameters, which can in turn be estimated using various techniques already available in the literature, such as the maximum likelihood method, the method of (trimmed) moments, and so on (e.g., Brazauskas et al., 2009; Kleefeld and Brazauskas, 2012; and references therein). We may also seek non-parametric estimators of b, which can be found in standard books on regression.

Case 2: linear regression and non-negative risks. Having mentioned non-negative risks, which are abundant in insurance and are called losses (e.g., Klugman et al., 2008), we now look at the case of non-negative risks X_i. Let the aggregate risk be the sum $S = X_1 + X_2 + \cdots + X_d$. Due to the obvious equations $\sum_i r_i(0) = \mathbf{E}[S \mid S = 0] = 0$ and the non-negativity of all the summands $r_i(0)$, we have $r_i(0) = 0$. This fact and linearity assumption (3.7) imply that the intercept of the regression line vanishes, that is, $a = 0$, and we thus in turn obtain the equation

$$b = \frac{\mathbf{E}[X_i]}{\mathbf{E}[S]} \tag{3.10}$$

because $\mathbf{E}[X_i] = \mathbf{E}[r_i(S)] = b\mathbf{E}[S]$. Consequently, WIPM equation (3.5) becomes

$$\begin{aligned}\Pi_w[X_i \mid S] &= \mathbf{E}[X_i] + \frac{\mathbf{E}[X_i]}{\mathbf{E}[S]}\big(\pi_w[S] - \mathbf{E}[S]\big) \\ &= \frac{\mathbf{E}[X_i]}{\mathbf{E}[S]}\pi_w[S]. \end{aligned} \tag{3.11}$$

Note that the 'beta' b given by equation (3.10) requires the existence of only the first moments of the risks X_i and S. This is in sharp contrast with the covariance-based betas that we encountered earlier in this paper.

4. Gini-type weighted insurance pricing model

There are many risk measures and risk capital allocation rules that are not covered in our previous discussion of $\pi_w[S]$ and $\Pi_w[X_i \mid S]$. The reason is that in a number of cases the weight function w acts not on the aggregate risk severity S but on its rank $F(S)$, where F is the cumulative distribution function (cdf) of S. Hence, we next turn $\pi_w[S]$ and $\Pi_w[X_i \mid S]$ into what we call the Gini-type weighted risk measure

$$\pi_{w,\text{Gini}}[S] = \frac{\mathbf{E}[Sw(F(S))]}{\mathbf{E}[w(F(S))]} \tag{4.1}$$

and the corresponding Gini-type risk capital allocation rule

$$\Pi_{w,\text{Gini}}[X_i \mid S] = \frac{\mathbf{E}[X_i w(F(S))]}{\mathbf{E}[w(F(S))]}. \tag{4.2}$$

To illustrate them, we use the weight functions $w : [0,1] \to [0,\infty)$ given by

$$w(t) = p(1-t)^{p-1},$$
$$w(t) = g'(1-t),$$
$$w(t) = e^{pt},$$
$$w(t) = \mathbf{1}\{t > p\},$$

where $p > 0$ is a parameter, and $g : [0,1] \to [0,1]$ is a 'distortion' function (e.g., $g(t) = t^p$; Wang, 1995, 1996). The above examples of the weight function w give rise to, respectively, the proportional hazards, distortion, Aumann-Shapley, and conditional tail expectation risk measures, as well as to the corresponding risk capital allocation rules. In addition, the weight function

$$w(t) = w_0(t)\mathbf{1}\{t > p\}$$

with some 'underlying' weight function $w_0 : [0,1] \to [0,\infty)$ leads to what we call the Gini-type conditional-tail weighted risk measure and the corresponding risk capital allocation rule (Furman and Zitikis, 2016a,b). For more extensive mathematical details on this topic, we refer to Furman et al (2017).

To derive the corresponding pricing model, we follow equations (3.3) with $w(F(S))$ instead of $w(S)$ and have

$$\Pi_{w,\text{Gini}}[X_i \mid S] = \mathbf{E}[X_i] + \beta_{i,w,\text{Gini}}\big(\pi_{w,\text{Gini}}[S] - \mathbf{E}[S]\big), \tag{4.3}$$

where the ratio of loadings is

$$\beta_{i,w,\text{Gini}} = \frac{\mathbf{Cov}[X_i, w(F(S))]}{\mathbf{Cov}[S, w(F(S))]}. \tag{4.4}$$

When the ratio of loadings $\beta_{i,w,\text{Gini}}$ does not depend on w, in which case we denote it by $\beta_{i,\text{Gini}}$, we arrive at the equation

$$\Pi_{w,\text{Gini}}[X_i \mid S] = \mathbf{E}[X_i] + \beta_{i,\text{Gini}}\big(\pi_{w,\text{Gini}}[S] - \mathbf{E}[S]\big) \tag{4.5}$$

that we call the G-WIPM equation. A remark on $\beta_{i,w,\text{Gini}}$ follows next, after which we discuss two cases when $\beta_{i,w,\text{Gini}}$ does not depend on w.

Interestingly, the ratio of covariances on the right-hand side of equation (4.4) has already appeared in the literature. Namely, when $w(t) = t$, the ratio is known as the Gini correlation coefficient and is usually denoted by $\Gamma[X_i, S]$. Its origins can be traced back to the work of C. Gini a hundred years ago. The ratio of covariances under the weight function $w(t) = p(1-t)^{p-1}$ is usually denoted by $\Gamma_p[X_i, S]$ and called the extended Gini correlation coefficient. It appeared and was thoroughly investigated several decades ago in the works of S. Yitzhaki and E. Schechtman, and we refer to the recent monograph by Yitzhaki and Schechtman (2013) for details and references on the topic. In our CAS technical report (Furman and Zitikis, 2016a), we discuss the weighted Gini-type correlation coefficient in the same form as presented on the right-hand side of equation (4.4). We next discuss two cases when $\beta_{i,w,\text{Gini}}$ does not depend on w.

Case 1: linear regression. A sufficient condition for the G-WIPM equation to hold is the linearity of the regression function $r_i(s)$, that is, when equation (3.7) holds. Indeed, in this case we can follow equations (3.8) with $w(F(S))$ instead of $w(S)$ and obtain that $\beta_{i,w,\text{Gini}}$ is equal to the slope b of the regression line, which is of course independent of w and can thus be called 'beta.' In this case, analogously to WIPM equation (3.9), we have the following G-WIPM equation

$$\Pi_{w,\text{Gini}}[X_i \mid S] = \mathbf{E}[X_i] + b\big(\pi_{w,\text{Gini}}[S] - \mathbf{E}[S]\big). \tag{4.6}$$

Case 2: linear regression and non-negative risks. When in addition to linearity of the regression function we also deal with non-negative risks, the G-WIPM equation turns into

$$\begin{aligned}\Pi_{w,\text{Gini}}[X_i \mid S] &= \mathbf{E}[X_i] + \frac{\mathbf{E}[X_i]}{\mathbf{E}[S]}\big(\pi_{w,\text{Gini}}[S] - \mathbf{E}[S]\big) \\ &= \frac{\mathbf{E}[X_i]}{\mathbf{E}[S]}\pi_{w,\text{Gini}}[S],\end{aligned} \tag{4.7}$$

which is an analogue of WIPM equation (3.11).

5. On the independence of loading ratios of w

In our explorations above, we have encountered three ratios of loadings: the most general $\beta_i(\Pi, \pi)$ in equation (2.7), the weighted ratio $\beta_{i,w}$ in equation (3.4), and the Gini-type weighted ratio $\beta_{i,w,\text{Gini}}$ in equation (4.4). As in the classical CAPM, neither of these ratios we want to depend on any 'subjective' function, such as the weight function w.

A parametric route for checking whether this is true or not has transpired in our considerations above. Namely, given data, we can start with a goodness-of-fit technique and choose an appropriate parametric distribution for (X_i, S). Then, mathematically, we would derive an expression for the regression function $r_i(s)$ and see whether it is linear or not (e.g., Furman and Zitikis, 2016a,b; and references therein).

The parametric approach, however, is usually quite time and energy consuming, given the mathematical complexities that naturally arise when calculating conditional densities and, in turn, the regression function $r_i(s)$. Hence, one would – at least initially – prefer a simpler non-parametric method (some kind of a 'rule of thumb') to determine whether the loading ratio could, or could not, be free of any 'subjective' function. We offer several thoughts on this issue that we think might be helpful in practice.

Ratio $\beta_{i,w}$: a computational formula. Perhaps the simplest way that we can think of for checking if the ratio $\beta_{i,w}$ might be independent of w would be to construct a non-parametric estimate of the ratio and then plug into it several specific weight functions w to see whether there would be any significant change in the obtained estimates. This is definitely a heuristic approach, but we believe it is fast and practically useful. Hence, suppose that we have n observed pairs $(x_{i,k}, s_k)$, $1 \le k \le n$. In this case,

$$\beta_{i,w} = \frac{\mathbf{E}[X_i w(S)] - \mathbf{E}[X_i]\mathbf{E}[w(S)]}{\mathbf{E}[S w(S)] - \mathbf{E}[S]\mathbf{E}[w(S)]} \approx \widehat{\beta}_{i,w}, \tag{5.1}$$

where

$$\widehat{\beta}_{i,w} = \frac{\sum_{k=1}^n x_{i,k} w(s_k) - \widehat{x}_i \sum_{k=1}^n w(s_k)}{\sum_{k=1}^n s_k w(s_k) - \widehat{s} \sum_{k=1}^n w(s_k)} \tag{5.2}$$

with

$$\widehat{x}_i = \frac{1}{n}\sum_{k=1}^n x_{i,k} \quad \text{and} \quad \widehat{s} = \frac{1}{n}\sum_{k=1}^n s_k. \tag{5.3}$$

The dependence of $\widehat{\beta}_{i,w}$ on w can now be explored numerically.

Ratio $\beta_{i,w}$: an alternative computational formula. There might be situations when in addition to realizations s_k, $1 \le k \le n$, there is also an estimate $\widehat{r}_i(s)$ of the regression function $r_i(s)$. In this case,

$$\beta_{i,w} = \frac{\mathbf{Cov}[r_i(S), w(S)]}{\mathbf{Cov}[S, w(S)]} \approx \widetilde{\beta}_{i,w}, \tag{5.4}$$

where

$$\widetilde{\beta}_{i,w} = \frac{\sum_{k=1}^{n} \widehat{r}_i(s_k) w(s_k) - \widetilde{x}_i \sum_{k=1}^{n} w(s_k)}{\sum_{k=1}^{n} s_k w(s_k) - \widehat{s} \sum_{k=1}^{n} w(s_k)} \tag{5.5}$$

with

$$\widetilde{x}_i = \frac{1}{n} \sum_{k=1}^{n} \widehat{r}_i(s_k). \tag{5.6}$$

Based on this, we can now explore the dependence of $\widetilde{\beta}_{i,w}$ on w numerically.

Ratio $\beta_{i,w,\text{Gini}}$: a computational formula. Suppose that we have observed pairs $(x_{i,k}, s_k)$, $1 \le k \le n$. Using them, we estimate $\mathbf{E}[X_i]$ and $\mathbf{E}[S]$ by $\widehat{x}_i$ (or $\widetilde{x}_i$) and $\widehat{s}$, respectively, whose definitions are given above. Then

$$\beta_{i,w,\text{Gini}} = \frac{\mathbf{E}[X_i w(F(S))] - \mathbf{E}[X_i] \int_0^1 w(u)\mathrm{d}u}{\mathbf{E}[S w(F(S))] - \mathbf{E}[S] \int_0^1 w(u)\mathrm{d}u} \approx \widehat{\beta}_{i,w,\text{Gini}} \tag{5.7}$$

with

$$\widehat{\beta}_{i,w,\text{Gini}} = \frac{\sum_{k=1}^{n} x^*_{i,k,n} w(k/n) - \left(\sum_{k=1}^{n} x_{i,k}\right) \int_0^1 w(u)\mathrm{d}u}{\sum_{k=1}^{n} s_{k:n} w(k/n) - \left(\sum_{k=1}^{n} s_k\right) \int_0^1 w(u)\mathrm{d}u}, \tag{5.8}$$

where $s_{1:n} \le s_{2:n} \le \cdots \le s_{n:n}$ are ordered s_k, $1 \le k \le n$, and where $x^*_{i,k,n}$, $1 \le k \le n$, are the first coordinates of the pairs $(x_{i,k}, s_k)$, $1 \le k \le n$, ordered according to the ascending second coordinates. In the statistical literature, $x^*_{i,k,n}$, $1 \le k \le n$, are called the induced (by s_k, $1 \le k \le n$) order statistics of $x_{i,k}$, $1 \le k \le n$. We can now explore the dependence of $\widehat{\beta}_{i,w,\text{Gini}}$ on w numerically.

Ratio $\beta_{i,w,\text{Gini}}$: an alternative computational formula. We may also proceed by connecting $\beta_{i,w,\text{Gini}}$ with so-called L-estimates. To somewhat simplify the presentation, we assume that the cdf F of S is a continuous function, in which case $F(S)$ is a uniform on $[0,1]$ random variable. Using the classical notation F^{-1} for the inverse (i.e., quantile) function of F, and with the regression function $r_i(s)$ defined by equation (3.6), we have the equations

$$\begin{aligned}
\beta_{i,w,\text{Gini}} &= \frac{\mathbf{E}[X_i w(F(S))] - \mathbf{E}[X_i] \int_0^1 w(u)\mathrm{d}u}{\mathbf{E}[S w(F(S))] - \mathbf{E}[S] \int_0^1 w(u)\mathrm{d}u} \\
&= \frac{\mathbf{E}[r_i(S) w(F(S))] - \mathbf{E}[X_i] \int_0^1 w(u)\mathrm{d}u}{\mathbf{E}[S w(F(S))] - \mathbf{E}[S] \int_0^1 w(u)\mathrm{d}u} \\
&= \frac{\int_0^1 r_i(F^{-1}(u)) w(u)\mathrm{d}u - \mathbf{E}[X_i] \int_0^1 w(u)\mathrm{d}u}{\int_0^1 F^{-1}(u) w(u)\mathrm{d}u - \mathbf{E}[S] \int_0^1 w(u)\mathrm{d}u}.
\end{aligned} \tag{5.9}$$

Given observed pairs $(x_{i,k}, s_k)$, $1 \le k \le n$, we estimate $\mathbf{E}[X_i]$ by $\widehat{x}_i$ (or $\widetilde{x}_i$) and $\mathbf{E}[S]$ by $\widehat{s}$. The integral

$$L_w = \int_0^1 F^{-1}(u) w(u)\mathrm{d}u$$

in the denominator on the right-hand side of equation (5.9) is known in the statistical literature as L-functional, and its estimate is

$$\widehat{L}_w = \sum_{k=1}^{n} s_{k:n} \int_{(k-1)/n}^{k/n} w(u)\mathrm{d}u,$$

where $s_{1:n} \le s_{2:n} \le \cdots \le s_{n:n}$ are ordered s_k, $1 \le k \le n$. As to the integral

$$L_w(r_i) = \int_0^1 r_i(F^{-1}(u))w(u)\mathrm{d}u$$

in the numerator on the right-hand side of equation (5.9), we have the estimate

$$\widehat{L}_w(\widehat{r}_i) = \sum_{k=1}^{n} \widehat{r}_i(s_{k:n}) \int_{(k-1)/n}^{k/n} w(u)\mathrm{d}u,$$

where $\widehat{r}_i(s)$ is an estimate of the regression function $r_i(s)$. In summary, we have

$$\widetilde{\beta}_{i,w,\mathrm{Gini}} = \frac{\widehat{L}_w(\widehat{r}_i) - \widehat{x}_i \int_0^1 w(u)\mathrm{d}u}{\widehat{L}_w - \widehat{s} \int_0^1 w(u)\mathrm{d}u}. \tag{5.10}$$

We may of course use $\widetilde{x}_i$ given by equation (5.6) instead of $\widehat{x}_i$ on the right-hand side of equation (5.10). We can now explore the dependence of $\widetilde{\beta}_{i,w,\mathrm{Gini}}$ on w numerically.

Acknowledgements

We thank the audience of our invited presentation on the topic at the 2016 Annual Meeting of the Casualty Actuarial Society (CAS) in Orlando, Florida, for enlightening discussions and suggestions, and we are grateful to CAS for an award that has enabled us to concentrate our time and energy on this project.

References

Brazauskas, V., Jones, B.L. and Zitikis, R. (2009). Robust fitting of claim severity distributions and the method of trimmed moments. Journal of Statistical Planning and Inference, 139, 2028–2043.

Furman, E. and Zitikis, R. (2008a). Weighted premium calculation principles. Insurance: Mathematics and Economics, 43, 263–269.

Furman, E. and Zitikis, R. (2008b). Weighted risk capital allocations. Insurance: Mathematics and Economics, 43, 263–269.

Furman, E. and Zitikis, R. (2009). Weighted pricing functionals with application to insurance: an overview. North American Actuarial Journal, 13, 483–496.

Furman, E. and Zitikis, R. (2010). General Stein-type covariance decompositions with applications to insurance and finance. ASTIN Bulletin, 40, 369–375.

Furman, E. and Zitikis, R. (2016a). The Capital Asset Pricing Model: An Insurance Variant. CAS Technical Report, Arlington, VA.

Furman, E. and Zitikis, R. (2016b). Beyond the Pearson correlation: heavy-tailed risks, weighted Gini correlations, and a Gini-type weighted insurance pricing model. (Submitted for publication.)

Furman, E., Wang, R. and Zitikis, R. (2017). Gini-type measures of risk and variability: Gini shortfall, capital allocations, and heavy-tailed risks. (Submitted for publication.)

Heilmann, W.-R. (1989). Decision theoretic foundations of credibility theory. *Insurance: Mathematics and Economics* 8, 77–95.

Kleefeld, A. and Brazauskas, V. (2012). A statistical application of the quantile mechanics approach: MTM estimators for the parameters of t and gamma distributions. European Journal of Applied Mathematics, 23, 593–610.

Klugman, S.A., Panjer, H.H. and Willmot, G.E. (2008). Loss Models: From Data to Decisions. (3rd edition.) Wiley, New Jersey.

Levy, H. (2011). The Capital Asset Pricing Model in the 21st Century: Analytical, Empirical, and Behavioral Perspectives. Cambridge University Press, Cambridge.

Owen, J., and Rabinovitch, R. (1983). On the class of elliptical distributions and their applications to the theory of portfolio choice. Journal of Finance, 38, 745–752.

Su, J. (2016). Multiple Risk Factors Dependence Structures with Applications to Actuarial Risk Management. Ph.D. Thesis, York University, Ontario.

Su, J. and Furman, E. (2017). A form of multivariate Pareto distribution with applications to financial risk measurement. ASTIN Bulletin, 47, 331–357.

Wang, S.S. (1995). Insurance pricing and increased limits ratemaking by proportional hazards transforms. Insurance: Mathematics and Economics 17, 43–54.

Wang, S.S. (1996). Premium calculation by transforming the layer premium density. ASTIN Bulletin 26, 71–92.

Yitzhaki, S. and Schechtman, E. (2013). The Gini Methodology: A Primer on a Statistical Methodology. Springer, New York.

Compendium of credit risk resources

Jean-Philippe Boucher, Mathieu Boudreault and Jean-François Forest-Desaulniers

March 13, 2017

Abstract

This compendium summarizes the various aspects of credit risk that are important to insurance companies in general, namely corporate credit risk (single and multi-name), typical credit-sensitive securities, credit risk for individuals (including mortgage insurance), municipal credit risk, sovereign credit risk, counterparty risk, and regulatory and enterprise risk management. The document also includes considerations for property and casualty insurers and about their practices. Finally, we also list and link to important resources for practitioners and graduate students.

Keywords

- Actuarial Applications & Methodologies
  - Capital management: Capital allocation, Capital requirements
  - Dynamic risk modeling: ALM, solvency analysis
  - Enterprise risk management: Analyzing/Quantifying risks, Financial Risks
  - Regulation and law: Rating agencies, Risk-based capital, Solvency
- Business Areas
  - Credit
  - Surety
- Financial and Statistical Methods
  - Asset and econometric modeling
    - Asset classes (ABS, Corporate bonds, Equities, MBS, Municipal bonds)
    - Credit spreads
- Practice Areas
  - Consulting
  - Risk management

Contents

Chapter 1

Fundamentals of Credit Risk

1.1 Credit risk

An organization (company, government, etc.) that has issued debt (or any other security) or has been extended credit and is unable to meet its obligations, partially or fully, is deemed to be insolvent. Thus, credit risk arises from the potential loss a lender may suffer due to the borrower's insolvency. Credit risk has two components: the uncertainty related to the timing of default (which may never occur) and the amount of loss at default (loss given default, which is the inverse of the recovery rate given default). A financial instrument (asset, derivative, etc.) with a price that depends directly or indirectly on the solvency of the underlying company is known as a credit-sensitive instrument (asset, security, etc.).

There are several types of credit risk, depending on the type of issuer of credit-sensitive securities:

- Corporate credit risk, which involves private and public companies, primarily through corporate bonds but also stocks, credit default swaps, etc.
- Consumer credit risk, which involves ordinary people through credit cards, lines of credit, loans, mortgages, etc.
- Sovereign, state/provincial/county, municipal credit risk, which arises primarily from bonds issued by countries and their governments or their entities (such as utilities);

Losses resulting from credit risk can be extremely large and can easily spread to impact the solvency of several companies and even an entire economy. Life insurance companies and pension plans are heavily invested in long-term bonds to match long-term cash flows and are thus particularly exposed to credit risk.

There are many of examples of insolvency in the recent past. For example, Lehman Brothers went bankrupt in September 2008, and General Motors filed for bankruptcy protection in June 2009 and reorganized the company thereafter. Russia defaulted on its debt in 1998, whereas Greece restructured its debt in 2012. Finally, the city of Detroit filed for bankruptcy in 2011.

1.2 Legal aspects of credit risk

In the credit risk literature, various terms are often employed to define a similar event: insolvency, default, bankruptcy, credit event, etc. However, there are subtle differences among these terms, and thus, this section will clarify some legal terms used to define credit risk.

A default results from the failure of a debtor to make a payment on a debt, whereas insolvency is the legal term equivalent to default. Bankruptcy is tied to an order from the court supervising an insolvent firm. Moreover, a credit event is a term often used when a credit derivative is established. Usually, the contract needs to specify all the events that trigger a payment. For participants belonging to the International Swaps and Derivatives Association (ISDA), credit events must be defined in the ISDA Master Agreement. Finally, technical default is a term used to designate a quasi-default or a default that has been avoided by government intervention, such as a distressed sale or exchange.

When a company nears insolvency or has defaulted on a payment, the company may seek protection from creditors through the courts. In the US, the Bankruptcy Code (which is technically known as *Title 11 of the*

United States Code) establishes the various types of bankruptcies:

- Chapter VII: bankruptcy for individuals and corporations, involving liquidation of assets
- Chapter IX: bankruptcy for cities
- Chapter XI: bankruptcy for individuals and corporations, involving restructuring of debt

In Canada, the *Bankruptcy and Insolvency Act* oversees the bankruptcy process for individuals and corporations.

Corporate defaults thus generally fit into chapters VII and XI, but most corporations first seek to renegotiate the terms of their debt while protected by the courts (Chapter XI bankruptcy). That was the case for General Motors in 2009, which reorganized its debt and eliminated several automotive brands. Lehman Brothers also initially filed for Chapter XI bankruptcy before Barclays acquired the investment bank the following day. Few companies go directly to a Chapter VII bankruptcy, although that was indeed the case for consumer electronics retailer Circuit City (acquired much later) and video game developer and publisher Acclaim. Many financial institutions and investment banks technically defaulted in 2008 after requesting help from the US government through the Troubled Asset Relief Program (TARP) to find a buyer for their toxic assets.

There are various types of creditors in a company who are entitled to the cash flows of the debt when a company goes into restructuring or liquidation. The most senior creditors are always repaid first when the company is insolvent, whereas junior creditors and stockholders usually receive what remains. This is known as the seniority of a debt, whereas the latter money distribution mechanism is known as the absolute priority rule. For example, when a Canadian individual goes bankrupt, any amount owed to the Canada Revenue Agency (the Canadian equivalent of the IRS) has to be paid first, followed by mortgages, credit cards and other types of loans. Senior debt is thus less risky and trades at a lower interest rate.

When countries, states or cities default, they can restructure their debt with their creditors or seek help from other countries. For example, in the period from 2010 to 2012, Greece received help from the International Monetary Fund (IMF) and other European countries in exchange for implementing drastic austerity measures. Privatization of national services and utilities can also be used to quickly raise money in the event of default.

1.3 Credit risk assessment

Given the extent of the risk involved, it is very important for an investor to assess the credit risk on a security. However, making such an assessment is far from easy. Although it is possible for an investor to use models and data to infer the quality of an asset, there are firms that specialize in evaluating the credit risk of any entity (corporation, municipality, country, etc.).

1.3.1 Credit Rating Agencies

A credit rating agency is a private corporation with the core business of assessing the quality of a debt contract. Investors demand information on the quality of the issuer's credit, and typically, the issuer pays the rating agency to evaluate the credit risk linked to a specific debt issue by considering the likelihood of the firm's insolvency, the loss at default and the seniority of the issue.

A credit rating is a grade or score associated to a specific debt issue. Rating agencies usually use letters, numbers, and plus and minus signs to differentiate the various types of ratings. Long- and short-term debt are usually evaluated differently. In the US, the three major credit rating agencies are Moody's, Standard & Poor's, and Fitch. The highest rating for long-term debt is AAA (S&P and Fitch) or Aaa (Moody's), which is considered a prime investment. Ratings between BBB- and AA+ (S&P and Fitch) or Baa3 and Aa1 (Moody's) are investment-grade investments, whereas lower ratings are designated as high-yield or speculative investments. DBRS (formerly known as Dominion Bond Rating Service) is a rating agency founded and headquartered in Canada; the company uses a scale resembling those of its American counterparts.

In the aftermath of the 2008 financial crisis, the rating agencies were highly criticized, notably by the Financial Crisis Inquiry Commission, as well as by countless journalists and economists.

1.3.2 Modeling, valuation and risk management

Credit risk assessment focuses on the likelihood of default (in a broad sense) and the distribution of the loss in the event of default. Such an assessment can be used for risk management and/or to price credit-sensitive instruments. Credit risk is complicated to model, and unfortunately (from the modeler's point of view), defaults occur very rarely, thereby further complicating the evaluation of credit risk.

Statistical approaches seek to uncover factors that can explain a firm's survival or default. They rely, for example, on financial statements, industry data, default counts, transitions (from one credit rating to another) and similar information. A probit regression is an example of a statistical model that can be used to approximate a company's default likelihood (probit models are discussed in Chapter 4). One of the earliest attempts to evaluate credit risk is Altman's Z-Score, which is based on financial ratios.

Actuarial approaches usually represent the total loss in a portfolio in a manner similar to aggregate loss models in actuarial mathematics. Refinements are usually necessary to account for the possible dependence relationship between the number of defaults in a portfolio and the individual loss on each debt issue. CreditRisk+ by CreditSuisse is a famous example of an actuarial credit risk model (details are provided in Chapter 3).

Actuarial and statistical approaches usually perform better when the portfolio (the number of different firms) is large and sufficient data are available. They are designed primarily for risk management purposes because they cannot be used to consistently price credit-sensitive financial instruments.

Credit risk models can also be designed with the intent of pricing credit-sensitive securities such as stocks, corporate bonds, and credit default swaps. It is important to insist that the term "pricing" is defined consistently with that in financial engineering, i.e., a price that prevents arbitrage opportunities. For example, in the earliest credit risk model, Merton (1974) views stockholder equity as a call option on the firm's assets. Further details on credit risk models can be found in Chapters 2-3.

1.4 References

- Altman, E.I. (2000). Predicting financial distress of companies: revisiting the Z-score and ZETA models. Stern School of Business, New York University. http://pages.stern.nyu.edu/~ealtman/Zscores.pdf
- Merton, R.C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance* 29, 449–470.

1.5 List of resources

See also Chapter 12.

1.5.1 Books

Presented in alphabetical order.

- Bielecki, T.R., M. Rutkowski (2002). Credit Risk: Modeling, Valuation and Hedging, Springer Finance.
  - As with other Springer Finance books, the treatment of credit risk models is highly technical, but this is one of the few that rigorously addresses the mathematics behind these models. A solid background in mathematics is required.
  - Chapter 1 takes a broad view of credit risk and is perhaps the easiest chapter to understand.
  - Technical/Mathematical level: Accessible/technical
- Bluhm, C., L. Overbeck, C. Wagner (2010). An Introduction to Credit Risk Modeling, Second Edition, CRC Press.
  - Well-rounded book that covers many areas of corporate credit risk.
  - Targeted to both professionals and academics.
  - Some basics of credit risk management are discussed in Chapter 1, while Chapter 6 examines default probabilities.
- Crouhy, M., D. Galai, R. Mark (2000). Risk management, McGraw-Hill.
  - Crouhy et al. (2000)'s book is a well-rounded book that discusses the regulatory system and capital requirements and exhaustively covers credit risk.
  - The same authors published a book in 2014 (Essentials of risk management) that does not seem to go into as much detail on credit risk.
  - Credit ratings and ratings agencies are discussed in Chapter 7.
  - Technical/Mathematical level: Very accessible
- De Servigny, A., O. Renault (2004). Measuring and Managing Credit Risk, McGraw-Hill.
  - An excellent book written for practitioners, devoted entirely to the topic of (corporate) credit risk.
  - Chapter 1 covers the fundamentals of credit risk in a broader context, whereas Chapter 2 examines credit ratings.
  - Technical/Mathematical level: Very accessible
- Duffe, D., K.J. Singleton (2003). Credit Risk: Pricing, Measurement, and Management, Princeton Series in Finance.
  - Book mostly targeted to academics and professionals with a solid background in mathematics.
  - Basics of credit risk are covered in Chapter 2.
  - Technical/Mathematical level: Accessible/technical
- Fabozzi, F.J., S. Mann (2012). The Handbook of Fixed Income Securities, 8th edition, McGraw-Hill.
  - A definite must-have book for those interested in various aspects of credit risk (especially corporate and municipal). At over 1800 pages, the book covers a very wide range of topics.
  - The fundamentals of credit risk are briefly discussed in Chapters 1 and 12.
  - Technical/Mathematical level: Very accessible
- Hull, J.C. (2014). Options, futures and other derivatives, 9th edition, Pearson.
  - Hull's book is always a good starting point for learning about credit risk, a classic finance text for MBA students.
  - Chapter 24 covers (corporate) credit risk in general with a few subsections on credit

ratings, historical default probabilities and recovery rates.
  - Technical/Mathematical level: Very accessible
- Hull, J.C. (2015). Risk Management and Financial Institutions, Wiley Finance.
  - Book mostly aimed at practitioners.
  - Chapter 19 is an excellent introduction to the various aspects of corporate credit assessment and valuation.
  - Technical/Mathematical level: Very accessible.

1.5.2 Websites and online reports

- US Bankruptcy Code for individuals and corporations (Official government website): http://www.uscourts.gov/services-forms/bankruptcy. Links to the various bankruptcy Chapters are clearly indicated.
  - US Code: Title 11 - Bankruptcy, Chapter 7 - Liquidation (through Cornell University Law School): https://www.law.cornell.edu/uscode/text/11/chapter-7
  - US Code: Title 11 - Bankruptcy, Chapter 11 - Reorganization (through Cornell University Law School): https://www.law.cornell.edu/uscode/text/11/chapter-11
  - US Code: Title 11 - Bankruptcy, Chapter 13 - Adjustment of Debts of an Individual with Regular Income (bankruptcy for individuals) through Cornell University Law School): https://www.law.cornell.edu/uscode/text/11/chapter-13

- Canadian bankruptcy laws:
  Summary from Industry Canada: https://www.ic.gc.ca/eic/site/cilp-pdci.nsf/eng/h_cl00021.html
  - *Bankruptcy and Insolvency Act* (Liquidation, corporations and individuals): http://laws-lois.justice.gc.ca/eng/acts/b-3/ (Official government website)
  - *Companies' Creditors Arrangement Act* (Reorganization): http://laws-lois.justice.gc.ca/eng/acts/C-36/ (Official government website)
  - US Department of Treasury, Troubled Asset Relief Program (TARP) (Official government website): http://www.treasury.gov/initiatives/financial-stability/TARP-Programs/Pages/default.aspx
- Ratings agencies' official websites:
  - Moody's: https://www.moodys.com/
    - Moody's Analytics – Moody's KMV: http://www.moodysanalytics.com/About-Us/History/KMV-History
  - Standard & Poor's: http://www.standardandpoors.com/en_US/web/guest/home
  - Fitch: https://www.fitchratings.com/
  - Dominion Bond Ratings Services (DBRS): http://www.dbrs.com/
- Ratings definitions:
  - Moody's (March 2015): https://www.moodys.com/researchdocumentcontentpage.aspx?docid=PBC_79004
  - Standard & Poor's (August 2016): https://www.standardandpoors.com/en_US/web/guest/article/-/view/sourceId/504352
  - Fitch (December 2014): https://www.fitchratings.com/web_content/ratings/fitch_ratings_definitions_and_scales.pdf
- Annual corporate default and recovery rate studies:
  - Moody's (2008 study, latest public study): https://www.moodys.com/sites/products/DefaultResearch/2007400000578875.pdf
  - Moody's (up to 2016, not public): https://www.moodys.com/Pages/GuideToDefaultResearch.aspx
  - Standard & Poor's (2014 study, in PDF format, unofficial source):

http://www.nact.org/resources/2014_SP_Global_Corporate_Default_Study.pdf (from the National Association of Corporate Treasurers (NACT).
- Standard & Poor's (2014 study, in HTML, official source): can be found through their Global Credit Portal https://www.globalcreditportal.com/ or through the Search tool from S&P's official website. Title of the report: Annual Global Corporate Default Study And Rating Transitions.
- Fitch (2014 study, through NRSRO Annual Certification): https://www.fitchratings.com/web_content/nrsro/nav/NRSRO_Exhibit-1.pdf
- DBRS (2014 study, official): http://www.dbrs.com/research/278497/2014-dbrs-corporate-rating-transition-and-default-study.pdf

1.5.3 Data

In addition to the major databases mentioned in Chapter 12, two additional databases focus on bankruptcy filings.

- UCLA (University of California, Los Angeles) LoPucki Bankruptcy Research Database (BRD): Chapter 7 and 11 filings for companies with over $100 million in assets: http://lopucki.law.ucla.edu/ Sample data available.
- New Generation Research Bankruptcy Data: http://www.bankruptcydata.com/

Historical credit ratings by companies, sorted by credit ratings agency (see Chapter 12).

1.5.4 Computer programs

Matlab programming language:

- Credit risk modeling with Matlab

Chapter 2

Single-name corporate credit risk models

A wealth of scientific and professional literature has been published on this subject, and this chapter summarizes the most significant approaches proposed to assess the credit risk of debt issuers.

2.1 Structural models

Structural credit risk models rely on an explicit definition of the company's assets, liabilities and equity. Default is triggered when assets are insufficient to meet the company's obligations, either at the exact moment the payment is due or before.

2.1.1 Merton (1974)

Merton's (1974) model is the cornerstone of modern credit risk assessment. The idea is that a firm is composed of risky assets (A_t) and has committed to paying a specific sum of money (F) on a known future date (T), which is the maturity of this debt. Before the debt matures, the company pursues normal business, regardless of the valuation of the assets. At debt maturity T, the firm is dissolved and the assets redistributed between debtholders and equityholders according to the absolute priority rule, whereby the creditors are paid in full first and the equityholders receive the remainder.

Using mathematical notation, we have

$$\text{payment to debtholders at } T = D_T = F - max(F - A_T, 0)$$
$$\text{payment to equityholders at } T = E_T = max(A_T - F, 0).$$

Thus, one can view equity as a call option on the firm's assets, whereas the payment to debtholders corresponds to the face value of the debt (the promised payment) minus a put option on the assets, which represents the debtholders' loss given default. Therefore, using put-call parity, we easily recover the fundamental accounting equation, which is

$$A_T = L_T + E_T$$
$$A_T = F - max(F - A_T, 0) + max(A_T - F, 0).$$

That is, the total value of assets is equal to the sum of liabilities and equity.

When the firm survives, the debtholders recover F (the put option is out-of-the-money), whereas the equityholders have a right to the excess of A_T over F. When the firm defaults, the debtholders have a right to the liquidated value of the firm (A_T), and debtholders suffer a loss of $F - A_T$. However, equityholders receive nothing (the call option is out-of-the-money).

Assuming that the market value of the assets of the company evolves as a geometric Brownian motion, which is the standard Black-Scholes assumption, one can use the Black-Scholes equations to value the company's equity and liabilities.

In computing the value of the equity (or of the liabilities), one important quantity often arises, which is

$$d_1^{\mu} = \frac{\ln\left(\frac{A_T}{F}\right)+\mu(T-t)}{\sigma\sqrt{T-t}}.$$

d_1^{μ} is known as a normalized "distance to default" metric[1]. Indeed, the ratio of the assets over the face value is some form of distance that has to be adjusted by the potential growth of the assets ($\mu(T-t)$) and by the volatility of the assets ($\sigma\sqrt{T-t}$). The original founders of the company KMV, Kealhofer, McQuown and Vasicek, designed an expected default frequency (EDF) based on the distance to default.

Merton's model is very intuitive and has done much to foster our understanding of the dynamics between the capital structure and the firm's credit risk. However, the default triggering mechanism is unrealistic, as most creditors would not wait until maturity to claim a fraction of the assets. Moreover, at least from a scientific point of view, the asset and liability dynamics are overly simplistic.

2.1.2 First-passage models

In first-passage credit risk models, default is triggered as soon as the value of the assets crosses a given barrier, known as the default barrier. This barrier is likely to be different from the value of the promised payment(s), as debtholders are risk-averse and would intervene to maximize the likelihood of being paid. First-passage models are similar in spirit to ruin theory, according to which the claim and premium arrival processes are used to represent the insurer's ruin.

To contrast the first-passage and Merton models, we will rely on Figure 2.1. Suppose that the debt matures in 20 years. In Merton's model, the firm easily survives because the assets (700) exceed the liabilities (200). However, we see that at approximately 10 years, the assets have fallen below the value of liabilities, and thus, in a first-passage model, that company would have defaulted.

First-passage models are also intended to quantify value for debtholders and equityholders. The idea is similar to Merton's, namely that equityholders and debtholders have a right to assets whenever there is a default. Thus, whenever the value of assets crosses the default barrier, the absolute priority rule applies, and debtholders are paid in full before equityholders. Instead of having plain vanilla call and put options to represent the equity value and the loss given default, in a first-passage model, we instead have barrier call and put options. The authors who pioneered the first-passage model and equity pricing include Black & Cox (1976) and Brockman & Turtle (2003). At present, most structural models are based on the first passage of assets across a default barrier.

A notable professional model was also inspired by the spirit of first-passage models, namely CreditMetrics by RiskMetrics (the model was formerly owned by JP Morgan). In CreditMetrics, the creditworthiness of a firm evolves according to a Markov chain, the role of which is to mimic changes

[1] Note that when d_1 is used to price equity and liabilities, μ has to be replaced with *r* to be consistent with absence of arbitrage arguments.

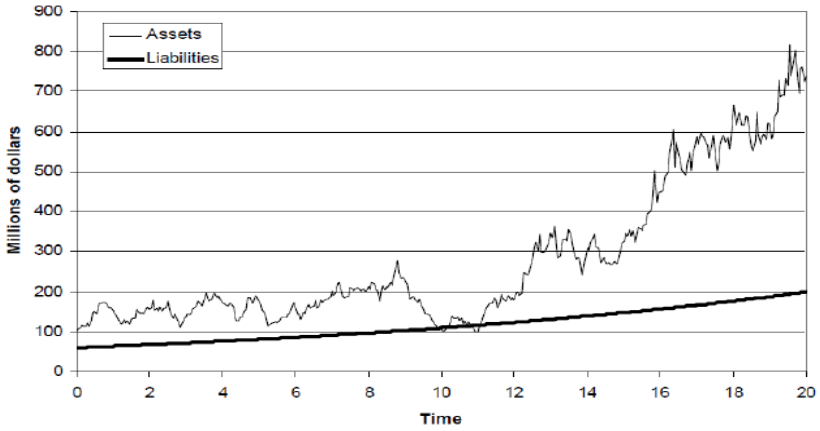


Figure 2.1: Evolution of assets and liabilities of a fake company

in the credit ratings of a company over time. For example, a firm can evolve from A to B, AAA or D (default). Further details can be found in Chapter 3.

2.1.3 Other structural models

Numerous extensions have been proposed to extend Black & Cox (1976) with the similar goal of evaluating the equityholder's and debtholder's value. For example, stochastic interest rates were introduced in Longstaff & Schwartz (1995) and Briys & de Varenne (1997). Collin-Dufresne & Goldstein (2001) introduced a structural model with additional debt issuance, thus allowing for mean reversion in leverage ratios under a stochastic interest rate environment. Finally, Leland (1994) and Leland & Toft (1996) introduced models in which debt is a (perpetual) coupon bond and the default trigger is endogenously determined by external frictions such as bankruptcy costs, and taxes.

2.1.4 Conclusion

Structural models are extremely useful for better understanding the interplay of assets and debt in valuing a firm and its default risk. However, most structural models are unable to replicate the level of credit spreads observed for short-term securities.

The main reason for this deficiency is that in a structural model, default or survival is predictable in the very short term. When the distance to default is large, which means that the firm's assets are currently far from the default barrier, multiple declines in asset value need to be generated in the short term to generate a default. This event is generally nearly impossible to include in a model, and hence, these models cannot replicate the observed level of credit spreads for short-term securities (see the discussion by Jones, Mason & Rosenfeld (1984) and more recently the books mentioned below). It is clear that default holds an element of surprise that increases short-term credit spreads.

When structural models are used to value debtholders' and equityholders' shares of the assets, the absolute priority rule is a crucial assumption. However, according to Eberhart et al. (1990) and more recently Bris et al. (2006), deviations from the absolute priority rule are common and are even expected and priced in the financial markets.

2.2 Reduced-form models

Instead of modeling the assets and the debt of a company to obtain the value of the debtholders and equityholders, a reduced-form credit risk model takes the completely opposite view. The idea is

that we directly model the instantaneous default probability (also known as default intensity, hence intensity models) as a stochastic process. This process is completely unrelated to the firm's capital structure. The stochastic process is chosen and estimated to replicate the observed prices of credit-sensitive assets as closely as possible.

Default intensity or forces of mortality are similar concepts in which we model the instantaneous probability of the event: default or death. Even if death probabilities can evolve over time with the policyholder's behavior, it is clear that default probabilities will vary considerably more over time. Therefore, reduced-form models have grown to be increasingly sophisticated to better fit prices of credit-sensitive securities.

The idea of a reduced-form model was pioneered by Jarrow & Turnbull (1995) and further extended in Jarrow et al. (1997). In the former article, a model similar to an exponential distribution was used for the moment of default, whereas in the latter, the rate of arrival of default was modeled as a continuous-time Markov chain, thus mimicking ratings transitions from credit rating agencies. Other important and early contributions to reduced-form models are Lando (1998) and Duffie & Singleton (1999).

From the latter articles, the number of reduced-form models has exploded over the years as authors have often changed the default intensity process. Because the *T*-year default probability (as opposed to default intensity) is computationally equivalent to the price of a zero-coupon bond in a stochastic interest rate model, the former literature has borrowed much from the latter. The paper by Duffie, Pan & Singleton (2000) helped many authors to find closed-form formulas for default probabilities. Because the speed of calibration and estimation is determined primarily by the speed at which these quantities are computed, analytical expressions can become substantial.

Whereas structural models use information from the firm's capital structure to deduce default probabilities and the price of credit-sensitive securities, reduced-form models use the observed prices of these securities to deduce the market's view of the firm's solvency. Therefore, parameters obtained from such a calibration are only valid to price other credit-sensitive instruments.

2.3 Hybrid models

Although the fit of reduced-form models to security prices is known to be much better than that of structural models, reduced-form models lack the financial intuition and justification of structural models. However, structural and reduced-form models can be viewed as a single, consistent type of model, depending on the amount of information observed by an investor. The information paradigm of credit risk models was introduced by Duffie & Lando (2001) and further justified by Jarrow & Protter (2004).

To be usable in practice, structural models rely on full and continuous knowledge of the market value of the company's assets and debt structure. However, only firm managers have that much information, whereas investors only receive partial and periodic information from the company, often in the form of accounting statements. Duffie & Lando (2001) demonstrate that for a given capital structure, when investors receive only partial information, default may prove to be partially predictable and thus better replicate short-term credit spreads. Contrarily, when investors have absolutely no information, the perceived credit risk may behave as in a reduced-form model. The result is what is known as a hybrid credit risk model, i.e., a model that features components of both structural and reduced-form credit risk models.

2.4 Loss given default

The vast majority of the previous models focus primarily on determining the likelihood of default, whereas another component of credit risk has an important impact on the prices of credit-sensitive securities: the loss given default (LGD) (or, conversely, the recovery rate). Altman et al. (2005) and Acharya et al. (2007) report an inverse relationship between default probabilities and recovery rates. In other words, solvent companies that default (say, for liquidity reasons) have higher

recovery rates (and lower LGDs), whereas insolvent companies have lower recovery rates[2]. Thus, the factors that explain the likelihood of default also seem to explain the LGD. Altman et al. (2005) report that losses can be highly underestimated when the relationship between the two is ignored.

Earlier attempts to integrate stochastic recovery rates possibly related to company solvency are Frye (2000), Jarrow & Yu (2001) and Jokivuolle & Peura (2003). For example, Frye (2000) and Jokivuolle & Peura (2003) focus on the value of the collateral on a loan, with recovery provided by the value of collateral on default. The negative relationship between recovery rates and default intensity/probability is integrated in the models of Gaspar & Slinko (2008), Das & Hanouna (2009), Bruche & Gonzalez-Aguado (2010), Bade et al. (2011) and Boudreault et al. (2013).

When recovery rates are stochastic and depend on the same factors as default, a term structure of recovery rates appears. In other words, the timing of default impacts the distribution of recovery rates in a credit-sensitive instrument. Boudreault et al. (2013) find that the term structure of recovery rates decreases for high-rated firms but increases for low-rated firms. For example, a AAA-rated firm can only maintain or downgrade its rating and hence, its recovery rate can only be stable or decrease in the future. For a CCC-rated firm, if it survives, its solvency will improve (or it will default and be excluded), and thus its recovery rate will increase.

2.5 References

- Acharya, V.V., S.T. Bharath, A. Srinivasan (2007). Does industry-wide distress affect defaulted firms? Evidence from creditor recoveries. Journal of Financial Economics 85, 787–821.
- Altman, E.I. (2006). Default recovery rates and LGD in credit risk modeling and practice: an updated review of the literature and empirical evidence. New York University, Stern School of Business.
- Altman, E.I., V.M. Kishore (1996). Almost everything you wanted to know about recoveries on defaulted bonds. Financial Analysts Journal 52, 57–64.
- Bade, B., D. Rösch, H. Scheule (2011). Default and recovery risk dependencies in a simple credit risk model. European Financial Management 17, 120–144.
- Black, F., J.C. Cox (1976). Valuing corporate securities: Some effects of bond indenture provisions. Journal of Finance 31, 351–367.
- Boudreault, M., G. Gauthier, T. Thomassin (2013). Recovery rate risk and credit spread in a hybrid credit risk model. Journal of Credit Risk 9, 3.
- Briys, E., F. de Varenne (1997). Valuing risky fixed rate debt: An extension. Journal of Financial and Quantitative Analysis 32, 239–248.
- Brockman, P., H.J. Turtle (2003). A barrier option framework for corporate security valuation. Journal of Financial Economics 67, 511–529.
- Bruche, M., C. Gonzalez-Aguado (2010). Recovery rates, default probabilities, and the credit cycle. Journal of Banking & Finance 34, 754–764.
- Collin-Dufresne, P., R.S. Goldstein, J.S. Martin (2001). The determinants of credit spread changes. Journal of Finance, 2177–2207.
- Das, S.R., P. Hanouna (2009). Implied recovery. Journal of Economic Dynamics and Control 33, 1837–1857.
- Duffie, D., K.J. Singleton (1999). Modeling term structures of defaultable bonds. Review of Financial studies 12, 687–720.
- Duffie, D., D. Lando (2001). Term structures of credit spreads with incomplete accounting information. Econometrica, 633–664.
- Eberhart, A.C., W.T. Moore, R.L. Roenfeldt (1990). Security pricing and deviations from the absolute priority rule in bankruptcy proceedings. Journal of Finance, 1457–1469.

[2] This was observed as early as in Altman & Kishore (1996) and in Table III of Elton et al. (2001).

- Elton, E. J., M.J. Gruber, D. Agrawal, C. Mann (2001). Explaining the rate spread on corporate bonds. Journal of Finance 56, 247–277.
- Frye, J. (2000). Collateral Damage: A Source of Systematic Credit Risk , Risk Magazine
- Gaspar, R.M., I. Slinko (2008). On Recovery and Intensity's Correlation – A New Class of Credit Risk Models. Journal of Credit Risk 4, 1–33.
- Jarrow, R.A., P. Protter (2004). Structural vs Reduced Form Models: A New Information Based Perspective, Journal of Investment Management 2.
- Jarrow, R.A., S.M. Turnbull (1995). Pricing derivatives on financial securities subject to credit risk. Journal of Finance 50, 53–53.
- Jarrow, R.A., D. Lando, S.M. Turnbull (1997). A Markov model for the term structure of credit risk spreads. Review of Financial studies, 10(2), 481–523.
- Jarrow, R.A., F. Yu (2001). Counterparty risk and the pricing of defaultable securities. Journal of Finance 56, 1765–1799.
- Jokivuolle, E., S. Peura (2003). Incorporating collateral value uncertainty in loss given default estimates and loan-to-value ratios. European Financial Management, 9, 299–314.
- Jones, E.P., S.P. Mason, E. Rosenfeld (1984). Contingent claims analysis of corporate capital structures: An empirical investigation. Journal of Finance, 611–625.
- Lando, D. (1998). On Cox processes and credit risky securities. Review of Derivatives research 2, 99–120.
- Leland, H.E. (1994). Corporate debt value, bond covenants, and optimal capital structure. Journal of Finance, 1213–1252.
- Leland, H.E., & K.B. Toft (1996). Optimal capital structure, endogenous bankruptcy, and the term structure of credit spreads. Journal of Finance, 987–1019.
- Longstaff, F.A., E.S. Schwartz (1995). A simple approach to valuing risky fixed and floating rate debt. Journal of Finance 50, 789–819.
- Merton, R.C. (1974). On the pricing of corporate debt: The risk structure of interest rates, Journal of Finance 29, 449–470.

2.6 List of resources

See also Chapter 12.

2.6.1 Books

Presented in alphabetical order.

- Ammann, M. (2002). Credit Risk Valuation: Methods, Models, and Applications, Springer
  - Mostly focused on single-name credit risk models, aimed at academics.
  - Chapters 1–5 examine traditional single-name models.
- Bielecki, T.R., M. Rutkowski (2002). Credit Risk: Modeling, Valuation and Hedging, Springer Finance.
  - The mathematics behind structural models (Chapters 2 and 3) and reduced-form models (Chapter 8) are rigorously described.
  - Technical/Mathematical level: Very technical.
- Bluhm, C., L. Overbeck, C. Wagner (2010). Introduction to Credit Risk Modelling, Second Edition, CRC Press.
  - Merton's model is covered in Chapter 3.
- Crouhy, M., D. Galai, R. Mark (2000). Risk management, McGraw Hill
  - Chapter 9 is dedicated to structural models, also known as the contingent claim approach. It is mostly focused on Merton's model and its commercial equivalent, Moody's-KMV.
  - Chapter 10 discusses reduced-form models.

  - Technical/Mathematical level: Very accessible
- De Servigny, A., O. Renault (2004). Measuring and Managing Credit Risk, McGraw-Hill.
  - Merton's model and Moody's-KMV models are described in Chapter 3.
  - It is interesting to note that this book is one of the few that examines credit scoring techniques, i.e., using regression models (statistics) to evaluate solvency (from Altman's Z-score to generalized linear models).
  - There is also an entire chapter devoted to loss given default (recovery rate) and its impact on credit risk management.
  - Technical/Mathematical level: Accessible
- Duffie, D., K.J. Singleton (2003). Credit Risk: Pricing, Measurement, and Management, Princeton Series in Finance.
  - Structural, reduced-form models and some statistical models are covered in Chapter 3. The authors contrast pricing and risk-management applications (physical and risk-neutral probabilities) in Chapter 5.
  - Technical/Mathematical level: Accessible/technical
- Fabozzi, F.J., S. Mann (2012). The Handbook of Fixed Income Securities, 8th edition, McGraw-Hill.
  - Single-name models are very well summarized in Chapter 45. The content is scientifically accurate but citations are centered on the chapter's co-authors.
  - Technical/Mathematical level: Very accessible
- Hull, J.C. (2014). Options, futures and other derivatives, 9th edition, Pearson.
  - Chapter 24 has some subsections on credit risk modeling, but coverage is very light. There is not much on the two families of credit risk models.
  - Technical/Mathematical level: Very accessible
- Lando, D. (2004). Credit risk modeling, Theory and Applications, Princeton University Press.
  - Lando's book reviews a wide range of credit risk models and methods, especially from an academic standpoint.
  - Structural models are exhaustively covered in Chapters 2 and 3, whereas reduced-form models are discussed in Chapter 5.
  - It is worth mentioning that there is an entire chapter devoted to credit scoring techniques using regressions (Chapter 4).
  - Technical/Mathematical level: Technical
- McNeil, A.J., R. Frey, P. Embrechts (2005). Quantitative Risk Management, Princeton University Press.
  - Textbook aimed at advanced undergraduates, graduates or professionals with an applied mathematics background. Covers a wide range of topics in quantitative risk management. Very well done if the reader has the technical knowledge.
  - Chapters 8 and 9 discuss some single-name and portfolio credit risk models in detail.
  - Technical/Mathematical level: Technical
- O'Kane, D. (2008). Modelling Single-name and Multi-name Credit Derivatives, Wiley Finance.
  - Book focusing specifically on pricing credit derivatives. The Wiley Finance series is usually accessible to practitioners.
  - Brief summary of single-name credit risk models in Chapter 3.
  - Technical/Mathematical level: Accessible

2.6.2 Computer programs

Matlab programming language

- Merton structural credit risk model

- Moody's KMV Credit Risk Model Probability of default

Chapter 3

Portfolio corporate credit risk models

In Chapter 1, we laid down the foundations of credit risk, whereas in Chapter 2, we addressed how to assess credit risk for a single corporation. We now extend some of the principles found in Chapter 2 to portfolio credit risk management. To adequately transition from Chapter 2 to this chapter, we must understand that a key aspect of portfolio credit risk is assessing the dependence between assets. Therefore, Section 1 summarizes some of the key concepts of dependence, whereas Sections 2 and 3 briefly review some of the most important classes of portfolio models.

3.1 Notions of dependence

3.1.1 Sources of dependence

From the probabilistic or statistical perspective, dependence between two random variables *X* and *Y* is defined as the influence that a specific realization of *X* exerts on the distribution of outcomes of *Y*. In finance and actuarial science, dependence can occur in several ways: as a direct influence of one asset on the other or as a result of a common source of shock or influence affecting both.

An example of dependence resulting from direct influence occurs in life insurance. The broken heart syndrome is a well-known phenomenon whereby the death of a spouse alters the lifetime distribution of the surviving spouse. Natural catastrophes in P&C (property and casualty) insurance provide many situations of common shocks that simultaneously affect the claims distribution of a portfolio of policyholders. For example, the occurrence of an earthquake can affect thousands of policyholders living close to the epicenter.

In credit risk modeling, dependence also occurs as either a direct influence or a result of exposure to common variables. The most common source of dependence in credit risk is exposure to common macroeconomic (state of the economy, interest rates, etc.) or industry variables (those that affect one sector as a whole but not another). An example of how defaults can directly influence the solvency of other companies occurs when an important insolvency on the asset side of an investor's balance sheet can also provoke its own bankruptcy. This phenomenon is known as a type of contamination process and can occur when large financial institutions default (systemic risk).

3.1.2 Dependence measures

To assess the impact of dependence on portfolio losses, we need to determine the level of dependence contained in the asset portfolio. To do so, we use what is known as a dependence measure. Such a measure evaluates the degree to which *X* influences *Y* and vice versa.

The attentive reader might have noticed that thus far we have discussed dependence without mentioning the word "correlation". Correlation, as defined by Pearson, is

$$Corr(X,Y) = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y}$$

where μ_X and σ_X are the mean and standard deviation of the random variable X, respectively. Correlation describes whether one variable is above its mean when the other variable is either above or below its mean. It is situated between -1 and 1.

Correlation is the unique dependence parameter in the joint normal distribution and plays a fundamental role in linear regressions. This is also why it is known as a linear dependence measure, as it can only capture linear types of dependence. Therefore, it can fail to measure many types of dependence relationships, as shown in Figure 3.1.

There are alternative dependence measures known as rank correlation. The idea of rank correlation is to measure the capability of generating small/large realizations when the other variable has generated a small/large outcome. However, instead of determining the relative size based on the mean, as is the case with Pearson's correlation, the idea is to compare ranks. The most well known rank correlation measures are Kendall's tau and Spearman's rho, two measures included in most statistical software such as R and Matlab.

Finally, a dependence measure that is extremely useful in finance is the tail dependence measure. Tail dependence is the ability of one variable to influence the other when either has generated an extreme outcome. For example, in times of crisis, stocks and other assets traded on financial markets tend to show more dependence than in more "normal" times. Because the tails of a normal distribution are very light (it is incapable of consistently generating large extremes), the joint normal distribution does not have tail dependence.

3.1.3 Creating dependence

To create dependence in a portfolio of credit-sensitive assets, the most popular but not necessarily most appropriate choice is to choose a multivariate normal distribution. In the latter

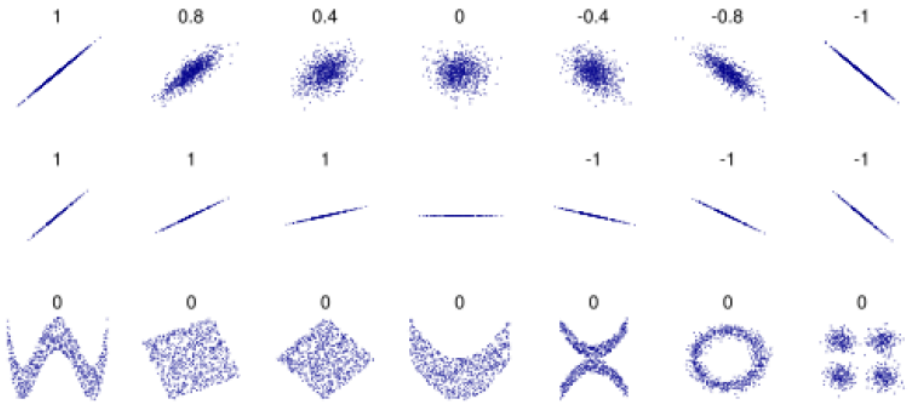


Figure 3.1: Illustration of how Pearson's correlation can fail in detecting several types of dependence (Source: Wikipedia.)

case, dependence between risks is summarized by each pairwise correlation.

Another popular approach used in finance to create dependence between random variables is to use what is known as copulas. A copula is a mechanism that creates dependence using the marginal cumulative distribution functions (c.d.f.s). Mathematically, suppose we have two risks X and Y with known c.d.f.s $F_X(x)$ and $F_Y(y)$. Instead of writing X as a function of Y or linking X and Y with another variable Z, the copula works by assuming that the joint c.d.f. $F_{X,Y}(x,y)$ is some function of $F_X(x)$ and $F_Y(y)$. Therefore,

$$F_{X,Y}(x,y) = C(F_X(x), F_Y(y))$$

where the bivariate function C is known as the copula. The copula creates dependence between X

and Y by linking each c.d.f. $F_X(x)$ and $F_Y(y)$.

Using copulas is very popular because it allows the modeler to separate the problem of fitting the distributions of X and Y and assessing the dependence in the variables. An excellent and accessible reference on the topic is Frees & Valdez (1998).

A very popular copula in finance and credit risk modeling is the Gaussian or normal copula[3]. The Gaussian copula links two or more random variables that are not necessarily normally distributed using the dependence structure that stems from a multivariate normal distribution. For example, X can be gamma distributed, and Y can be distributed as a Weibull, but we can create a dependence structure between X and Y based on their correlation.

The Gaussian copula is tractable; correlation between variables is relatively easy to estimate, but as mentioned earlier, it lacks appropriate tail dependence. The problem remains that for large portfolios, there are many correlations to estimate (large correlation matrix).

There are alternatives to reduce the number of correlations to estimate. One can assume that correlation is constant for subgroups of firms (block constant correlation matrix). Another popular approach is to use factor models. The idea is that risk X can be regarded as the sum of random common factors (common to all firms) and an idiosyncratic element, which is firm specific. The one-factor model is very popular, and when risk factors are Gaussian, we retrieve the Gaussian copula.

The widespread use of Gaussian copulas (and some factor models) in the pricing of collateralized debt obligations (CDOs) (see Chapter 7) was highly criticized in the aftermath of the 2008 financial crisis. However, realistic alternatives to Gaussian copulas for large portfolios (of more than 5–10 assets!) are unfortunately not readily available either. The Student-t copula is one of the few realistic alternatives that are also based on correlations, with a feature that allows tail dependence to be gauged.

3.2 Professional models

Professional portfolio credit risk management models are designed primarily to compute the credit value-at-risk (VaR), i.e., the VaR related to credit losses. Two popular approaches used in the industry are CreditMetrics (by RiskMetrics, MSCI) and CreditRisk+ (by Crédit Suisse). The two methods are briefly described below, with further details provided in their respective technical documents, also below.

3.2.1 CreditMetrics

CreditMetrics is mostly based on ratings provided by agencies such as Standard & Poor's and Moody's. The idea is (1) to evaluate each bond (or any credit-sensitive asset) held in the investment portfolio for each possible ratings transition (upgrade, downgrade and default) and (2) to do so using joint transition probabilities to calculate the credit VaR.

The first step is to compute the hypothetical price of each bond for every possible rating. This is usually done with the 1-period forward yield curves specific to each possible rating.

The second step is to compute the joint transition probabilities, e.g., the probability that issuer no. 1 is upgraded to AA while issuer no. 2 is downgraded to B. While the ratings agencies provide transition matrices, this only applies to one company, and thus yield marginal transition probabilities. To compute joint transition probabilities, CreditMetrics maps ratings to a normal distribution such that the Gaussian copula can be used to compute the joint probabilities.

For large portfolios, CreditMetrics is best handled with a simulation. The multivariate

[3] The Archimedean family of copulas is also well known in statistics (including Frank, Clayton & Gumbel), but we will not discuss these further. This is because when attempting to link more than two random variables in an Archimedean copula, it is difficult to ensure that dependence between all pairs differs from one pair to the next.

Gaussian copula is very easy to simulate, even on large scales. With this simulation, retrieving the transitions for each bond and thus the gain or loss is straightforward.

Correlation is a critical input in CreditMetrics, and theoretically, we require a correlation for each specific pair of firms in the portfolio. For 100 firms, there are 5,000 correlations to estimate. This is a daunting task that can be simplified by using constant correlations across sectors. For example, two companies from the same sector have the same correlation, and the correlation between any two companies in each of the two sectors is also constant (block constant correlation matrix). CreditMetrics suggests using equity correlation, i.e., correlations of stock returns, although this is only an approximation, as stock return dynamics do not necessarily reflect the credit risk dynamics (see. e.g., Boudreault et al. (2015)).

3.2.2 CreditRisk+

The core of the CreditRisk+ model is actuarial in nature. The simplest version assumes a very large portfolio with small individual default probabilities. In this case, the number of defaults in the portfolio can be reasonably approximated by a Poisson distribution. Typical severity distributions can be used to approximate the loss on each default such that the aggregate loss can be evaluated. This simplistic framework does not account for dependence in defaults or individual bond characteristics.

This framework can be extended to account for (a) heterogeneity in the bond loss distribution and (b) dependence in default occurrences. However, doing so means addressing individual default occurrences and every possible loss through a Monte Carlo simulation. Models similar to CreditRisk+ are presented in Chapter 9 of McNeil et al. (2005).

3.3 Academic models

Although they can also be used for risk-management purposes, most academic portfolio credit risk models are also designed for pricing. Pricing in this context is taken in the financial engineering sense, meaning that we are required to find the price that eliminates arbitrage opportunities. In this pricing framework, models are able to price basket credit default swaps (such as *k*-th to default), collateralized debt obligations, and so forth (further details on these products are provided in Chapter 7).

A range of multivariate extensions of structural and reduced-form credit risk models exist. In a multi-name context, the idea is that each company's assets and liabilities are bound using dependence models such as copulas. This will create clusters of defaults in times of crisis. Similarly, in multi-name reduced-form models, the default intensity of each company has some form of dependence, which increases the likelihood of default clusters. Often-cited papers are Li (2000), Duffie & Garleanu (2001), Andersen & Sidenius (2004), and Hull, Predescu & White (2010). Special models of contagion and contamination have appeared in the literature in which the default of one company provokes a sequence of defaults (see, for example, Davis & Lo (2001) and Jarrow & Yu (2001)).

The financial crisis also prompted additional research on the topic of systemic risk, the risk that one very large institution provokes the insolvency of other financial institutions. When investigating banks and insurance companies, a notable finding is that the insolvency of major banks may cause the insolvency of insurance companies but not the opposite. Key papers on that topic are Billio et al. (2012) and Chen et al. (2014). There is also an entire scientific journal devoted to the issue: Journal of Financial Stability (launched in 2004).

Finally, it is well known that recovery rates are inversely proportional to firms' solvency (see Chapter 2, Altman et al. (2006) and Acharya et al. (2007)). During recessions, the number of defaults rises, and we should expect recovery rates to decrease as solvency deteriorates. While this effect on aggregate losses has seldom been investigated, this double-whammy effect is examined in Boudreault et al. (2014).

3.4 References

- Acharya, V.V., S.T. Bharath, A. Srinivasan (2007). Does industry-wide distress affect defaulted firms? Evidence from creditor recoveries. Journal of Financial Economics 85, 787–821.
- Altman, E.I. (2006). Default recovery rates and LGD in credit risk modelling and practice: an updated review of the literature and empirical evidence. New York University, Stern School of Business.
- Andersen, L., Sidenius, J. (2004). Extensions to the Gaussian copula: Random recovery and random factor loadings. Journal of Credit Risk 1
- Billio, M., M. Getmansky, A.W. Lo, L. Pelizzon (2012). Econometric measures of connectedness and systemic risk in the finance and insurance sectors. Journal of Financial Economics 104, 535–559.
- Boudreault, M., G. Gauthier, T. Thomassin (2014). Contagion effect on bond portfolio risk measures in a hybrid credit risk model, Finance Research Letters 11, 131–139.
- Boudreault, M., G. Gauthier, T. Thomassin (2015). Estimation of correlations in portfolio credit risk models based on noisy security prices. Journal of Economic Dynamics and Control 61, 334–349.
- Chen, H., J.D. Cummins, K.S. Viswanathan, M.A. Weiss (2014). Systemic risk and the interconnectedness between banks and insurers: An econometric analysis. Journal of Risk and Insurance 81, 623–652.
- Davis, M., V. Lo (2001). Infectious defaults. Quantitative Finance 1, 382–387.
- Duffie, D., N. Garleanu (2001). Risk and valuation of collateralized debt obligations. Financial Analysts Journal 57, 41–59.
- Frees, E.W., E.A. Valdez (1998). Understanding relationships using copulas. North American Actuarial Journal 2, 1–25.
- Hull, J.C., M. Predescu, A. White (2010). The Valuation of Correlation-Dependent Credit Derivatives Using a Structural Model, Journal of Credit Risk 6, 99–132.
- Jarrow, R.A., F. Yu (2001). Counterparty risk and the pricing of defaultable securities. Journal of Finance 56, 1765–1799.
- Li, D.X. (2000). On Default Correlation: A Copula Function Approach. Journal of Fixed Income 9, 4354.
- McNeil, A.J., R. Frey, P. Embrechts (2005). Quantitative Risk Management, Princeton University Press.

3.5 List of resources

3.5.1 Books

Presented in alphabetical order.

- Bielecki, T.R., M. Rutkowski (2002). Credit Risk: Modelling, Valuation and Hedging, Springer Finance.
  - Dependence in portfolio structural models is briefly described toward the end of Chapter 3, while this topic is also discussed in Chapters 9 and 10.
  - Technical/Mathematical level: Very technical.
- Bluhm, C., L. Overbeck and C. Wagner (2010). Introduction to Credit Risk Modelling, Second Edition, CRC Press.
  - Correlated defaults are discussed in Chapter 2, and CreditRisk+ is covered in Chapter 4.
- Crouhy, M., D. Galai, R. Mark (2000). Risk management, McGraw Hill

  - Chapters 7 and 8 examine the building blocks of CreditMetrics and the evaluation of the credit VaR in this model.
  - Moody's-KMV, which is often used in portfolio models, is discussed in Chapter 9.
  - CreditRisk+ is presented as an actuarial model in Chapter 10.
  - Technical/Mathematical level: Very accessible
- De Servigny, A., O. Renault (2004). Measuring and Managing Credit Risk, McGraw-Hill.
  - Two complete chapters are devoted to the issue of portfolio credit risk modeling. Chapter 5 examines the sources and measures of dependence, while portfolio models are discussed in Chapter 6.
  - Technical/Mathematical level: Accessible
- Duffie, D., K.J. Singleton (2003). Credit Risk: Pricing, Measurement, and Management, Princeton Series in Finance.
  - Correlated defaults are covered in Chapter 10.
  - Technical/Mathematical level: Accessible/technical
- Hull, J.C. (2014). Options, futures and other derivatives, 9th edition, Pearson
  - Chapter 22 discusses default correlation and credit VaR, which is well explained.
  - Technical/Mathematical level: Very accessible
- Hull, J.C. (2015). Risk Management and Financial Institutions, Wiley Finance.
  - Book mostly targeted at practitioners.
  - Credit VaR is covered in Chapter 21.
- Lando, D. (2004). Credit risk modelling, Theory and Applications, Princeton University Press.
  - Entire chapter devoted to dependence modeling, from correlation and copulas to network dependence.
  - Technical/Mathematical level: Technical
- McNeil, A.J., R. Frey, P. Embrechts (2005). Quantitative Risk Management, Princeton University Press.
  - Textbook targeted at advanced undergraduates, graduates or professionals with an applied mathematics background. Covers a wide range of topics in quantitative risk management. Very well done if the reader has the technical skills.
  - Chapter 5 addresses copulas and dependence.
  - Chapters 8 and 9 discuss in detail some single-name and portfolio credit risk models.
  - Technical/Mathematical level: Technical
- O'Kane, D. (2008). Modelling Single-name and Multi-name Credit Derivatives, Wiley Finance.
  - Chapters 13–16 discuss portfolio models and how they can be used to price multi-name derivatives.

3.5.2 Websites and online reports

- CreditMetrics' (official) technical document (2007), RiskMetrics, MSCI: https://www.msci.com/resources/technical_documentation/CMTD1.pdf
- CreditRisk+ (official) document (1997), Credit Suisse | First Boston: http://www.csfb.com/institutional/research/assets/creditrisk.pdf

3.5.3 Computer programs

R Programming Language

- crp.CSFP package: According to CRAN “Modelling credit risks based on the concept of “CreditRisk+”, First Boston Financial Products, 1997 and “CreditRisk+ in the Banking Industry”, Gundlach & Lehrbass, Springer, 2003.”
- CreditMetrics package: According to CRAN “A set of functions for computing the CreditMetrics risk model”

Chapter 4

Credit risk for individuals

This chapter discusses credit risks that are sold by insurance companies. Instead of proposing how to model credit risk on the asset side, we discuss credit risk on the liability side.

In the case of short-term credit insurance products, for example insurance on credit cards, mortgages[1], surety insurance, and similar products of this type, actuaries can use the traditional approach of P&C insurance. Insurance companies can then operate under a micro-level approach, where each insured is observed over several years. The risk characteristics, the presence or absence of default risk, as well as the date and the amount paid for those defaults are recorded in the insurance company database.

In the analysis of such insurance products, P&C actuaries are on familiar ground, and the usual pricing techniques can be used. This chapter will summarize the statistical approaches used in pricing in general insurance. As opposed to the other chapters of this compendium, a more technical and theoretical approach is needed in this chapter, as the knowledge of the P& C actuaries is more advanced.

4.1 Credit Scoring

The basis for the credit risk modeling for individuals is the notion of risk scoring. Scoring is a classic segmentation technique used in finance, marketing or actuarial sciences. This technique is to assign an overall rating to an individual based on some of his personal characteristics, such as his age, his sex, his marital status or his income but can also be based on past activities or past defaults. These characteristic variables can be categorical or continuous and may also be used independently or in interaction. As we shall see, the construction of a score uses predictive modeling techniques.

The main idea of credit score modeling is to use a historical database, in which we have the characteristics of each individual, a variable indicating the number of credit defaults, and ideally, the cost involved in each of his bankruptcies. Table 4.1 shows a small sample of a database illustrating the type of observations with which a scoring analyst could work[2]:

# Observation	Sex	Age	Civil Status	…	Number of defaults	Loss given default
1	F	51	Married	…	0	.
2	M	45	Single	…	1	15,733
3	M	27	Single	…	0	.
…	.	.	…	…	…	…
155,000	F	71	Divorced	…	2	7,845

[1] Mortgage insurance has average durations of 4-7 years, with some policies extending beyond 7 years.
[2] In the U.S., many of those rating variables represent a protected class and therefore cannot be used for pricing in practice.

Table 4 1: Sample of a scoring database

The general challenge in scoring techniques is therefore to determine a way to estimate the probability of default, or the number of future defaults, and the loss given default for some specific individuals having various characteristics.

4.2 Basic Modeling and Actuarial Techniques

Table 1 shows a database for credit risk, but could also represent a database for conventional insurance products, such as automobile or property insurance. Indeed, the risk characteristics used to price an insurance product, such as gender, the age, or the marital status, are used in ratemaking to predict the realization of a random variable. In the pricing of auto insurance or home insurance, an actuary attempts to predict the number of claims and the cost of each of these claims. In credit risk analysis, the analyst will instead attempt to predict the number of defaults and amount of loss given default. In essence, the situations are almost identical. Thus, P&C actuaries already have the basic knowledge to perform modeling for this type of product.

4.2.1 Minimum Bias

Ratemaking, pricing and using segmentation variables is an old problem in actuarial science, for which many solutions and techniques have been proposed in the literature. Historically, the minimum bias technique that was introduced by Bailey and Simon (1960) and Bailey (1963) were used to find the parameters of some classification rating systems. These techniques are intended to find parameters that minimize the bias of the premium using iterative algorithms.

With the development of regression models and statistical theories, theoretical models have been used instead of the minimum bias. As we will see in the following section, advanced predictive models based on generalized linear models (often simply called GLMs) are used to estimate the parameters of a rating system. Interestingly, note that it has been shown that the results obtained from GLM theory are very similar to those obtained by the minimum bias technique (see, for example, Brown (1988)).

4.2.2 Statistical Approaches

The statistical approaches to ratemaking, as well as those that can be used for scoring risk ratings, are based on the following pure premium expression:

$$PP_i = \sum_{j=1}^{N_i} S_{i,j}.$$

By convention, we suppose that PP=0 when N=0, meaning that there is no related cost when there are no claims. This models the total amount of claims paid by the insurer; specifically, for an insured *i*, for a given level of insurance coverage (such as insurance protection against defaults) is the sum paid *S* on all N_i claims. It can be shown that the expected value of *PP* can be expressed as the product of the expected value of *N* times the expected value of *S*. In other words, the pure premium is equal to the frequency times the severity. Having defined the total potential cost to be assumed by the insurer, it is now necessary to specify how actuaries must model and segment the random variables N, S and PP.

4.2.3 Number of defaults, N

The starting point for modeling the number of events, claims or defaults, a random variable denoted *Y*, is the Poisson distribution. The Poisson distribution is the basis for almost all analysis of count

data. The Poisson distribution has some interesting properties, such as an equidispersion property, which means that the expected value of the distribution is equal to its variance. By examining a database, it is then quite simple to verify whether the Poisson distribution is appropriate by comparing the empirical mean and the empirical variance.

The characteristics of the insured that should influence the premium, such as age, sex or marital status, can be included as regressors affecting the parameter representing the count distribution. As in classic regression models, the exogenous information can be coded with binary variables. For example, we can model sex with a variable *x* that takes value 1 if the insured is a man and 0 otherwise. In statistical modeling, we use the link function $h(x_i'\beta)$, where $\beta' = (\beta_0, \beta_1, \dots, \beta_k)$ is a vector of regression parameters for the binary explanatory variables $x_i' = (x_{i,0}, x_{i,1}, \dots, x_{i,k})$. Usually in insurance, the link function *h()* is an exponential function. Consequently, we have a mean parameter $\lambda_i = t_i \, exp\,(x_i'\beta) = exp\,(x_i'\beta + \ln(t_i))$, where t_i represents the risk exposure of the insured *i*. Because the mean parameter can be seen as the premium to be charged to an insured, there is a substantial advantage from using an exponential function. Indeed, it allows the insurer to construct a premium based on multiplicative relativities:

$$\begin{aligned}\lambda_i &= t_i \, exp(\beta_0 + \beta_1 \, x_{i,1} + \cdots + \beta_1 x_{i,k}) \\ &= t_i \, exp(\beta_0) \, exp(\beta_1 x_{i,1}) \dots exp(\beta_1 x_{i,k}) \\ &= \; t_i p_0 \times r_{i,1} \times \dots \times r_{i,k},\end{aligned}$$

where p_0 can be viewed as the base premium and $r_{i,j}, j = 1, \dots, k$ as the relativities applied to insureds having the property *j* (i.e., $x_{i,j} = 1$).

Note that because the Poisson distribution is a member of the exponential family, it has some useful statistical properties (see the list of resources at the end of this chapter for further details). One such property is the fact that GLM theory can be used to estimate the parameters. Thus, instead of manually maximizing the sum of the log-probability function, one can use packages in R or SAS procedures to estimate the parameters. The properties of the MLE are well known and allow us to compute the variance of the estimators.

Because the Poisson distribution has some severe drawbacks, such as its equidispersion property, that limit its use, some generalizations are often needed. Indeed, we can still use the estimated parameters obtained by the Poisson distribution assumption and simply *correct* the overdispersion of the Poisson distribution by adding a multiplicative factor ϕ to the variance of the estimators. We can estimate the ϕ parameter using several techniques that are often included in R packages and SAS procedures, for example. The statistical significance of the estimators can be tested using classic Wald or likelihood ratio tests. Deviance can also be used to verify the fit of the Poisson because this distribution is a member of the linear exponential family.

Other properties of the Poisson distribution

The scientific literature proposes many other models that can correct problems related to the use of the Poisson distribution (see the list of resources at the end of this chapter for further details). An interesting property of the Poisson distribution refers to the time between two defaults. If we define the time between the $i-1^{th}$ and the i^{th} default as a random variable, we can define the arrival time of the j^{th} default by summing all the j^{th} times between defaults. By examining the relationship between the count process of the time between defaults, we can show the if we suppose that the waiting time between two defaults is exponentially distributed with mean $1/\lambda$, the number of defaults before time t can be expressed as a Poisson distribution of mean λt. Note that because the hazard function does not depend on *t* (the memory less property of the exponential distribution), the Poisson does not imply duration dependence, meaning that a default does not modify the expected waiting time until the next default. Another important property of the Poisson distribution that

follows the waiting time interpretation is that the mean parameter of the model is proportional to the observed time length. Normally, *t* is regarded as the number of years of coverage. For example, an insured covered for 6 months will have a premium half as high as if he were insured for 1 year because his mean parameter of the Poisson distribution will be 0.5λ, compared to λ. Some papers consider waiting times that are not exponentially distributed.

An interesting generalization of the Poisson distribution can be constructed by adding a heterogeneity term to the mean parameter of the Poisson distribution. Intuitively, we suppose that the overdispersion of the insurance data is caused by the omission of some important classification variables. Consequently, by supposing that the insurance portfolio is heterogeneous, we can generalize the Poisson distribution by adding a random heterogeneity term:

$$Pr(Y = y) = \int_0^\infty Pr[Y = y|\theta]g(\theta)d\theta,$$

where $Pr[Y = y|\theta]$ is the conditional distribution of Y, and $g(\theta)$ is the density of Θ. The introduction of a heterogeneity term means that the mean parameter is also a random variable. When the random variable Θ follows a gamma distribution of mean 1 (to ensure that the mean heterogeneity is equal to 1), the resulting mixed model is then a negative binomial distribution. This count distribution is also well known in actuarial science. It can be proved that the negative binomial distribution accounts for overdispersion. As for the Poisson distribution, we can estimate the parameters of the distribution by maximum likelihood by maximizing the log-probability function, or by using R and SAS, as they include preprogrammed packages and procedures for this distribution. Obviously, other heterogeneity distributions can be used with the Poisson distribution, such as the Inverse Gaussian distribution, the lognormal distribution, or mixtures of continuous distributions. However, in those cases, parameter inference is more difficult.

Other count distributions

The actuarial literature, as well as the statistical literature, and studies on the number of defaults propose many other types of count distributions. We refer the interested reader to the book by Cameron and Trivedi (2013) or that by Denuit et al. (2007) for an overview of all possible count distributions. It could be interesting for a credit risk modeller to fit many count distributions to his datasets and compare the fit of the model, as well as its predictive ability.

4.2.4 Loss given defaults, S

To model the loss given defaults, i.e., the amount of claims, denoted S, we restrict ourselves to strictly positive distributions. As for the count distributions, many parametric distributions can be used to model the amount, but we will restrict ourselves to a few of them. The distributions were chosen because of their importance in the actuarial literature (see the list of resource for further details).

1. The gamma distribution is usually the first distribution used to model the cost of claims in P&C insurance and should also be the first distribution to use to model the loss given default. The gamma distribution is a member of the linear exponential family, and as done with the Poisson distribution, GLM theory can be used. Consequently, regressors can be added quite easily in the mean parameter of the distribution.
2. The inverse-Gaussian distribution is an interesting alternative to the gamma distribution. As it is also a member of the linear exponential family, covariates and regressors can be added easily in the mean parameter of the distribution, and GLM theory can be used. The inverse-Gaussian distribution has a heavier tail that the gamma distribution.
3. The beta-prime distribution, also called the inverse-beta distribution or second-type beta distribution, is also very popular in actuarial science. The distribution is quite flexible, but it is not a member of the linear exponential family. Consequently, covariates and regressors

can still be used in the model, but maximum likelihood estimators are difficult to identify, as the log-likelihood function should be maximized numerically. Like the inverse-Gaussian distribution, the beta-prime distribution can be used for heavy-tail data.

4. The Pareto distribution is another interesting distribution to consider. This distribution is widely used in the literature (in actuarial sciences and statistics) to model extreme values.
5. The Champernowne distribution is less known by actuaries but possesses many advantages. For example, the tail of the Champernowne distribution converges toward a Pareto distribution.

Parametric modeling allows the actuary to summarize data into a distribution with a small number of parameters. However, the analysis of the loss given default can be performed by first exploring the data. Histograms and kernel estimators, for example, should be used to understand the data. Finally, goodness-of-fit tests should be used to verify the fit, and information criteria can be used to compare distributions. Classic statistical books can provide further details.

4.3 Typical Credit Risk Products Sold by Insurers

As indicated in the introduction of this chapter, the statistical approach developed by actuaries for traditional P&C products is perfectly suitable for credit products sold by P&C insurers. To complete the overview, we present some typical insurance products and the approaches employed by actuaries and risk modelers.

Mortgage Credit Risk

As mentioned in Mrotek and Schmitz (2010), before the financial crisis of 2008, many financial analysts based their risk evaluations of mortgage credit risk on ratings agencies and securities brokers. Since the crisis, more techniques and approaches have been proposed for actuaries to analyze such risks.

There are numerous scientific papers discussing the financial crisis of 2008. These papers often attempt to explain how such a crisis occurred and frequently discuss the techniques used to quantify and manage mortgage credit risk. Regarding actuarial science, let us mention the overview of the subprime mortgage crisis by Donnely and Embrechts (2010). The authors explain how actuaries quantified risk at that time. The paper discusses some of the actuarial models that were used in the pricing of credit derivatives. The authors explain how important it is to correctly model the dependence between risks. Indeed, since mortgages all depend on the same economic and social conditions, it is necessary to rule out the independence assumption between risks. However, as the authors note, adding dependency between risks is not a simple task, and it is important to understand how to model dependent random variables.

A data source, and significant expertise in the modeling of mortgage credit risk in the United States, is offered by the Federal Housing Finance Agency (FHFA) (Dunsky et al. (2014)). This agency provides an overview of the mortgage market in the United States, and the FHFA notes the following:

"The motivation to build the FHFA Mortgage Analytics Platform derived from the Agency's need for an independent empirical view on multiple policy initiatives. Academic empirical studies may suffer from a lack of high quality data, while empirical work from inside the industry typically represents a specific view. The FHFA maintains several vendor platforms from which an independent view is possible, yet these platforms tend to be inflexible and opaque. The unique role of the FHFA as regulator and conservator necessitated platform flexibility and transparency to carry out its responsibilities."

Surety Insurance

Jiang and Dunn (2013) offer a survey of techniques that were historically used to model credit card risks, as well as mortgage default risk. As previously mentioned, statistical approaches based on GLM or panel data modeling are used. The paper lists several databases that can be used to adjust models. Dunn and Kim (1999) also use regression techniques to explore the important covariates that can be used to explain credit card default.

Surety insurance modeling is less popular, and few papers can be found in the literature. However, a paper published by actuaries, Alwis and Steinbach (2003), explains the product:

Surety is unique in the insurance industry in that it is the only three-party insurance instrument. It is a performance obligation, meaning it is a joint undertaking between the principal and the surety to fulfill the performance of a contractual obligation.

Alwis and Steinbach (2003) also describe the historical context and propose intuitive techniques to model risk. Advanced techniques using statistical approaches should be developed by the actuarial community. However, these products do not seem to have attracted interest from the actuarial community or academics.

Credit Scoring and Credit Cards

A credit score corresponds to a numerical value to represent the credit risk associated with an individual or company. This score is used to better express the creditworthiness or the probability of default of an individual or a company. Usually, as mentioned at the beginning of this chapter, the credit score is based on individual characteristics and the past financial operations carried out by the person or company concerned. Such information is often held by credit bureaus. Banks and credit companies use the credit score to assess the risk of default posed during a financial loan and when issuing credit cards. There is vast literature on credit scoring. For example, two books, Siddiqi (2012) and Thomas (2009), explore this domain. More recently, Koh et al. (2015) propose data mining techniques to construct credit scoring models.

In actuarial science, credit scoring is directly used as a covariate in many insurance products, such as automobile or homeowner insurance. Actuaries should not have to model credit scores directly, but they should understand how such scoring works. Credit scoring is highly related to bonus-malus systems in automobile insurance. As mentioned in Denuit et al. (2007), bonus-malus systems are class systems in which the insured's level increases or decreases depending on the number of reported accidents. A specific entry level is determined for new insured individuals. Each year, the level of each insured is adjusted depending on that person's claim experience. The number of claims is usually considered in determining the level of the BMS, but the amount of the claim could also be considered.

4.4 References

- Athula Alwis, A. C. A. S., & Steinbach, C. M. (2003, March). Credit & Surety Pricing and the Effects of Financial Market Convergence. In The Casualty Actuarial Society Forum Winter 2003 Edition Including the Data Management Call Papers and Ratemaking Discussion Papers (p. 139).
- Bailey, R. A. (1963), “Insurance Rates with Minimum Bias,” Proceedings of the Casualty Actuarial Society 50, pp.4-11.
- Bailey, R. A., and L. Simon (1960), “Two Studies in Automobile Insurance Ratemaking,” ASTIN Bulletin 1, 1960, pp. 192-212

- Brown, R. L (1988)., "Minimum Bias with Generalized Linear Models," Proceedings of the Casualty Actuarial Society 75, pp. 187-217.
- Cameron, A. C., & Trivedi, P. K. (2013). Regression analysis of count data (Vol. 53). Cambridge university press.
- Denuit, M., Marechal, X., Pitrebois, S., & Walhin, J. (2007). Actuarial Modelling of Claims Count. John Wiley&Sons.
- Donnelly, C., & Embrechts, P. (2010). The devil is in the tails: actuarial mathematics and the subprime mortgage crisis. Astin Bulletin, 40(1), 1-33.
- Dunn, L. F., & Kim, T. (1999). An empirical investigation of credit card default. Ohio State University, Department of Economics Working Papers, (99-13).
- Dunsky, R.M., Zhou, X., Kane, M. Chow, M., Hu, C. and Varrieur, A., FHFA Mortgage Analytics Platform, FHFA, July 10, 2014.
- Jiang, S. S., & Dunn, L. F. (2013). New evidence on credit card borrowing and repayment patterns. Economic Inquiry, 51(1), 394-407.
- Koh, H. C., Tan, W. C., & Goh, C. P. (2015). A two-step method to construct credit scoring models with data mining techniques. International Journal of Business and Information, 1(1).
- Schmitz, M. and Mrotek, K (2010). An Analysis of the Limitations of Utilizing the Development Method for Projecting Mortgage Credit Losses and Recommended Enhancements Casualty Actuarial Society E-Forum, Fall 2010-Volume 2

4.5 List of resources

See also Chapter 12.

4.5.1 Books

Presented in alphabetical order.

- Anolli, M., Beccalli, E., & Giordani, T. (Eds.). (2013). Retail credit risk management. Palgrave Macmillan.
- Cameron, A. C., & Trivedi, P. K. (2013). Regression analysis of count data (Vol. 53). Cambridge university press.
- Charpentier, A. (Ed.). (2014). Computational Actuarial Science with R. CRC Press.
- De Jong, P., & Heller, G. Z. (2008). Generalized linear models for insurance data (Vol. 136). Cambridge: Cambridge University Press.
- Frees, E. W. (2009). Regression modeling with actuarial and financial applications. Cambridge University Press.
- Frees, E. W., Derrig, R. A., & Meyers, G. (Eds.). (2014). Predictive Modeling Applications in Actuarial Science (Vol. 1). Cambridge University Press.
- Kleiber, C., & Kotz, S. (2003). Statistical size distributions in economics and actuarial sciences (Vol. 470). John Wiley & Sons.
- Klugman, S. A., Panjer, H. H., & Willmot, G. E. (2012). Loss models: from data to decisions (Vol. 715). John Wiley & Sons.
- Molenberghs, G, & Verbeke, G. (2005). Models for Discrete Longitudinal Data. Springer Series in Statistics. Springer.
- Siddiqi, N. (2012). Credit risk scorecards: developing and implementing intelligent credit scoring (Vol. 3). John Wiley & Sons.
- Thomas, L. C. (2009). Consumer Credit Models: Pricing, Profit and Portfolios: Pricing, Profit and Portfolios. OUP Oxford.
- Wu, D. D., Olson, D. L., & Birge, J. R. (2011), Computational Risk Management.

4.5.2 Scientific Publications

Presented in alphabetical order.

- Abad, R. C., Fernndez, J. M. V., & Rivera, A. D. (2009). Modelling consumer credit risk via survival analysis. SORT: statistics and operations research transactions, 33(1), 3-30.
- Boucher, J. P., & Guillen, M. (2009). A survey on models for panel count data with applications to insurance. RACSAM-Revista de la Real Academia de Ciencias Exactas, Fisicas y Naturales. Serie A. Matematicas, 103(2), 277-294.
- Boucher, J. P., Denuit, M., & Guillen, M. (2008). Models of insurance claim counts with time dependence based on generalization of Poisson and negative binomial distributions. Variance, 2(1), 135-162.
- Boucher, J. P., Denuit, M., & Guilln, M. (2007). Risk Classification for Claim Counts: A Comparative Analysis of Various Zeroinated Mixed Poisson and Hurdle Models. North American Actuarial Journal, 11(4), 110-131.
- Dunn, L. F., & Kim, T. (1999). An empirical investigation of credit card default. Ohio State University, Department of Economics Working Papers, (99-13).
- Frees, E. W., & Valdez, E. A. (2008). Hierarchical insurance claims modeling. Journal of the American Statistical Association, 103(484), 1457-1469.
- Getter, D. E. (2006). Consumer credit risk and pricing. Journal of Consumer Affairs, 40(1), 41-63.
- Hand, D. J. (2001). Modelling consumer credit risk. IMA Journal of Management mathematics, 12(2), 139-155.
- Rules, O. A. T., & Orphanides, A. (2007). Finance and Economics Discussion Series Divisions of Research & Statistics and Monetary Affairs Federal Reserve Board. Washington, DC, January.
- Thomas, L. C. (2009). Modelling the credit risk for portfolios of consumer loans: Analogies with corporate loan models. Mathematics and Computers in Simulation, 79(8), 2525-2534.
- Wekesa, O. A., Samuel, M., & Peter, M. (2012). Modelling Credit Risk for Personal Loans Using Product-Limit Estimator. International Journal of Financial Research, 3(1), 22.

4.5.3 Database

- Canadian Statistics for Credit Cards Defaults
- Canadian Statistics for Mortgage and Credit Cards Defaults
- Statistics for Credit Cards Defaults
- Canadian consumer credit trends, Equifax Analytical Services
- Consumer Finance Monthly

4.5.4 Computer programs

- ActuR package
- Arthur Charpentier's website: http://freakonometrics.hypotheses.org/
- Credit scoring demo

Chapter 5

Single-name credit-sensitive assets

In the first part of this chapter, we discuss three credit-sensitive assets traded on a company, i.e., stocks, corporate bonds and credit default swaps, and in the second part, we address how to use these securities to evaluate a company's creditworthiness[3].

5.1 Securities

A company that needs to raise money generally has two choices: issue stocks or issue bonds. In the first case, it cedes part of the ownership in the company in exchange for money. In the second case, it acquires a loan that needs to be repaid with interest. For the lender, the loan bears credit risk tied to the uncertainty over whether the borrower will fully repay the loan.

5.1.1 Stocks

A stock is a title of ownership in a company. Its value evolves randomly over time and may or may not pay dividends. A dividend is an income stream paid for each share owned. Standard stock valuation methods rely on discounting cash flows such as dividends and earnings during some holding period. Fundamental and technical analyses are methods used by investors to assess the relative value of a stock, whether it is overpriced or underpriced. Fundamental analysis relies on financial ratios such as earnings per share and price to earnings, whereas technical analysis relies mostly on patterns in historical stock prices. For further details, see Bodie et al. (2013) and Brealey et al. (2013).

The standard stock valuation approaches rarely explicitly discuss the issue of credit risk and how stocks can reflect the financial health of a firm. In fact, changes in earnings due, for example, to an economic downturn will affect a firm's solvency and credit risk. However, the structural models described in Chapter 2 (which are inspired by Merton (1974)) often view stocks as a contingent claim or a derivative on the assets of the firm, incorporating the capital structure as a whole. The two approaches thus fundamentally differ.

5.1.2 Corporate bonds

A corporate bond is essentially a loan contracted between the borrower and thousands of investors on the financial markets and is part of a larger family known as fixed-income securities. The price of a bond is expressed on a common basis, which is known as the par or face value. This value is often set at 100 by convention. Corporate bonds entail credit risk in the sense that cash flows are not guaranteed, i.e., the borrower may or may not fully repay the bond's principal and interest. When one or more payments are missed, we say that the borrower is in default.

The price of a corporate bond is also often expressed in terms of yields or spreads. The yield, or yield-to-maturity, is the unique interest rate that reprices a corporate bond. In other words, given the future cash flow structure and assuming that the bonds are paid with certainty, the yield-to-maturity is the unique interest rate used to discount cash flows such that it matches the current

[3] This is in contrast with statistical tools that are based on default frequencies, ratings transitions or financial ratios.

observed price. The yield is unique to a specific corporate bond, as it depends not only on the general level of interest rates or the credit risk of the issuer but also on the cash flow structure. Conversely, credit spreads are firm-specific measures that do not necessarily depend on the bond's characteristics.

The relationship between the duration of a loan and its annualized interest rate is known as the term structure of interest rates. Such a term structure can be obtained on default-free securities (such as those issued by a highly rated country) or on (defaultable) corporate bonds. The credit-spread curve is the difference between the term structures of interest rates on a defaultable bond and the equivalent on a default-free bond. Thus, the credit spread is the compensation required by bondholders to invest in a risky bond for a given maturity. In theory, credit risk should be the main driver (along with interest rates) of corporate bond prices (and spreads).

However, this is not entirely the case empirically. Various authors have investigated the question of what proportion of corporate bond spreads is truly attributable to credit risk. The first authors examining this issue are Elton et al. (2001), who find, quite surprisingly, that the proportion of corporate spreads related to default is at most 25%. Using a different methodology and dataset, Dionne et al. (2010) find that this proportion can vary between 30% and 75%, depending on the credit rating and period. Other notable contributions in this area are Collin-Dufresne et al. (2001), Eom et al. (2004) and Huang & Huang (2012), who obtain similar results while considering different datasets, periods and methodologies. The consensus is that (1) credit risk alone cannot fully explain corporate bond spreads, (2) liquidity can also be an important factor, especially for long-term bonds that are held by insurance companies and rarely traded, and (3) the proportion of credit spreads explained by credit risk decreases as a firm's credit quality increases.

A difficulty that may arise with these studies is the presence of embedded options (also known as optionalities or covenants) in corporate bonds. These options are included in corporate bonds to give bondholders further protections, making the bonds more attractive to the market. However, these covenants also affect the price of the bond and the corporate spread. Most bonds have numerous embedded options such as callable (redeemable), putable, exchangeable, extendible, and so forth.

For a more thorough treatment of corporate bonds and other fixed-income securities, the reader should consult Fabozzi & Mann (2012).

5.1.3 Credit default swaps

A credit default swap (CDS) is a credit derivative that offers protection against the default (or any credit event) of the underlying company in exchange for a premium. A company may take on the role of the protection seller (which receives a premium but assumes default risk) or of the protection buyer (which pays a premium in exchange for protection). CDSs are generally available with maturities of 1 to 10 years, with 5 being the most traded maturity and 1 being the least traded.

Credit default swaps are quoted with a reference bond that is employed in calculating the amount of loss paid on default. When the CDS is settled in cash, a sum of money equivalent to the loss on the reference bond is paid on default. When physical settlement is chosen, on default, the insurance buyer cedes the defaulted bond to the seller, and the latter also pays the notional amount on the bond.

A CDS is essentially an insurance contract against default and is used chiefly by investors who face default risk on their assets. CDSs are therefore used mostly as a hedging instrument. However, following the financial crisis of 2008, CDSs have been in the spotlight due to lack of CDS market regulation. A first reason is that the protection buyer is not required to have an equivalent insurable interest in the company for which it seeks protection, thus allowing for major speculation. For example, an investor holding $100M in bonds in company ABC is allowed to buy CDS protection of more than $100M in notional value. Second, to avoid its own bankruptcy (also known as counterparty risk), the issuer of a CDS (protection seller) needs to set aside sufficient funds in case it has to pay for a default.

Although much less common since the 2008 financial crisis, CDSs can also be issued on structured financial products (such as asset-backed securities). In this case, default is interpreted as the inability of the investment to deliver the totality of the promised payments. American Insurance Group (AIG) suffered very significant losses in 2008 from CDSs issued on collateralized debt obligations on securitized subprime mortgages.

The valuation of CDSs is similar in spirit to the valuation of *T*-year term insurance with premiums payable 4 times a year until the end of the contract or on default, whichever comes first. The main difference, which is essential, is that pricing is subject to the absence of arbitrage arguments. A CDS combined with a corporate bond should yield the risk-free rate, as the combination is supposed to be risk free. Market makers will exploit arbitrage opportunities, and thus, expectations should be taken under the pricing (risk-neutral) probability measure. The quarterly premium is determined based on the equivalency principle, meaning that the expected benefits should be equivalent to the expected premiums. In theory, the spread or premium paid by the protection seller should be determined by the firm's default risk and interest rates, given that payments can be as far as 10 years in the future.

However, in practice, when we study the determinants of CDS premiums, it is found that default risk and interest rates are indeed major components of the spreads (see Ericsson et al. (2009)). However, other components—liquidity and counterparty risk (see Arora et al. (2012))—are also significant, especially since the financial crisis, although their overall impact on the spreads can be small.

5.2 Credit risk assessment using security prices

To assess the credit risk of a firm using a credit risk model, one needs to estimate the model parameters. An increasingly popular approach is to use security prices such as stocks, corporate bonds and/or CDSs[4]. The idea is that when the parameters are known, most credit risk models can be used to calculate the price of securities. Thus, the opposite can also hold, that is, to take observed security prices as given and use them to estimate a model.

An intuitive way to do so is to minimize the squared difference between the theoretical (model) and observed prices over time and/or a cross-section of securities. This method is better known as calibration, which is popular among practitioners and often performed daily on multiple securities at once, yielding a different set of parameters every day. The quality of the fit is assessed on the basis of the pricing errors. Despite being very intuitive, it is difficult, albeit not impossible, to assess the uncertainty associated with a given parameter.

The scientific community has turned instead to statistically based estimation approaches such as maximum likelihood or the generalized method of moments (GMM), which are based on the frequentist paradigm (as opposed to the Bayesian paradigm). In the maximum likelihood approach, one seeks to identify the parameters that maximize the probability of observing a given sample. The GMM, as its name indicates, generalizes the method of moments by using more moments than the number of parameters or by placing different weights on the moments (such as assigning greater weight to tail measures).

The estimation of credit risk models using maximum likelihood with security prices was pioneered by Duan (1994, 2000). The idea is that the main driver of a company's creditworthiness, that is, the market value of assets, the default intensity, etc., is not directly observed but indirectly observed through the value of a security (such as a stock, corporate bond or CDS). Since the latter is a transformation of the former, the likelihood function has to account for this link between the two, yielding a slightly modified likelihood function. One recovers all the usual properties of the maximum likelihood by following Duan (1994).

The disadvantage of this approach is that it assumes that all changes in the price of the security

[4] Other estimation approaches will be discussed throughout the compendium. Portfolio models are also further discussed in Chapter 3.

are attributable to changes in creditworthiness, which is not necessarily the case. For example, Duan & Fulop (2009) report that asset volatility in Merton's model can be overstated if model error is not considered. Therefore, we need to disentangle the changes in the security price that are due to changes in the fundamentals (the true credit risk of the firm) and those due to other factors. This is the role of filtering techniques such as non-linear Kalman filters or particle filters. For a review of these methods, see Boudreault et al. (2015).

More recently, academics and practitioners have focused on models that include a random recovery rate. The estimation of such models presents a challenge because of the identification issue. The price of a credit-sensitive security is affected by two variables: the likelihood of default and the loss given default. Depending on the model and the data used, it is possible to vary the two variables in opposite directions to obtain the same price. This identification issue can be solved by two methods: (1) modify how recovery payments occur in the model (see Pan & Singleton (2008)) and/or (2) use various types of securities (or datasets) that are exposed differently to recovery rate risk in the estimation process.

5.3 References

- Arora, N., P. Gandhi, F.A. Longstaff (2012). Counterparty credit risk and the credit default swap market. Journal of Financial Economics 103, 280–293.
- Bodie, Z., A. Kane, A.J. Marcus (2013). Investments, 10th edition, McGraw Hill.
- Boudreault, M., G. Gauthier, T. Thomassin (2015). Estimation of correlations in portfolio credit risk models based on noisy security prices. Journal of Economic Dynamics and Control 61, 334–349.
- Brealey, R.A., S.C. Myers, F. Allen (2013). Principles of Corporate Finance, McGraw-Hill
- Collin-Dufresne, P., R.S. Goldstein, J.S. Martin (2001). The determinants of credit spread changes. Journal of Finance, 2177–2207.
- Dionne, G., G. Gauthier, K. Hammami, M. Maurice, J.G. Simonato (2010). Default risk in corporate yield spreads. Financial Management 39, 707–731.
- Duan, J.C. (1994). Maximum likelihood estimation using price data of the derivative contract. Mathematical Finance 4, 155–167.
- Duan, J.C., A. Fulop (2009). Estimating the structural credit risk model when equity prices are contaminated by trading noises. Journal of Econometrics 150, 288–296.
- Elton, E. J., M.J. Gruber, D. Agrawal, C. Mann (2001). Explaining the Rate Spread on Corporate Bonds. Journal of Finance 56, 247277.
- Eom, Y. H., J. Helwege, J.Z. Huang (2004). Structural models of corporate bond pricing: An empirical analysis. Review of Financial Studies 17, 499–544.
- Ericsson, J., K. Jacobs, R. Oviedo (2009). The determinants of credit default swap premia. Journal of Financial and Quantitative Analysis 44, 109–132.
- Fabozzi, F.J., S.V. Mann (2012). The Handbook of Fixed Income Securities. McGraw Hill Professional.
- Huang, J.Z., M. Huang (2012). How Much of the Corporate-Treasury Yield Spread Is Due to Credit Risk?. Review of Asset Pricing Studies 2, 153–202.
- Merton, R.C. (1974). On the pricing of corporate debt: The risk structure of interest rates. Journal of Finance 29, 449–470.
- Pan, J., K.J. Singleton (2008). Default and recovery implicit in the term structure ofsovereign CDS spreads. Journal of Finance 63, 2345–2384.

5.4 List of resources

See also Chapter 12.

5.4.1 Books

Presented in alphabetical order.

- Ammann, M. (2002). Credit Risk Valuation: Methods, Models, and Applications, Springer
  - Chapter 6 discusses the pricing of single-name credit derivatives.
- Bluhm, C., L. Overbeck, C. Wagner (2010). Introduction to Credit Risk Modelling, Second Edition, CRC Press.
  - Single-name and multi-name credit derivatives (excluding CDOs) are introduced in Chapter 7.
- Crouhy, M., D. Galai, R. Mark (2000). Risk management, McGraw Hill.
  - Credit (single-name and portfolio) derivatives are discussed in the context of hedging credit risk in Chapter 12.
  - Technical/Mathematical level: Very accessible
- De Servigny, A., O. Renault (2004). Measuring and Managing Credit Risk, McGraw-Hill.
  - There is a brief discussion on credit default swaps and other single-name credit derivatives in Chapter 9.
  - Technical/Mathematical level: Accessible
- Duffie, D., K.J. Singleton (2003). Credit Risk: Pricing, Measurement, and Management, Princeton Series in Finance.
  - The pricing of corporate bonds and credit default swaps is covered in Chapters 6 and 8, respectively.
  - Technical/Mathematical level: Accessible/technical
- Fabozzi, F.J., S. Mann (2012). The Handbook of Fixed Income Securities, 8th edition, McGraw-Hill.
  - Chapter 66 primarily examines at the characteristics of CDSs, whereas Chapter 67 discusses their valuation.
  - Technical/Mathematical level: Very accessible
- Hull, J.C. (2014). Options, futures and other derivatives, 9th edition, Pearson.
  - Chapters 22 and 23 discuss single-name and portfolio credit derivatives. The method proposed to estimate Merton's model is outdated and provides biased estimates of the parameters[5]. Very good introduction to CDSs.
  - Technical/Mathematical level: Very accessible
- Lando, D. (2004). Credit Risk Modeling, Theory and Applications, Princeton University Press.
  - CDSs are discussed in Chapter 8.
  - Technical/Mathematical level: Technical
- O'Kane, D. (2008). Modelling Single-name and Multi-name Credit Derivatives, Wiley Finance.
  - Book specifically focusing on the pricing of credit derivatives. The Wiley Finance series is usually accessible to practitioners.
  - Chapters 4 and 5 describe corporate bonds and CDSs in detail. Chapter 6 addresses pricing CDSs, while risk management of CDSs is discussed in Chapter 8.
  - Technical/Mathematical level: Accessible

5.4.2 Computer programs

R programming language

- creditr package: According to CRAN, it "Provides tools for pricing credit default swaps

[5] See Ericsson, J. and J. Reneby (2005), "Estimating Structural Bond Pricing Models," Journal of Business 78, 707-735.

using C code for the International Swaps and Derivatives Association (ISDA) CDS Standard Model"

- credrule package: According to CRAN, "It provides functions to bootstrap Credit Curves from market quotes (Credit Default Swap - CDS - spreads) and price Credit Default Swaps - CDS."

Matlab programming language

- CDS Pricer

Chapter 6

Municipal securities

6.1 Introduction

Municipal securities are bonds issued by government entities, such as a city, province/state or public utility. These products are better known as short-term municipal securities, or munis. The fact that munis are generally tax-free makes them well suited to investors taxed in the highest brackets. Government entities issue munis to finance expansion projects (airports, waterworks, parks, schools, hospitals, etc.) or simply to fund daily operations (infrastructure maintenance, landscaping, etc.). The main buyers of municipal securities are individuals, mutual funds, commercial banks and P&C insurance companies.

6.2 Type and characteristics

Municipal bonds come in many forms. They include zero-coupon bonds and bonds with fixed or variable rates. Munis also fall into two categories: general obligation bonds and revenue bonds. General obligation bonds are bonds secured by the taxation power of the issuer (often considered infinite). By comparison, revenue bonds repay bondholders directly from revenues generated by the infrastructure being funded (airport, schools, hospitals, power stations, maritime ports, etc.).

6.3 Ratings by recognized credit rating agencies

Until the 1980s and 1990s, investors considered ratings issued by rating agencies (Moody's, Standard & Poor's, etc.) to be infallible. In 1994, the case of Orange County created upheaval when the county was forced to seek protection under the bankruptcy act; its munis had previously been very highly rated. This case also led investors to ask more questions about fund management by municipalities and quashed the popular belief that muni fund managers invest prudently.

More recently, on July 18, 2013, the city of Detroit declared bankruptcy, and investors subsequently lost a total of $7 billion. In mid-2015, the island of Puerto Rico defaulted on a payment of nearly $58 million.

In the wake of all these improbable yet actual events, financial institutions have become more suspicious of the credit ratings of munis. They now use them as a starting point and continue their analysis using internal models. Note that rating agencies use the same rating system for munis as for corporate bonds. In addition, agencies use more than one rating style to represent several types of municipal bonds (see Moody's, Standard & Poor's).

6.4 Tax issues

Although munis are generally considered nontaxable, it is important to understand that this status applies only at the federal level and not necessarily at the state, provincial and municipal levels.

Evaluating munis is complex; options and many other factors need to be considered. In addition, the rate of return (nontaxable) of munis must be compared with that of corporate bonds

(taxable) to determine how profitable they are. Because the muni rate is below the rate of return of equity, it is important to ensure that the investment is worthwhile (given the investor's taxation rate). Therefore,

$$\text{Taxable bond rate} \approx \frac{Muni\ rate}{1-T}$$

where T is the investor's marginal taxation rate. Note that this is a very rudimentary way to evaluate the tax impact on the total bond yield.

6.5 Insurance on municipal bonds

One method to reduce a muni's borrowing rate is by using insurance. This insurance guarantees that the investor will be paid the coupon and principal in full if the issuer defaults on the payment. There are two groups of municipal bond insurers: monoline and multiline. Monoline insurers provide only financial guarantees, whereas multiline insurers are P&C insurance companies. Although this insurance may appeal to issuers of lower-quality bonds, the financial crisis of 2008 sent insurance companies into turmoil. As a result, the addition of this insurance does not necessarily improve the quality of the bond due to the counterparty risk of the insurance company.

6.6 Factors determining credit risk

Before investing in munis, investors must analyze the clauses of the official document to determine their risk exposure. According to the SEC's Office of Investor Education and Advocacy, there are at least four factors to consider before investing in munis (SEC (2012)):

1. Type of municipal bond: as mentioned above, there are two main types of munis and, therefore, two different credit risks. For general obligation bonds, even if the taxation power is considered unlimited, it is important to verify the real taxation power of the entity issuing the bond. For revenue bonds, revenue may be more at risk if the future project is risky, and thus, this type of muni has a higher probability of default than general obligation bonds do.
2. Non-recourse financing: Some revenue bonds have a non-recourse financing clause, which means that the investor obtains nothing if the product does not generate revenue. This greatly increases the credit risk.
3. Financial condition of the issuer: The credit rating is a good starting point to anticipate the outlook for a municipal bond issuer. Evidently, the worse the rating is, the higher the default risk. As mentioned above, secondary analysis of the issuer is crucial to anticipate possible future issues.
4. Other sources to pay the principal and interest on the bond: Sometimes, the funds to repay the municipal bond come from an uncertain source, which makes the issuer's capacity to derive constant revenue from that source uncertain. One example is funds coming from taxes on the sale of a product in declining demand due to a social or economic cause, such as tobacco.

To begin investigating a municipal bond, one can consult the official release on the Electronic Municipal Market Access (EMMA) website at www.emma.msrb.org.

6.7 Credit risk model

As we have seen, and as Fama (1977) and Miller (1977) confirmed, municipal bonds are not risk-free, which is why their yields are higher than typical government bonds adjusted for tax considerations. An adjustment is made for credit risk. In practice, explaining the part of the spread attributed to credit risk is called the muni puzzle: this calculation is not easy because the taxation

rate is not fixed for each individual. Few models of credit risk calculation exist. A recent approach proposes the use of a neural network to classify munis (Hjek 2011). Starting with the rating given to munis by an agency and through the use of several qualitative and quantitative variables, the algorithm attempts to find links between variables and performs a classification. With this type of model, we can see whether the chosen variables are representative of the evolution of credit risk and then predict the future rating of a muni. Chalmers (1998) demonstrates that default risk does not suffice to solve the muni puzzle. This means that another source of risk must be incorporated into the calculation.

6.8 Liquidity risk

One of the risks that should not be overlooked in evaluating municipal bonds is liquidity risk. According to Wan et al. (2005), the "Liquidity premium explains about 7 to 13 percent of the observed municipal yields for AAA bonds, 7 to 16 percent for AA/A bonds and 8 to 20 percent for BBB bonds with different maturities." Based on muni price data, Wang et al. (2005) use a model to find the percentage of return attributable to liquidity, default, taxes and the risk-free rate of a muni. Ang et al. (2014) offer a more recent study on this topic.

6.9 List of resources

6.9.1 Books

Presented in alphabetical order.

- Fabozzi, F.J. (2012). The Handbook of Fixed Income Securities, McGraw-Hill
  - Chapters 11 and 44 cover several aspects of munis.
  - This is one the most important references used to write this chapter.
  - Technical/Mathematical level: Very Accessible
- Fabozzi, F.J., S.G. Feldstein (2008). The Handbook Of Municipal Bonds, Wiley
  - Qualitative book addressing several aspects of municipal bonds.
  - Technical/Mathematical level: Very Accessible
- Mysak, J., M.R. Bloomberg (2010). Handbook for Muni-Bond Issuers, Wiley
  - Although this book is not intended for investors, we found it useful to include a reference on muni issuers because this subject is not very well known.
  - Technical/Mathematical level: Very Accessible
- SIFMA (2011). The Fundamentals of Municipal Bonds, 6th Edition, Wiley Finance
  - SIFMA is a group of private companies (banks, hedge funds, etc.) interested in munis.
  - Chapters 6 and 7 cover risk and credit analysis of munis, respectively.
  - Technical/Mathematical level: Very Accessible
- Handbook of Finance (2008). Financial Markets and Instruments, Wiley.
  - This three-volume work addresses a number of diverse topics in finance.
  - Chapter 22 covers municipal bonds.
  - Technical/Mathematical level: Very Accessible

6.9.2 Data

- The Municipal Securities Rule Making Board (MSR) is a company that protects municipal bond investors. Although the MSRB has compiled many databases on munis, they are not free to use.
- J.J. Kenny Drake, Inc.: "J.J. Kenny Drake, Inc. operates as a municipal bonds inter-dealer. It offers market data that provides regional and national coverage with approximately 500 screen pages of market activity; offerings and bid-wanted; coverage by region, state, and

bond type; and descriptions on various advertised issues through Web and digital feeds." (Bloomberg Finance website).

6.9.3 Bibliography

- Ang, A., V. Bhansali, Y. Xing (2014). The Muni Bond Spread: Credit, Liquidity, and Tax, Columbia Business School Research Paper No. 14–37.
- Chalmers, J.M.R. (1998). Default risk cannot explain the muni puzzle: Evidence from municipal bonds that are secured by US Treasury obligations, Review of Financial Studies 11, 281–308.
- Fama, E.F. (1977). A Pricing Model for the Municipal Bond Market, University of Chicago, Mimeo.
- Hjek, P. (2011). Municipal credit rating modelling by neural networks, Decision Support Systems, Vol. 51(1), pp. 108–118.
- Miller, M. (1977). Debt and Taxes, Journal of Finance, 32, 261–275.
- Moody's Standing Committee (2015). Moody's Rating Symbols and Definitions.
- PRMIA (2001). Orange County, Sungard.
- SEC (2012). Municipal Bonds: Understanding Credit Risk, Office of Investor Education and Advocacy, Access. Investor Bulletin.
- Standard & Poor's (2012). Standard & Poor's Ratings Definitions.
- Wang, J., C. Wu, F. Zhang (2005). Liquidity, Default, Taxes and Yields on Municipal Bonds, Finance and Economics Discussion Series Divisions of Research & Statistics and Monetary Affairs Federal Reserve Board.

Chapter 7

Portfolio credit risk derivatives and other structured assets

Multi-name credit-sensitive products are, in general, investments with values that depend on the credit risk of an entire portfolio of corporations or individuals. In this chapter, we will summarize some of the most popular multi-name credit-sensitive products, namely the basket CDS, collateralized debt obligations and other asset-backed securities.

7.1 Basket Credit Default Swaps

A basket credit default swap (CDS) is similar to a standard single-name CDS, but the credit event is triggered with the *k*-th default in a portfolio of reference assets. For example, the first-to-default CDS on a 100-name portfolio implies that a payment is expected whenever one reference in the portfolio defaults.

The behavior of the basket CDS depends on the dependence relationship in the portfolio. A first-to-default CDS will be more costly when assets in the portfolio are independent, whereas a last-to-default CDS will be the cheapest under independence.

Pricing, in the financial engineering sense (absence of arbitrage), requires a portfolio credit risk model, as described in Chapter 3. Because the distribution of the 1st, 2nd or *k*-th to default is rarely tractable in large portfolios, prices are determined by simulation.

7.2 Asset-Backed Securities

Asset-backed securities (ABS) are financial assets with values that are derived from the cash flows of a portfolio of obligations. Most of these instruments are created using what is known as securitization, which is the process of pooling non-tradable obligations, such as mortgages or credit cards, into a tradable security. An intermediary handles the legal aspect of the creation of the ABS by collecting the premium from the investors on one side and the cash flows from the obligations on the other. The obligations most commonly used in ABS are mortgages, auto loans, credit cards and student loans. We will discuss mortgage-backed securities (MBSs) and collateralized debt Obligations (CDOs) in further detail. MBSs are covered very extensively in Fabozzi & Mann (2012).

7.2.1 Mortgage-Backed Securities

We define as MBSs the subset of ABS based on a pool of residential or commercial mortgages. MBSs issued by either Fannie Mae (FNMA), Freddie Mac (FHLMC) or Ginnie Mae (GNMA) are known as agency MBSs, whereas non-agency MBSs are mostly issued by private investment banks. As of late 2014 and according to the Securities Industry and Financial Markets Association (SIFMA), US agencies have approximately $6 trillion in outstanding MBSs, whereas non-agencies have approximately $1.5 trillion in outstanding MBSs.

The most significant types of MBSs are pass-through securities, collateralized mortgage

obligations (CMOs) and stripped MBSs. A pass-through security backed by mortgages simply repays a share of the cash flows of a pool of mortgages (residential or commercial). CMOs represent all structured securities (backed by mortgages) with cash flow designs that are meant to distribute prepayment risk differently among investors (see below). They are legally established by a special purpose entity. Finally, stripped MBSs are securities that redistribute part of the interest or principal of a pool of mortgages.

MBSs are affected by changes in interest rates, the borrower's prepayment behavior and credit risk. Prepayment means that a borrower repays its mortgage faster than initially planned. This occurs primarily when the mortgage is refinanced at a lower rate or whenever the home is sold. Prepayment is a risk to the investor since it changes the timing of cash flows and, in the case of refinancing, lowers the rate at which the cash flows are reinvested.

CMOs offer investors various ways to cope with prepayment risk. Principal can be paid sequentially to investors (sequential pay) or based on a planned amortization schedule with bounded repayment rates. For example, in the first design, an investor entitled to the second 25% of the principal will start receiving payments when the first 25% has been paid. Therefore, when prepayment is faster (slower) than expected, the investor will also receive its share of the cash flows faster (slower). The second design is known as planned amortization class (PAC). Whenever prepayments are within some given range, the investor assumes prepayment risk, and thus, the cash flows will be very stable. Otherwise, an excess or shortage of prepayments will be compensated by companion tranches. CMOs are popular assets for insurance companies to match the duration of their liabilities. More CMO structures are discussed in Chapters 26–29 of Fabozzi & Mann (2012).

The borrower's credit risk is tied to the borrower's ability to fully repay its mortgage. In many instances, it is considered negligible because lenders investigate the credit history of their borrowers, agencies provide guarantees against homeowner default, and the homes serve as collateral. However, the subprime mortgage crisis of 2008 proved that many financial institutions relied too heavily on the home equity to provide mortgages to insolvent individuals. When the subprime bubble burst, many borrowers defaulted, the value of the houses plunged, and Fannie Mae and Freddie Mac were placed in conservatorship by the US Treasury.

7.2.2 Collateralized Debt Obligations

A CDO is a type of ABS that was highly popular in the 2000s. It was used to securitize and repackage loans and tranches of other MBSs. CDOs have been blamed for fuelling the subprime mortgage crisis in 2008-2009. At its 2007 peak, the CDO market size was as large as $1.4 trillion but collapsed following the recession. As of the end of 2014, the market size of the CDO market was approximately $800B (total CDOs outstanding, SIFMA).

A CDO is established through a special purpose vehicle, and its cash flows are distributed in tranches using a waterfall structure, i.e., a set of priority rules. To illustrate how tranches are structured in a CDO, let us assume a simple two-tranche CDO, i.e., senior and junior tranches. According to the waterfall structure, the senior-tranche investors are first in line to receive the cash flows from the collateralized debts. Once the senior-tranche investors have been paid in full, the junior-tranche investors receive the surplus until they are also paid in full. The latter structure provides senior-tranche investors with better credit risk protection than junior-tranche investors, and therefore, the senior tranche trades at a lower premium.

The value of CDOs depends on the creditworthiness of their internal obligations, but it is highly dependent on the degree of diversification within the portfolio. From a statistical standpoint, the strength of the dependence between the constituents is a major driver of CDO value, as illustrated in Table 7.1.

In Table 7.1, a "1" implies a default, whereas "0" is survival. It is important to note that the sum over columns leads to the same default probabilities in both scenarios of high or low dependence. Thus, there are 9 defaults in each part of the table.

High dependence generates clusters of defaults, i.e., scenarios in which there are either 0 or 4 defaults, whereas under low dependence, in all scenarios, there are at most 1 or 2 defaults. Therefore, under low dependence, it is highly unlikely that one would observe multiple defaults.

The impact on a CDO is direct. Under low dependence, senior tranche holders are highly protected whereas junior tranche holders will most likely assume all defaults. While not numerous, these defaults are regular. Under high dependence, all tranche holders are paid in full under certain scenarios, but under others, there are multiple defaults that also affect senior tranche holders. Therefore, dependence affects the variability of outcomes for all tranche holders in non-trivial manner. Tranches were often designed to obtain a very safe rating such as AAA, whereas as shown in the

High dependence						
Scenario/Company	1	2	3	4	5	Total
ω_1	0	0	0	0	1	1
ω_2	1	1	1	0	1	4
ω_3	0	0	0	0	0	0
ω_4	0	1|	1	1	1	4
ω_5	0	0	0	0	0	0
Default prob.	0.2	0.4	0.4	0.2	0.6	9
Low dependence						
Scenario/Company	1	2	3	4	5	Total
ω_1	1	0	0	1	0	2
ω_2	0	1	0	0	1	2
ω_3	0	0	1	0	0	1
ω_4	0	1	0	0	1	2
ω_5	0	0	1	0	1	2
Default prob.	0.2	0.4	0.4	0.2	0.6	9

Table 7.1: Scenarios of defaults in a portfolio under low or high dependence

previous illustration, in a high-dependence economy, even the safest tranches are affected.

In the CDO-related enthusiasm that preceded the financial crisis, CDOs squared (CDO^2) and CDOs cubed were also issued. A CDO squared is a CDO issued on a CDO, i.e., tranches of a CDO are further broken down into tranches to form the cash flows of the CDO^2. Similarly, a CDO cubed is a CDO issued on a CDO squared. In both cases, valuation was highly complex due to the tranching mechanism applied over other tranches.

Pricing, again in the financial engineering sense, requires a portfolio credit risk model, as described in Chapter 3. The Gaussian copula was very popular in pricing CDOs, in large part due to David X. Li's paper in 2000. Li provided the first tractable approach to pricing CDOs and therefore linked hundreds or thousands of loans using correlations. In the aftermath of the financial crisis, the Gaussian copula and its misuse in CDO pricing was highly criticized.

7.3 References

- Fabozzi, F.J., S.V. Mann (2012). The Handbook of Fixed Income Securities. McGraw Hill

Professional.
- Li, D.X. (2000). On Default Correlation: A Copula Function Approach. Journal of Fixed Income 9, 4354.

7.4 List of resources

See also Chapter 12.

7.4.1 Books

Presented in alphabetical order.

- Bielecki, T.R., M. Rutkowski (2002). Credit Risk: Modelling, Valuation and Hedging, Springer Finance.
  - Chapter 9 examines the mathematical mechanics behind basket credit derivatives.
  - Technical/Mathematical level: Very technical.
- Bluhm, C., L. Overbeck, C. Wagner (2010). Introduction to Credit Risk Modeling, Second Edition, CRC Press.
  - CDOs are thoroughly covered in Chapter 8.
- Crouhy, M., D. Galai, R. Mark (2000). Risk management, McGraw Hill.
  - Credit (single-name and portfolio) derivatives are discussed in the context of hedging credit risk in Chapter 12.
  - Technical/Mathematical level: Very accessible
- De Servigny, A., O. Renault (2004). Measuring and Managing Credit Risk, McGraw-Hill.
  - CDOs are introduced in Chapter 9.
  - Technical/Mathematical level: Accessible
- Duffie, D., K.J. Singleton (2003). Credit Risk: Pricing, Measurement, and Management, Princeton Series in Finance.
  - CDOs are covered in Chapter 11.
  - Technical/Mathematical level: Accessible/technical
- Fabozzi, F.J., S. Mann (2012). The Handbook of Fixed Income Securities, 8th edition, McGraw-Hill.
  - The coverage of the various MBSs is very thorough, with 10 chapters dedicated to several specific topics (Chapters 24–32, 41).
  - Moreover, ABSs on credit cards, auto and student loans are discussed in Chapters 33 and 34.
  - Technical/Mathematical level: Very accessible
- Hull, J.C. (2014). Options, futures and other derivatives, 9th edition, Pearson.
  - Chapter 23 briefly discusses basket credit derivatives such as basket CDSs, CDOs and their valuation.
  - Technical/Mathematical level: Very accessible
- O'Kane, D. (2008). Modelling Single-name and Multi-name Credit Derivatives, Wiley Finance.
  - Chapters 12 and 22 address portfolio credit derivatives, while risk management of these products is discussed in Chapter 17.
  - Technical/Mathematical level: Accessible

7.4.2 Computer programs

Matlab programming language
- Fast Computation of the Expected Tranche Loss of CDO Credit Portfolio
- Li's Copula model for CDS and CDO default intensities and loss function

Chapter 8

Counterparty risk

8.1 Introduction

According to the Bank for International Settlements, counterparty risk is (BIS (2005))
[...] the risk that the counterparty to a transaction could default before the final settlement of the transaction's cash flows. An economic loss would occur if the transactions or portfolio of transactions with the counterparty has a positive economic value at the time of default [...]

Counterparty risk is almost inevitable for an entity that wants to take a position in the financial market. As the definition states, counterparty risk is the risk that the entity holding the opposite position from that of the investor fails to meet its obligations. It can result from the default of the issuer of an option, the inability of the protection seller of a CDS to deliver the promised compensation, or the failure of the other side of the swap to make payments, and so forth.

For an insurance company, counterparty risk is not limited to the derivatives that it acquires to manage its risks. Reinsurers may also file for bankruptcy, which represents a major counterparty risk for insurers. For further details, see section 8.4.

To avoid default risk and reinforce the credibility of financial markets, governments and financial institutions have introduced counterparty policies. As one example, clearinghouses have been introduced; acting as intermediaries, they ensure that a financial security's cash flow streams are available when one of the two parties involved in a transaction demands settlement. Although clearinghouses are intermediaries that reduce the risk of payment default, the possibility that the clearinghouses themselves default is quite present; this risk must therefore also be considered.

Because all financial transactions involve a random future value, the credit risk of the counterparty must be assessed by means of calculation techniques stipulated by Basel, Solvency, etc. to determine the value of exposure to this risk.

8.2 Credit Risk

In a simple loan between a bank and an individual, counterparty risk is assumed by the bank. The bank loses money if the borrower does not repay the loan. It is important to understand that credit risk is quite different for a derivative. Hull (2012) (Chapter 23) notes three possible situations:

- "The contract is always a liability to the financial institution
- The contract is always an asset to the financial institution
- The contract can become either an asset or a liability to the financial institution."

In the first case, the financial institution bears no counterparty risk because the bank owes money to another entity (such as when a bank sells options). In the second case, the financial institution assumes counterparty risk because the product it trades has positive value (such as when a bank buys options). The last case refers to a situation in which the financial institution would have positive or negative counterparty risk depending on the product's characteristics (such as forwards).

It is therefore important for a financial institution to incorporate a premium in all the financial products that it sells, to cover this risk. By denoting the value of the financial product at

time t_i by f_i, the (risk-neutral) probability of default at time t_i by q_i and the recovery rate by *R*, we can define the following equation:

$$\sum_{i=1}^{T} q_i(1-R)PV(E[max(f_i,0)])$$

where PV(.) denotes the present value.

We therefore obtain the value of the CVA (credit value adjustment), an adjustment made to all financial products to account for counterparty risk. This calculation is very rudimentary but represents the broad outlines very well. Several factors come into play in CVA valuation, including the addition of some clauses to the contracts that financial institutions issue. Three typical clauses help mitigate counterparty risk (see, for example, Hull (2012) (Chapter 23), Caouette et al. (1998)):

1. **Netting:** Netting is a clause that stipulates that if entity A has several contracts with entity B and if one of the two entities defaults on one of its contracts, the entity in question will have to close all its positions with the other party. Let us assume, for example, that A has two contracts with B. The first contract has a value of -$10 and the second a value of $15 for A. If A defaults on the first contract, it will receive $15 from the second contract if there is no netting clause. In the opposite case, A will receive only $5 because the value of the two products will be combined.
2. **Collateralization:** This clause refers to what we mentioned earlier in Section 8.1. The value of a security changes over time, and to ensure that funds are indeed available at maturity when the security becomes payable, the entity that owes money must set a particular amount aside. This amount is generally administered by clearinghouses. This clause is very similar to the margins that clearinghouses require on futures.
3. **Downgrade Triggers:** This clause stipulates that if the credit rating of one of the parties is downgraded, then the contract can be closed immediately at its market value to avoid a future default.

Counterparty risk can be modeled using classic credit risk models such as those presented in a previous chapter. Because the counterparty may default at any time, a first-passage structural model is a good starting point to model counterparty risk (see Chapter 2). Several other models can also be used to model this risk (see Brigo et al. (2013) for a more exhaustive list). According to Brigo et al. (2013), counterparty risk in a derivative product can be divided into three subcategories:

- Credit exposure (CE): the value of the product today, if the counterparty defaults (market value)
- Expected exposure (EE): the value of expected future losses
- Potential future exposure (PFE): the possible future risk with a given confidence interval; this risk can be assessed using VaR (value-at-risk)

8.3 Credit Default Swap (CDS)

Counterparty risk can be mitigated with CDSs (see also Chapter 5), which are similar to a conventional swap, where a periodic premium is paid by the protection buyer in exchange for a payment by the protection seller in the event that the reference entity defaults. The use of these products has expanded considerably since 2008 following the bankruptcy of Lehman Brothers.

Note that even if these products largely cover credit risk, CDSs are also exposed to the counterparty risk between the buyer and issuer of this product. The counterparty risk linked to CDSs has been increasingly studied since AIG sustained major losses with its CDSs, valued at over $ 182.5 billion (Longstaff et al. (2010)). It is therefore important to consider this risk when pricing CDSs. Classic finance models can serve as the basis for CDS valuation.

8.4 Reinsurer credit risk

Counterparty risk may exist between an insurer and its reinsurer. The role of the reinsurer is to cover the losses of an insurance company that exceed a pre-defined limit. We present two different viewpoints on the link between the insurer and reinsurer (Britt & Krvavych (2009) and Cummins & Weiss (2010)).

Britt & Krvavych (2009) argue that several factors increase counterparty risk between insurers and reinsurers:

- Given that the number of reinsurers on the market is small, counterparty risk has become very concentrated.
- The correlation between the insurer and reinsurer is very high because they both provide insurance (they trade the same type of financial products). In addition, the strong link between the insurer, reinsurer and catastrophic events raises the counterparty risk. When a disaster occurs, the insureds of several insurance companies will file insurance claims, and their insurers will in turn file a claim with the reinsurer. This increases the reinsurer's probability of default.

Therefore, it is difficult to evaluate the mitigation of counterparty risk between an insurer and its reinsurer. A Monte Carlo simulation should be used to assess the impact of this risk.

It is important to mention that counterparty risk is a precursor to credit risk contagion. In our insurer/reinsurer relationship, default contagion exists if a reinsurer is unable to cover the losses of several insurers following a natural disaster, for example. This point is raised by Cummins & Weiss (2010) who analyze insurance data to analyze their correlations. They determined that a reinsurer's insolvency crisis would trigger an intra-sector crash. However, their study found only a weak correlation between insurers and the financial market, which implies that a crisis affecting insurers would not necessarily lead to a financial crisis. The authors conclude that insurers can nonetheless create systemic risk on financial markets when they use options (such as CDSs) in their non-core activities (as in the case of AIG, see Section 8.3).

8.5 References

- Brigo, D., M. Morini, A. Pallavicini (2013). Counterparty Credit Risk, Collateral and Funding – With Pricing Cases for All Asset Classes, Wiley Finance.
- Britt, S., Y. Kravych (2009). Reinsurance Credit Risk Modelling – DFA Approach, in ASTIN Colloquium 2009.
- Caouette, J.B., E.I. Altman, P. Narayanan (1998). "Managing Credit Risk: The Next Great Financial Challenge," Wiley.
- Cummins, J.D., M.A. Weiss (2010). Systemic Risk and the US Insurance Sector, Journal of Risk and Insurance 81, 489–528.
- Hull, J.C. (2014). Options, futures and other derivatives, 9th edition, Pearson.
- Longstaff, F.A., P. Gandhi, N. Arora (2010). Counterparty Credit Risk and the Credit Default Swap Market, Journal of Financial Economics 103, 280–293.

8.6 List of resources

8.6.1 Books

Presented in alphabetical order.

- Brigo, D., M. Morini, A. Pallavicini (2013). Counterparty Credit Risk, Collateral and Funding – With Pricing Cases for All Asset Classes, Wiley Finance.
  - Covers a wide range of financial models
  - Technical/Mathematical level: Technical

- Caouette, J.B., E.I. Altman, P. Narayanan (1998). "Managing Credit Risk: The Next Great Financial Challenge," Wiley.
    - Chapter 5 discusses clearinghouses at greater length;
- Cesari, G., J. Aquilina, N. Charpillon, Z. Filipovic, G. Lee, I. Manda (2009). Modelling, Pricing, and Hedging Counterparty Credit Exposure, Springer.
    - Concentrates on modeling counterparty risk for plain vanilla and exotic derivatives.
    - Technical/Mathematical level: Technical
- Gregory, J. (2011). Counterparty Credit Risk: The new challenge for global financial markets, Wiley finance.
    - Discusses several topics in counterparty risk, from qualitative and quantitative perspectives.
    - Technical/Mathematical level: Technical
- Gregory, J. (2015). The xVA Challenge: Counterparty Credit Risk, Funding, Collateral, and Capital, 3rd Edition, Wiley.
    - Extensively discusses counterparty risk
    - Presents many financial products
    - Technical/Mathematical level: Accessible
- Hull, J.C. (2014). Options, futures and other derivatives, 9th edition, Pearson.
    - Chapter 24 discusses counterparty risk as part of credit risk.
    - Technical/Mathematical level: Very accessible
- Norman, P. (2011). The Risk Controllers: Central Counterparty Clearing in Globalised Financial Markets, Wiley.
    - Discusses clearinghouses.
    - Refers to many historical facts.
    - Technical/Mathematical level: Accessible

Chapter 9

Sovereign credit risk

9.1 Introduction

After the restriction on cross-border cash flows was lifted in the late 1960s, Argentina and Mexico nearly defaulted on their loans due to oil shocks. Since these events, sovereign credit risk has been an important element to consider when one country decides to lend money to another. Sovereign credit risk is the risk that a (sovereign) country that borrowed money internationally will not repay its creditors (interest and/or principal). Whereas corporate bond default must be covered with the company's assets, international loans carry a higher risk because if the issuer defaults, the borrowing country is not necessarily obliged to repay its debt.

Although it may seem advantageous for an issuing country to default on its debt, this choice has serious economic repercussions. Take, for example, a country (which we will call A) that exports wheat. Country A may experience years of abundance and shortages. During the lean years, A has no choice but to borrow internationally to meet the needs of its population. Conversely, during good years, the country has a surplus and can repay its debt. The fact that A repays this debt cements good relations with its trading partner nations. If A refused to repay the loan, relations between the countries would deteriorate, and the underwriting countries would stop lending money to A. Over the long term, this may create enormous problems for A if it were to experience a shortage lasting several consecutive years.

9.2 Sovereign credit risk factors

Frenkel et al. (2004) identified five factors to explain sovereign credit risk:

1. Debt service ratio: a country's ratio of debt service payments (principal + interest) to export earnings. A country with a high ratio is considered risky.
2. Import ratio: A country's ratio of total imports to its total foreign exchange reserves. Foreign exchange reserves are assets that central banks hold to cover their debt. The higher this ratio is, the higher the country's risk.
3. Investment ratio: A set of accounting ratios related to a country's financial health. These ratios must be examined carefully to determine their impact on credit risk.
4. Variability of export revenue: Export revenue varies with the quantity and price of goods and services. Risk therefore comes from the law of supply and demand for this good/service. The more variable the revenue is, the higher the credit risk.
5. Domestic money supply growth: Domestic money supply is the total monetary value of the assets of an economy. If growth is good, the credit risk is lower because the country's economy is improving.

In addition to the factors identified by Frenkel et al. (2004), we can add power of taxation. The power of taxation encompasses income tax, goods/services sold by the government, sales tax, and more. A country that supplies a very expensive service to its population for free always has the option of privatizing this service to release itself from these financial obligations. In addition, a government may increase the income tax to reduce its international debt.

9.3 Modeling and pricing

One of the principal references for sovereign credit risk modeling is Eaton & Gersovitz (1981). They introduced a model for the amount of debt observed and determined that this amount is the minimum between the amount of debt sought by the borrowing country and the maximum quantity of loans permitted by the other countries. Their model allows for the possibility of default, but this must be to the borrower's advantage; moreover, after a single default, the country may no longer borrow abroad. Research on payment incentives has also been conducted by Grossman et al. (1988)), Bulow & Rogoff (1989a, b), Atkeson (1991), Dooley & Svenson (1994), Cole & Kehoe (1996, 2000), Dooley (2000), and many others.

Edwards (1984) offers another viewpoint on sovereign credit risk. He sought to identify the factors that influence the difference (spread) between a country's borrowing interest rate and the London Interbank Borrowing Rate (LIBOR). Because this difference should naturally be tied to default risk, default probability factors were determined empirically. Edwards found that the spread is positively correlated with the debt/GDP ratio and debt service. In addition, the spread is negatively correlated with the international reserves/GDP ratio and the propensity to invest. Empirical studies on the determinants of the lending rate spread over LIBOR have been performed by Edwards (1986, 2002), Berg & Sachs (1988), Boehmer & Megginson (1990), Duffie et al. (2003), and Zhang (2003).

Pan & Singleton (2007) demonstrate that the spread of the borrowing rate over LIBOR may be caused by common global factors. In line with these findings, Longstaff et al. (2007) perform a principal component analysis (PCA) on several CDS spreads and find that 50% of those spreads are "[…] more related to the US stock and high-yield bond markets, global risk premia, and capital flows than they are to their own local economic measures." More recently, Longstaff & Ang (2013) study the sensitivity to systemic risk of American and European issuers and find it to be quite heterogeneous.

9.4 Default history or sovereign insolvencies

9.4.1 Russia (1998)

The Russian financial crisis occurred on August 17, 1998, triggering the devaluation of the Russian ruble and the government's default on its obligations. This crisis may be explained by a decline in productivity, recurrent budget deficits and the high exchange rate between the ruble and other currencies. Productivity plunged following the Asian financial crisis in early 1997 and the slump in demand for crude oil and nonferrous metals. In addition, Russia's insistence on keeping the exchange rate with the US dollar relatively stable led the Russian central bank to spend over US $27 billion to maintain this exchange rate.

Note that due to default on Russian bonds, Long-Term Capital Management (LTCM) sustained a loss of over 44% of its $125 billion in assets in a single month. The case of LTCM demonstrated that the use of VaR (value-at-risk) has its limits and that stress testing is crucial to identify and hedge against risks with catastrophic repercussions.

9.4.2 European crisis

The European crisis is a series of events that occurred in late 2009, the effects of which are still being felt. First, it is worth mentioning that this crisis resulted from several phenomena. The globalization of finance, easy access to credit and the bursting of the real estate bubble are just a few of the reasons for this crisis. The catalyst, however, was the financial crisis of 2008. This crisis seriously undermined government bonds because investor confidence in the stability of issuing countries plummeted. This fear accentuated the context of an economy with massive government deficits. Note that the acronym PIGS (Portugal, Ireland, Greece, Spain) was widely used to

designate the countries hit hardest by the crisis. Here are some important events (by country) related to the crisis.

Portugal

From the 1970s to the financial crisis of 2008, several senior authorities in the public sector received huge sums of cash as bonuses and other compensation. In addition, the government hired more public servants than necessary. The crisis of 2008, followed by a downgrading by Moody's in 2010, led Portugal to seek a €78 billion bailout from the IMF (International Monetary Fund). After major changes within the government and the public sector, the country emerged from the rescue package in May 2014 but still bears a heavy financial burden.

Ireland

The real estate bubble of 2007 is the primary cause of the debt crisis in Ireland. The government guaranteed the six largest banks involved in the bubble. The banks had lost nearly €100 billion, and the country's rating dropped sharply. With bond rates continuing to increase, the Irish government had no choice but to request a bailout package from the EU and the IMF of nearly €70 billion in late 2010. Fortunately for Ireland, its financial woes ended in December 2013, and the country put the rescue measures behind it.

Greece

Greece is the country that suffered the most from this crisis. In the early 2000s, Greece already had a very large deficit. With its main sources of revenue being exports and tourism, the financial crisis of 2008 decimated the country's revenue. The country requested several loans and bailout plans between 2010 and 2015 from the European Commission, European Central Bank and IMF that resulted in downgrades of the country's rating and the deterioration of the Euro and the Greek stock market. Greece went on to announce several austerity measures as a condition for the bailout plans. The austerity measures triggered a recession in 2010-2011. In March 2012, the Greek government defaulted on its obligations. There were even rumors that Greece would be ousted from the euro zone shortly thereafter (an event known as Grexit). After several negotiations and rescue measures, Greece finally rebounded in 2014 (three consecutive quarters of positive economic growth). But in early 2015, the Greek economy deteriorated again, forcing the country to default on a payment to the IMF. Budgetary measures are still in place, and the story continues.

Spain

Despite having the smallest debt of the PIGS countries, Spain's credit rating plunged when the government had to spend massively to save banks following the real estate bubble. In 2012, Spain agreed to be rescued by the ESM (European Stability Mechanism) in exchange for an unlimited bond purchase plan. After several tax measures, Spain emerged from the rescue plan in January 2014.

9.5 List of resources

9.5.1 Books

Presented in alphabetical order.

- Andritzky, J. (2012). Sovereign Default Risk Valuation, Springer.
  - Discusses a model for the evaluation of sovereign obligations.

  - Compares bonds and CDSs during the crisis.
  - Technical/Mathematical level: Very Accessible
- Duffie, D., K.J. Singleton (2003). Credit Risk: Pricing, Measurement, and Management, Princeton University Press.
  - Chapters 6 and 7 discuss a method to determine the price of a sovereign bond and its spread.
  - Technical/Mathematical level: Technical
- Frenkel, M., A. Karmann, B. Scholtens (2004). Sovereign Risk and Financial Crises, Springer.
  - Identifies five factors that explain sovereign credit risk.
  - Qualitative book on sovereign credit risk during financial crises.
  - Technical/Mathematical level: Technical/Accessible
- Gaillard, N. (2012). A Century of Sovereign Ratings, Springer.
  - Discusses rating of sovereign debt.
  - Technical/Mathematical level: Very Accessible
- Kolb, R.W. (2011). Sovereign Debt: From Safety to Default, Wiley.
  - Discusses the global impact of sovereign debt.
  - Technical/Mathematical level: Accessible
- Pepino, S. (2015). Sovereign Risk and Financial Crisis: The International Political Economy of the Eurozone, Palgrave Macmillan.
  - Book on the European crisis.

9.5.2 Website and online report

- IMF website: http://www.imf.org/
- ESM website: http://www.esm.europa.eu/

9.5.3 Database

- IMF Public Debt Database.
- Thomson Reuters Datastream Professional is a "powerful tool that integrates economic research and strategy with cross-asset analysis to seamlessly bring together top down and bottom up in one single, integrated application."[6]
- Bloomberg: Pricing data for sovereign CDSs.
- GFI "is a globally recognized information hub for credit derivative products."[7]

9.6 Bibliography

- Atkeson, A. (1991). International Lending with Moral Hazard and Risk of Repudiation, Econometrica 59, 1069–1089.
- Berg, A., J. Sachs (1988). The Debt Crisis: Structural Explanations of Country Performance, Journal of Development Economics 29, 271–306.
- Boehmer, E., W.L. Megginson (1990). Determinants of Secondary Market Prices for Developing Country Syndicated Loans, Journal of Finance 45, 1517–1540.
- Bulow, J., K. Rogoff (1989a). A Constant Recontracting Model of Sovereign Debt, Journal of Political Economy 97, 155–178.
- Bulow, J., K. Rogoff (1989b). Sovereign Debt: Is to Forgive to Forget?, American Economic Review 79, 43–50.

[6] Commercial flyer from Thomson Reuters.

[7] http://www.gfigroup.co.uk/market-data/fixed-income/credit-derivatives.aspx

- Cole, H.L., T.J. Kehoe (1996). A Self-Fulfilling Model of Mexico's 1994–95 Debt Crisis, Journal of International Economics 41, 309–330.
- Cole, H.L., T.J. Kehoe (2000). Self-Fulfilling Debt Crises, Review of Economic Studies 67, 91–116.
- Dooley, M.P. (2000). A Model of Crisis in Emerging Markets, The Economic Journal 110, 256–272.
- Dooley, M.P., L.E.O. Svensson (1994). Policy Inconsistency and External Debt Service, Journal of International Money and Finance 13, 364–374.
- Duffie, D., L.H. Pedersen, K.J. Singleton (2003). Modeling Sovereign Yield Spreads: A Case Study of Russian Debt, Journal of Finance 58, 119–159.
- Eaton, J., M. Gersovitz (1981). Debt with Potential Repudiation: Theoretical and Empirical Analysis, The Review of Economic Studies, Vol. 48, No. 2, pp. 289–309.
- Edwards, S. (1984). LDC Foreign Borrowing and Default Risk: An Empirical Investigation, 1976–80, American Economic Review. 74, 726–734
- Edwards, S. (1986). The Pricing of Bonds and Bank Loans in International Markets: An Empirical Analysis of Developing Countries' Foreign Borrowing, European Economic Review 30, 565–589.
- Edwards, S. (2002). The Argentine Debt Crisis of 2001–2002: A Chronology and Some Key Policy Issues, Working paper, UCLA.
- Frenkel, M., A. Karmann, B. Scholtens, eds. (2004). Sovereign Risk and Financial Crises, Springer.
- Grossman, H.I., J.B. Van Huyck (1988). Sovereign Debt as a Contingent Claim: Excusable Default, Repudiation, and Reputation, American Economic Review 78, 1088–1097.
- Lane, P. (2012). The European Sovereign Debt Crisis, The Journal of Economic Perspectives, Vol. 26, No. 3, pp. 49–67.
- Longstaff, F.A., J. Pan, L.H. Pedersen, K.J. Singleton (2007). How Sovereign Is Sovereign Credit Risk?, American Economic Journal: Macroeconomics, Vol.3(2), pp.75–103.
- Longstaff, F.A., A. Ang (2013). Systemic sovereign credit risk: Lessons from the US and Europe, Journal of Monetary Economics, vol. 60(5), pages 493–510.
- Pan, J., K.J. Singleton (2007). Default and Recovery Implicit in the Term Structure of Sovereign CDS Spreads, Review of Financial Studies, Vol. 63(5), pp. 2345–2384.
- Zhang, F.X. (2003). What did the Credit Market Expect of Argentina Default? Evidence from Default Swap Data, AFA 2004 San Diego Meetings, Federal Reserve Board.

Chapter 10

Regulatory environment

This section presents the most significant agreements, laws and financial regulations in the world. Section 10.1 covers agreements and laws for banks (from Basel accords and the Federal Reserve), whereas Section 10.2 covers guidelines for insurance companies, both American (NAIC) and European (Solvency).

10.1 Banks

10.1.1 Basel I, II, III

The Basel I Accord was initially drafted in 1988 in Basel, Switzerland, by the Basel Committee on Banking Supervision (BCBS). The accord was intended to provide a basis regarding the capital that banks must hold to hedge against financial and operational risks. Basel I mainly concerned credit risk and risk weighting of assets (RWA). Overall, the accord requires banks with an international presence to hold capital of at least 8% of the sum of their RWA (also known as the Cook ratio).

The simplicity of this ratio means that it can be implemented in countries with different legislation and regulations. The two main objectives of Basel I were consequently achieved: maintain the stability of the banking system and apply internationally. The RWA is calculated by assuming the weight relative to the following risks:

Weight	Category
0%	Liquid cash, gold ingots, national bonds
20%	Securitized assets
50%	Municipal bonds, home mortgages
100%	Corporate bonds
Not measured	Other

It is important to note that the required capital must come from tier 1 or 2. In addition, at least 50% of the capital must be tier 1 capital. The tiers are composed as follows (PRMIA):

1. Tier 1 capital: funds, equity and retained earnings.
2. Tier 2 capital: subordinated debt (long term), and other eligible hybrid instruments and reserves (such as loan loss provisions).

The simplicity of this formula is also its main weakness because the regulation is not directly aligned with institutions' risk characteristics. Thus, the Basel II Accord was created in 2004 to correct the shortcomings of the first accord. Basel II is divided into three pillars: 1) minimum capital requirements, 2) supervisory review and 3) market discipline. The ideas behind Basel I were revisited, and the definition of RWA was changed to better represent the risk sensitivity of the institution. In addition, instead of a single risk measurement method, three more complex methods were introduced. Regarding pillar 2, a series of guiding principles were adopted concerning the

revision of the necessary capital. Stress testing and other risk evaluation methods were introduced. Finally, pillar 3 encourages adequate disclosure of capital levels and risk exposure.

Regrettably, the implementation of Basel II was not completed before the financial crisis of 2008. Consequently, stricter laws have since been adopted, and Basel III was ultimately created in 2014 with implementation scheduled for 2019. The general idea behind Basel III is to complete the guidelines of Basel I and II by emphasizing the risk of bank runs. Basel III also includes an important chapter on the obligation to adjust capital to credit quality (credit value adjustment – CVA) because the BCBS observed that during the financial crisis of 2008, most of the losses stemmed from credit downgrades and not necessarily defaults.

CVAs are losses tied to defaults and the downgrading of counterparties for the entire portfolio duration. Note that this is not a reserve but rather a profit and loss (P&L) measure. To understand the impact of CVA on P&L, we must not consider each product individually but rather the set of products traded with the same counterparty. A Monte Carlo method can be used to evaluate CVA, but this method is not optimal given the sometimes very low default probability. The classical method, shown below, may be more effective in some cases:

$$CVA_t = (1 - R) \sum_{k=0}^{\infty} EPE(k)PD(k)$$

where *R* is the recovery rate, *PD* the default probability and *EPE* the expected positive exposure at default. Note that companies may reduce their CVA risk by using a netting agreement (legal agreement recognized by the ISDA) or by collateral arrangements.

10.1.2 Federal Reserve

In the early 2000s, the Federal Reserve introduced a set of measures intended to reduce the insolvency risk of banks and other financial institutions. The main objective of the latter measures is to assess whether the company has sufficient funds to survive an important financial stress or adverse economic conditions. The first measure, performed annually, is known as the Comprehensive Capital Analysis and Review (CCAR), whereas the second measure is the Dodd-Frank Act Stress Test (DFAST).

10.2 Insurance companies

10.2.1 NAIC – RBC

The NAIC (National Association of Insurance Commissioners) is an American organization that facilitates discussions on standard-setting and regulatory support. The organization was founded in 1871 and is governed by all 50 state insurance regulators. The NAIC establishes standards and best practices, which are then adopted, or not, by each state (according to their official website). Since the 1990s, one of the key roles of the NAIC has been to determine the required capital that an insurance company must hold to avoid default risk (risk-based capital – RBC). This is done by determining the quantity of risk that an insurance company may assume. This measure takes the size and risk style of the insurance company into account.

Different models of RBC have therefore been created for the following types of insurance: life, P&C (property & casualty) and health. The possible risk categories and the RBC formulas for each type of company are summarized by the American Academy of Actuaries' Joint Risk Based Capital Task Force (1999) (see references). Note that in Canada, RBC is determined by the Office of the Superintendent of Financial Institutions (OSFI).

The NAIC also defines the SAP (statutory accounting principles), a presentation guide for financial statements that insurance companies must follow. The SAP ensures the homogeneity of

insurance companies' financial statements, thereby simplifying the work and increasing the effectiveness of the state insurance department in auditing solvency and company reserves.

10.2.2 Solvency I and II

In 1973, the European commission decided to harmonize the risk management practices of all European insurance companies. The goal of Solvency I was to link the amount of capital required for insurance companies directly to their respective risk. Solvency II went into effect in January 1, 2016, and fills the gaps of Solvency I. It comprises three pillars (European Commission – Solvency II, 2015):

- "Pillar 1 consists of the quantitative requirements (e.g. amount of capital an insurer must hold)
- Pillar 2 sets out requirements for the governance and risk management of the insurer, and for effective supervision of the insurer
- Pillar 3 focuses on disclosure and transparency requirements."

Solvency II's Pillar 1 comprises Solvency and Minimum Capital Requirements (SCR and MCR) with formulas similar to that of RBC. However, Solvency II does not have the same risk separation as RBC. According to the CAS Risk-Based Capital (RBC) Research Working Parties (2012), for a typical P&C insurance company, Solvency II's classification corresponds to the following:

1. "Underwriting Risk: Premium (loss ratio) risk, excluding catastrophe risk, Reserve (loss development) risk, excluding catastrophe risk, Catastrophe risk
2. Default (Counterparty) Risk: Non-diversified counterparties, most significantly reinsurance counterparties, Diversified counterparties, most significantly agents balances and other receivables
3. Market Risk: Interest rate risk, Equity risk, Real estate (property) risk, Spread risk, Currency risk, Concentration risk, Illiquidity risk
4. Operational Risk."

10.4 List of resources

10.4.1 Books

Presented in alphabetical order.

- Barfield, R.E. (2011). A practitioner's guide to Basel III and beyond, Sweet & Maxwell.
  - Complete guide to Basel III written by a team of professionals from Price Waterhouse Coopers.
  - Covers the qualitative and calculation aspects of Basel III.
  - Technical/Mathematical level: Technical/Very Accessible
- Buckham, D., J. Wahl (2010). Stuart Rose, Executives Guide to Solvency II, Wiley.
  - Detailed examination of Solvency II, from qualitative and quantitative perspectives.
  - Technical/Mathematical level: Technical/Very Accessible
- Crouhy, M., D. Galai, R. Mark (2014). The Essentials of Risk Management, Second Edition, McGraw-Hill.
  - Chapter 3 covers various aspects of financial regulation
  - Technical/Mathematical level: Very Accessible
- Cruz, M. (2009). The Solvency II Handbook, Risk Books.
  - This book discusses methods to implement internal models, pillar 1 requirements, etc. It also provides several practical examples of the three pillars.
  - A companion to this book was published in 2014: The Solvency II Handbook: Practical

Approaches to Implementation.
  - Technical/Mathematical level: Very Accessible
- Doff, R. (2014). The Solvency II Handbook: Practical Approaches to Implementation, Risk Books
  - This book is mainly intended for professionals but can also be used by consultants and students.
  - It presents a multitude of practical problems that insurance professionals will face throughout their careers.
  - This book is a companion book to the Solvency II Handbook.
  - Technical/Mathematical level: Very Accessible
- Engelmann, B., R. Rauhmeier (2011). The Basel II Risk Parameters: Estimation, Validation, Stress Testing – with Applications to Loan Risk Management (2011), Springer.
  - Covers calculation methods specific to Basel II that banks must use to comply with the standards.
  - Addresses three fundamental elements of Basel: probability of default (PD), loss given default (LGD), and exposure at default (EAD).
  - Technical/Mathematical level: Technical
- Hull, J.C. (2015). Risk Management and Financial Institutions, Wiley Finance
  - Chapters 15–17 cover Basel I, II, III and Solvency II.
  - Technical/Mathematical level: Very accessible.
- NAIC (2006). NAIC Risk-Based Capital Forecasting & Instructions, NAIC
  - Comes in four volumes: Fraternal, Health, Life and Property & Casualty.
  - Note that a CD-ROM is included for the purpose of projecting RBC.
- PRMIA (2004). PRMIA handbook volume III.
  - Good reference for credit risk in general. The other volumes are also very accessible and provide good basic knowledge. Note that a more recent version is available.
  - Chapter B contains information on credit risk.
  - Section 6 of chapter III.B addresses Basel.
  - Technical/Mathematical level: Very Accessible
- Webb, B.L., C.C. Lilly (1994). Raising the Safety Net: Risk-Based Capital for Life Insurance Companies, NAIC, Free book.
  - While not very recent, this free book is recommended by the NAIC and provides a good introduction to the fundamentals of RBC.
  - Many numerical examples are provided.
  - Technical/Mathematical level: Very Accessible

10.4.2 Website and online reports

- Basel accords:
  - Basel II: http://www.bis.org/publ/bcbsca.htm (main document available in English, French, German, Italian and Spanish)
  - Basel III: http://www.bis.org/bcbs/basel3.htm
- NAIC RBC:
  - Overview and links to important resources: http://www.naic.org/cipr_topics/topic_risk_based_capital.htm
- Solvency I and II: Links from the European commission
  - Solvency II Overview – Frequently asked questions
  - Solvency II FAQ
  - All information required for insurance and pension (including technical standards and guidelines)

10.5 Bibliography

- American Academy of Actuaries (1999). Joint Risk Based Capital Task Force, Comparison of the NAIC Life, P&C and Health RBC Formulas – Summary of Differences.
- Casualty Actuarial Society (2012). Solvency II Standard Formula and NAIC Risk-Based Capital (RBC), Report 3 of the CAS Risk-Based Capital (RBC) Research Working Parties Issued by the RBC Dependencies and Calibration Working Party (DCWP), E-Forum Fall 2012.
- PRMIA handbook volume II, PRMIA (2004).
- Sokol, A. (2012). A Practical Guide to Fair Value and Regulatory CVA, Numerix/CompatibL, PRMIA Global Risk Conference.

Chapter 11

Enterprise risk management

Enterprise risk management (ERM) is one of the most important concepts in this compendium. It is covered in the last chapter mostly because it builds on content discussed in previous chapters.

The traditional view of risk management has been to consider each sector of activity as independent (in a silo). ERM responds to companies' desire to integrate all risks into a single risk approach that considers the possible correlation among activity sectors. It creates a link between credit, market, liquidity, operational, information technology and other risks. Without a centralized risk management system, the chief risk officer may not be able to evaluate all of a firm's risks effectively, owing to the use of different valuation techniques in each sector.

The Casualty Actuarial Society (CAS) Committee on ERM defines ERM as (see CAS (2003)) "[...] the discipline by which an organization in any industry assesses, controls, exploits, finances, and monitors risks from all sources for the purpose of increasing the organization's short- and long-term value to its stakeholders." The authors also underline several aspects of this definition:

- Discipline: a process fully supported by the organization to an extent that it becomes part of the company's culture;
- Any industry: not just P&C insurers;
- Increasing value: ERM focuses on both risk mitigation and value creation;
- Stakeholders: includes shareholders, debtholders, management, employees and customers.

CAS (2003) is an excellent source of information for P&C actuaries considering ERM, and some important elements of this document are summarized in this chapter.

11.1 Reasons leading to ERM

What sums up ERM is the view that risk management should aggregate all risks faced by a company, rather than each risk being administered independently. It requires a strong commitment from the firm, and thus, risk management is addressed at a much higher level within the company's structure. According to CAS (2003), there are at least six reasons that enterprises have shifted their views on risk management:

1. Complexifying risks that have arisen from globalization, technology, financial innovation, sophistication of laws, etc.
2. External pressures: the media highlighting corporate mismanagement, along with ratings agencies, regulators, etc., forcing firms to mitigate these risks considerably to limit repercussions on their image, solvency, etc.
3. Portfolio view: companies are increasingly aware of links between departments and want to take this into account (as explained in the previous section) and benefit from these links.
4. Quantification: the growing desire to quantify everything is accentuating the implementation of ERM; companies are increasingly using value-at-risk (VaR) as a risk measure.
5. Information sharing and benchmarking: the flow of information is easier and faster as it now allows a large proportion of risk managers and other stakeholders to read, interpret and compare the financial performance of companies.

6. Alternative view of risk: recent trend in business to perceive risk as an opportunity rather than as a dangerous element. When it is controlled effectively and managed by the implementation of an ERM approach, risk is in fact beneficial for a company and can contribute to projects that generate higher actual net present value (NPV).

Although the implementation of an ERM approach is very beneficial for a company, it requires a significant investment of time and money. In addition, this process may take years to deploy and demands constant management and oversight. Generally, the position of CRO (Chief Risk Officer) must be created when implementing an ERM approach. This person specializes in risk management and reports directly to the CEO (Chief Executive Officer). Finally, ERM allows for better risk analysis during the implementation of new projects and, therefore, allows firms to choose the most profitable endeavors, namely those with the highest RAROC (risk-adjusted return on capital).

11.2 Components of an effective ERM framework

According to Lam (2003), there are seven components of an effective ERM implementation:

- **Corporate governance:** Top management and the Board must define their risk appetite to enable the CRO to work effectively. In addition, they must ensure that the people assigned to each risk are knowledgeable in this area and that the firm's structure is well aligned with the objective of implementing ERM.
- **Line management:** Following up on each line of business is important because it represents the firm's revenue. An exhaustive description of plausible losses must be compiled with the help of each line's managers.
- **Portfolio management:** Firms must diversify their risks with active portfolio management.
- **Risk transfer:** Firms must identify risks that should be transferred to an insurance company. Extreme risks can alter the firm's financial health or trigger bankruptcy. Some risks may be mitigated through the use of derivatives.
- **Risk analytics:** The use of mathematics and statistics to improve future risk management. This can be achieved internally (CRO) or externally with specialized consultants. Complex methods must add value.
- **Data technology and resources:** Centralization of each sector's IT system into a single one. Keeping up to date will surely improve the efficiency of the company in general.
- **Stakeholder management:** Transparency toward stakeholders is crucial to convey good firm management. This lowers credit risk because it eases rating agencies' uncertainty about the business.

11.3 Types of risks and mitigation

When implementing ERM, the CRO must identify and apply a rigorous approach to mitigate risks. According to CAS (2003), there are four major categories of risks:

- **Hazard risks**: fire, theft, vandalism, natural catastrophes, disability, illness, etc.
- **Financial risks**: typical market risks such as asset value, interest rates, foreign exchange rate, credit risk, liquidity risk, inflation, etc.
- **Operational risks**: all risks linked to the failure of a company process such as production, IT, reporting, etc.
- **Strategic risks**: all risks tied to competitors, laws, technological advances, etc.

It is important to note that the ERM Task Force of the Actuarial Standards Board has developed specific Standards of Practice tied to risk evaluation and treatment (see ASOPs 46 and 47) in the context of ERM.

11.4 Modeling

One of the CRO's tasks is to model risks, as discussed in the previous section. There is a large spectrum of tools that can be used to assess these risks and Appendix B of CAS (2003) provides a detailed account of the methods. They can be classified as methods solely based on data, expert opinion, or a combination of both. The categories are summarized as follows:

- **Purely analytical**: In this technique, only data are considered, and the approach is appropriate if there is a large history of data. Examples of tools are direct empirical analysis, estimation of probability density functions or time series (including stochastic differential equations), regressions, extreme value theory, etc.
- **Purely expert opinion**: This technique is often used when the data required are not available or too limited to be deemed reliable (e.g., the creation of a new product). Simple probabilities may be calculated by asking for expert opinions and by creating a distribution of possible results based on their responses. A greater number of scenarios and richer results can be obtained by adding constraints to questions asked to experts.
- **Mix of analytical and expert opinion**: The mix of the two techniques is useful when the quantity of data available is not too large.

11.5 Risk management process

Although the modeling techniques used to represent these risks can be very different, CAS (2003) proposes broad outlines representing the risk management process. The seven steps are as follows:

1. **Establish Context**: Determine the company's internal and external environments.
2. **Identify Risks**: Identify and document the threats that can impact the company.
3. **Analyze/Quantify Risks**: Use the methods listed in the previous section to analyze the financial impact of these risks along with probabilities.
4. **Integrate Risks**: Determine the link between these risks and aggregate the risks to determine the real impact on the company.
5. **Assess/Prioritize Risks**: Prioritize all the risks to identify those with the greatest loss potential, and deal with those first.
6. **Treat/Exploit Risks**: Take the necessary actions to mitigate the potentially most dangerous risks.
7. **Monitor & Review**: This step is the most important. Although the risks were mitigated, it is important to constantly control them and to return to step 1 if the actions taken are not effective or if other risks arise at the company level.

11.6 References

- Casualty Actuarial Society (CAS) (2003). Overview of Enterprise Risk Management, Enterprise Risk Management Committee.
- Lam, J. (2003), Enterprise Risk Management: From Incentives to Controls, Wiley Finance.

11.7 List of resources

11.7.1 Books

Presented in alphabetical order.

- Fraser, J., B. Simkins, K. Narvaez (2014). Implementing Enterprise Risk Management: Case Studies and Best Practices, Wiley Finance.
  - Presents case studies of companies that have adopted ERM.

  - Technical/Mathematical level: Accessible
- Jorion, P. (2006). Value at Risk, 3rd Ed.: The New Benchmark for Managing Financial Risk, McGraw-Hill Education.
  - Good reference for ERM.
  - Deals mainly with VaR (value-at-risk) assessment methods in an ERM context.
  - Technical/Mathematical level: Accessible/Technical
- Olson, D.L., D. Wu (2010). Enterprise Risk Management Models.
  - Discusses the fundamental steps in the ERM process.
  - Uses advanced techniques to model risks related to the supply chain and describes the advantage of using ERM for this purpose.
  - Technical/Mathematical level: Accessible/Technical
- Ross, S., R. Westerfield, J. Bradford (2015). Fundamentals of Corporate Finance, McGraw-Hill.
  - Chapter 23 discusses ERM.
  - Important reference in the corporate finance sphere.
  - Technical/Mathematical level: Accessible
- Sweeting, P. (2011). Financial Enterprise Risk Management, Cambridge University Press
  - Describes a range of qualitative and quantitative techniques to identify, model and measure risks in ERM.
  - Technical/Mathematical level: Accessible/Technical

11.7.2 Websites and online reports

- Casualty Actuarial Society (CAS) (2003). Overview of Enterprise Risk Management, Enterprise Risk Management Committee.
  - Excellent summary of ERM from the point of view of P&C insurers. It also provides an exhaustive list of resources on ERM.
  - A must-read. Very accessible.
- Actuarial Standards of Practice (related to ERM):
  - ASOP 46
  - ASOP 47
- CAS ERM:
  - Articles and press releases written by CAS members who discuss several aspects of ERM.
  - As a practice area

Chapter 12

General resources

In this chapter, we provide an overview of resources that might cover multiple areas of credit risk.

12.1 Research papers

Online repositories:

- Social Sciences Research Network (SSRN) http://papers.ssrn.com/sol3/DisplayAbstractSearch.cfm: online repository for articles in the social sciences, which include finance and economics. The entire library comprises over 500,000 papers, of which over 7,000 relate to credit risk. Note that the articles are freely available but are not peer-reviewed.
- DefaultRisk.com http://www.defaultrisk.com/: online repository for articles specifically focusing on corporate credit risk, with over 1,500 freely available papers. Articles are not peer-reviewed unless linked to a specific journal. This website was very active until 2013. The page http://www.defaultrisk.com/papers.htm groups papers by area of research.

Professional organizations:
Professional organizations for actuaries and other risk professionals conduct research in many areas, and the resulting papers are generally available on their respective websites (public or for members only).

Research papers:

- CAS: http://www.casact.org/research/index.cfm?fa=currentresearch
- SOA: https://www.soa.org/research/research-projects/default.aspx
- CIA: http://www.cia-ica.ca/research/research-projects
- CFA Institute: Financial Analysts Journal http://www.cfainstitute.org/learning/products/Pages/index.aspx
- GARP: White Papers http://www.garp.org/#!/risk-intelligence

Conference proceedings and online repositories:

- CAS: http://www.casact.org/research/index.cfm?fa=researchresources
- SOA: https://www.soa.org/BrowsePublication/BrowsePublication.aspx. The search tool on the upper-right part of the main website also points to the same publications.
- CIA: http://www.cia-ica.ca/publications/search-publications
- CFA Institute: http://www.cfainstitute.org/learning/products/Pages/index.aspx

Peer-reviewed scientific journals:
Generally accessible through subscriptions, but universities usually have access to articles published in these journals. The readership is usually academic researchers, and hence, many articles may not be technically accessible to most practitioners. Journals that also target practitioners are indicated with a *.

- Finance and quantitative finance: Journal of Finance, Journal of Financial Economics,

Review of Financial Studies, Journal of Financial and Quantitative Analysis, Journal of Business and Economic Statistics, Journal of Banking and Finance, Journal of Derivatives*, Journal of Fixed Income*, Journal of Credit Risk*, Risk*.

- Actuarial science: North American Actuarial Journal*, Variance*, Insurance: Mathematics & Economics, Scandinavian Actuarial Journal, European Actuarial Journal, ASTIN Bulletin.

12.2 Data

We list the most frequently used databases for various aspects of credit risk modeling and management. The largest databases require a large subscription fee.

- Center for Research in Security Prices (CRSP): maintained by the University of Chicago Booth Business School, the CRSP data contains stock, indexes, mutual fund, Treasury and real estate market prices. http://www.crsp.com/products/research-products
- Datastream (Thomson Reuters): "Datastream is a global financial and macroeconomic database covering equities, stock market indices, currencies, company fundamentals, fixed income securities and key economic indicators for 175 countries and 60 markets." (European University Institute) Prices of credit derivatives can be found in Datastream (http://financial.thomsonreuters.com/en/products/data-analytics/market-data/evaluated-pricing-data.html)
- Compustat (Standard & Poor's): Database of market and accounting information for US and international companies.
- Markit: provides prices of credit default swaps, corporate and municipal bonds and equity volatility data. They maintain the well-known CDX indices. https://www.markit.com/Product/
- Bloomberg http://www.bloomberg.com/enterprise/data/reference-data-services/
- Moody's Analytics http://www.moodysanalytics.com/
- Standard & Poor's Capital IQ https://www.spcapitaliq.com/client-solutions/data

12.3 Computer programs

The online programming community can share its codes on repositories. It might be possible to find credit risk valuation tools or other mathematical tools useful for credit risk modeling, pricing and risk management. It is important to note that those are user-generated code and are not necessarily validated (accuracy, bugs, etc.).

- Matlab: Matlab Central File Exchange http://www.mathworks.com/matlabcentral/fileexchange/
- R: CRAN Packages https://cran.r-project.org/web/packages/
- Python: Python Package Index https://pypi.python.org/pypi
- GitHub: online code repository service https://github.com/. Typing keywords provides available codes, sorted by programming language. Credit risk returns 34 programs, including 6 in R, 2 in Python, 2 in Java and 2 in C#.

12.4 Other resources

Material available online about credit risk can be found in the following:

- Master's and PhD theses: most universities post electronic versions of their students' theses. This can be a valuable, free peer-reviewed resource. Searching each university's website can be tedious. However, Canadian and US theses are available (or can be

requested) through national libraries.
  - US Library of Congress http://www.loc.gov/rr/main/alcove9/education/theses.html: links to several portals
  - Library and Archives Canada, Theses Canada Portal: http://www.bac-lac.gc.ca/eng/services/theses/Pages/theses-canada.aspx
- University course material: many professors put their teaching material online, which can be accessed through a standard Google search. Massively online open courses (MOOC) can also be a very useful resource:
  - https://www.coursera.org/: collaborates with hundreds of universities around the world including top universities such as Princeton, Stanford, and Columbia.
  - https://www.edx.org/: founded by MIT and Harvard, collaborates with other large US and Canadian universities.
  - https://www.class-central.com/: aggregates MOOCs from Coursera, edX, and many other providers of MOOCs. Therefore, readers can search for courses on various topics from various MOOC providers.
- Online courses and webinars (webcasts): professional organizations often provide webinars for their members and non-members that can be viewed online.
  - CAS: http://www.casact.org/education/webinar/
  - SOA: https://www.soa.org/professional-development/archive/webcast-recordings.aspx
  - CIA: http://www.cia-ica.ca/professional-development/webcasts
  - PRMIA: http://www.prmia.org/webinars (webinars) http://www.prmia.org/training/online (online courses)
    - Note that there is an online course on credit risk management on this website: http://www.prmia.org/online-course-group/credit-risk-management
  - GARP: https://www.garp.org/#!/risk-intelligence
  - CFA Institute: http://www.cfainstitute.org/learning/products/Pages/index.aspx. The CFA Institute offers online courses and webcasts.
- Presentation online repository: speakers at conferences, professors, and so forth can share their presentations on these specialized repositories.
  - SlideShare: the most well-known online service, owned by LinkedIn. http://www.slideshare.net

The Mathematics of On-Leveling

Ian Deters

1 Abstract

The mathematical foundation of on-leveling premium is explicitly stated. This is combined with an appropriate set of assumptions to derive the formulae for on-leveling premium by rate book (described within) and for using the Parallelogram Method. It is demonstrated in an appendix that this foundation subsumes all works in the bibliography. It is observed that rate book on-leveling has fewer assumptions than the Parallelogram Method and thus, if database granularity permits, the Parallelogram Method should be abandoned.

2 Introduction

On-leveling premium is an important part of any ratemaking exercise. Typical references (e.g. [2] p. 132, 142; [6] p. 73, 80) give three different methods: extension of exposures, the Parallelogram Method, and aggregating policies based on applicable rate level (i.e. rate book on-leveling). Of these, the first is the ideal. Extension of exposures is, by definition, the correct way to on-level. The challenge of the method is operational. It requires a well designed database and maintenance of rating engines. Rate book on-leveling requires an assumption involving the impact of rate changes and database queries of moderate complexity. Finally, the Parallelogram Method, while requiring virtually nothing more beyond periodic reports of earned premium and exposure, requires many assumptions in order to be tractable.

The shortcomings of actuarial literature with respect to on-leveling are fourfold. First, there is a lack of explicitly stated assumptions. Some works (e.g. [2] p. 133; [3] p. 76; [6], p. 73, 75) verbally state assumptions, but do not translate those assumptions into equations. Consequently, there is no demonstration that the verbally stated assumptions are sufficient to derive a formula with which to on-level one's premium. This concern is not merely theoretical. Such papers remark that there are times that the assumptions of the Parallelogram Method do not hold. If this is true, which it undoubltably is at times, then the mathematical formulations of the assumptions are needed so that they may be adjusted to the situation and a new formula derived with which to on-leveled the premium.

Second, explicit formulae are often omitted. For instance, [2], [3], and [6] each illustrate the Parallelogram Method (p. 133 - 141, p. 103 - 108, and p. 74 - 79 respectively) but never explicitly state a formula. The best job is done by Ross ([5]) who has a general integral formula which is then applied to various examples.

Third, there is an unacknowledged use of model functions. While Ross ([5]) explicitly states that the derivative of his written exposure function is constant, it is not observed that derivative of any written quantity function is 0 almost everywhere (this is demonstrated later in this paper). Hence, when proving claims about a function with a non-zero derivative, it is not true that these claims are being made for some class of well behaved written quantity functions, rather they are being made for no written quantity functions at all. Consquently, the work of Miller and Davis ([4]) and Bill ([1]), which use the work of Ross, have this same problem.

Finally, there is a misplaced emphasis on exposure writing and growth. The works ([1], [4], [5]) which sought to provide a theoretical basis for on-leveling devoted much of their text to assumptions regarding the writing and subsequent earning of exposures. However,

when on-leveling, one is not concerned with exposure. One is concerned with premium. Assumptions about exposure are only useful insofar as they allow one to calculate various quantities of premium.

The practical applications of this paper are two-fold. First, it provides an explicit record in the actuarial literature for the mathematical foundation of on-leveling. Hence, one may combine assumptions germane to one's situation with this foundation to obtain appropriate on-leveling formulae. Second, it demonstrates that rate book on-leveling has fewer assumptions than the Parallelogram Method. Hence, if one is able to determine in which period each amount of premium is written, one may jetison the Parallelogram Method and obtain a more precise estimate of on-leveled earned premium using rate book on-leveling. While a great deal of mathematics and formulae follow, it is not necessary to examine it in order to make use of this second application. In practice, while application of the extension of exposures method may prove difficult due to differeing structures of the rate order of calculation at different times, it is wholly conceivable that, due to the existence of databases containing transactional data, rate book on-leveling may be used and the subsequent exposition be interesting only to those with mathematical proclivities.

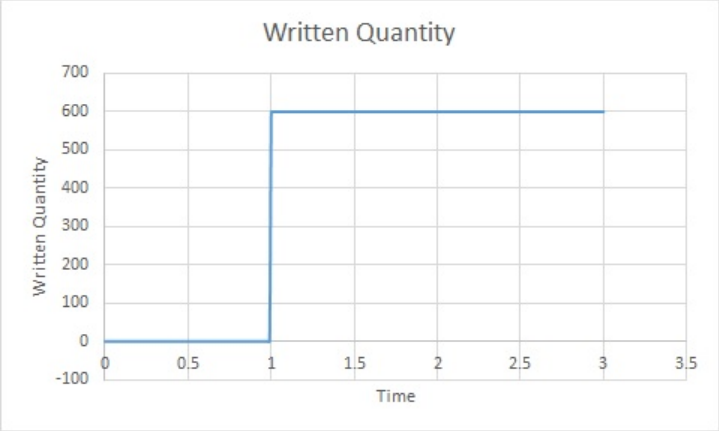
3 Aggregation Of Atomic Transactions

To begin, observe that in insurance, while one is often concerned with policies, policies are not typically the most granular piece of data possessed by the insurance company. Consider the following example.

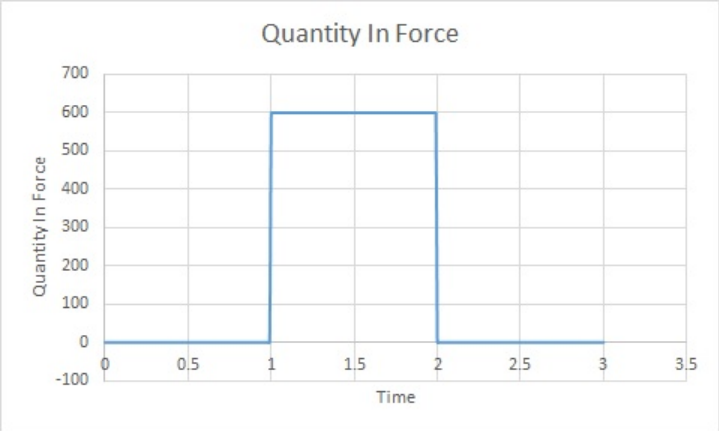
Date	Transaction
01/01/2016	Purchases annual auto policy for $600
04/01/2016	Adds second identical auto for $450
07/01/2016	Adds towinig to first auto for $25
10/01/2016	Cancels second auto's coverage

Three additional changes occured to this policy besides initial writing: a car was added, an endorsement, and a cancelation. The most granular data that a company possesses encompasses transactions such as initial writing of the policy, endorsements to the policy, and cancelation of the policy. The most granular data a company possesses shall, in this paper, be known as **atomic transactions**. In this paper, such transactions will be the basis of considerations of on-leveling premium.

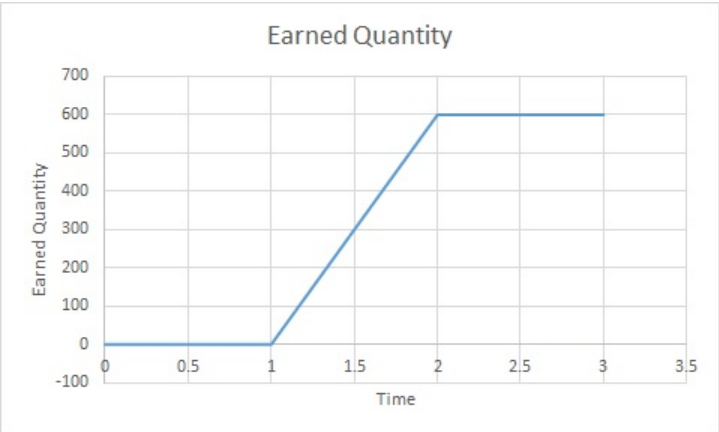
In order to proceed, one must define the set of indicator functions. For all $E \subseteq \mathbb{R}$ define $I_E : \mathbb{R} \to \mathbb{R}$ such that $I_E(x)$ is 1 if $x \in E$ and 0 otherwise. In this way I_E indicates if x is in the set E. For each transaction, one is concerned with the time of the transaction, some amount of some quantity (e.g. exposure or premium), and some length of time (e.g. the term length). Consider an atomic transaction at time $t = t_0$ for amount a and term length τ. Define the amount of written quantity for this transaction from the beginning of the company, at time $t = 0$, through time t, denoted as w, by $w(t) = aI_{([t_0,\infty))}(t)$. A graph of w for the first atomic transaction in the initial example (where time $t = 1$ corresponds to 01/01/2016) is below.



Define the quantity in force for this transaction at time t, denoted by f, by $f(t) = w(t) - w(t-\tau)$. A graph of f is below.



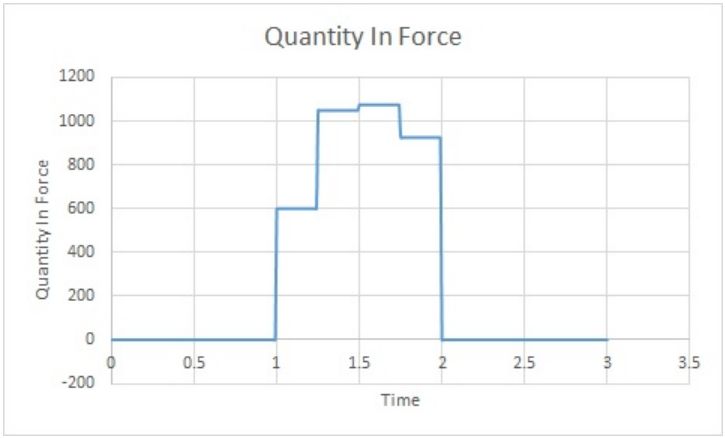
Finally, define the amount of earned quantity for this transaction through time t, denoted as e, by $e(t) = \frac{a}{\tau}(\max(t-t_0, 0) - \max(t-(t_0+\tau), 0))$. A graph of e is below.

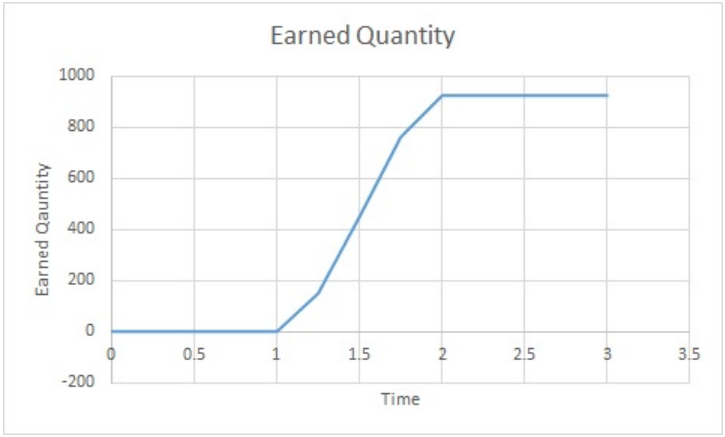


Observe that in the initial example there are four atomic transactions and thus four written quantity, quantity in force, and earned quantity functions. If one subscripts functions corresponding to the ith transaction with i, then the functions are given in the following table.

Written Quantity	Quantity In Force	Earned Quantity
$w_1(t) = 600 \cdot I_{[1,\infty)}(t)$	$f_1(t) = 600 \cdot I_{[1,2]}(t)$	$e_1(t) = 600(\max(t-1) - \max(t-2))$
$w_2(t) = 450 \cdot I_{[1.25,\infty)}(t)$	$f_2(t) = 450 \cdot I_{[1.25,2]}(t)$	$e_2(t) = 450(\max(t-1.25) - \max(t-2))$
$w_3(t) = 25 \cdot I_{[1.5,\infty)}(t)$	$f_3(t) = 25 \cdot I_{[1.5,2]}(t)$	$e_3(t) = 25(\max(t-1.5) - \max(t-2))$
$w_4(t) = -150 \cdot I_{[1.75,\infty)}(t)$	$f_4(t) = -150 \cdot I_{[1.75,2]}(t)$	$e_4(t) = -150(\max(t-1.75) - \max(t-2))$

Thus, the total written quantity for the policy is $w_1 + w_2 + w_3 + w_4$, the total quantity in force is $f_1 + f_2 + f_3 + f_4$, and the total earned quantity is $e_1 + e_2 + e_3 + e_4$. The graphs of these three functions are shown below.





More generally, suppose that n atomic transactions have occured since the beginning of the company and denote the kth specific term length, written quantity, quantity in force, and earned quantity functions by τ_k, w_k, f_k, and e_k respectively. Denote the company's written quantity, quantity in force, and earned quantity functions as w, f, and e respectively and define them as follows

$$w = \sum_{k=1}^{n} w_k, \ f = \sum_{k=1}^{n} f_k, \text{ and } e = \sum_{k=1}^{n} e_k.$$

Since both w and f are linear combinations of indicator functions and each of those functions has a zero time derivative almost everywhere (with respect to Lebesgue measure), then w and f have a zero derivative almost everywhere. This observation becomes important when deriving the Parallelogram Method.

In order to define the policy period quantity functions, fix a period $[t_1, t_2]$. Suppose that there are n atomic transactions pertaining to policies originating in the period $[t_1, t_2]$ and let $\hat{w}_k$, $\hat{f}_k$, and $\hat{e}_k$ be the respective written quantity, quantity in force, and earned quantity functions pertaining to the kth atomic transaction. Similar to how the calendar term functions were defined, denote the company's $[t_1, t_2]$ policy term written quantity, quantity in force, and earned quantity functions as $\hat{w}$, $\hat{f}$, and $\hat{e}$ respectively and define them as follows

$$\hat{w} = \sum_{k=1}^{n} \hat{w}_k, \ \hat{f} = \sum_{k=1}^{n} \hat{f}_k, \text{ and } \hat{e} = \sum_{k=1}^{n} \hat{e}_k.$$

4 On-Leveling

Define $t_0 = 0$ and suppose that the company has implemented n rate changes at times $t_1, t_2, ..., t_n$ where $t_k < t_{k+1}$ for $0 \leq k \leq n-1$. Let $t_{n+1} = \infty$ and $\hat{w}_k$, $\hat{f}_k$, and $\hat{e}_k$ be the policy period written quantity, quantity in force, and earned quantity functions respectively for the period $[t_k, t_{k+1})$ for $0 \leq k \leq n$. For $0 \leq k \leq n$ suppose that there are n_k atomic transactions pertaining to policies originating in the period $[t_k, t_{k+1})$ and for $1 \leq j \leq n_k$ let $\hat{w}_{k,j}$, $\hat{f}_{k,j}$, and $\hat{e}_{k,j}$ be the respective written quantity, quantity in force, and earned quantity functions. Since every atomic transaction is associated to a policy originating in exactly one of these periods, observe that

$$w = \sum_{k=0}^{n} \sum_{j=1}^{n_k} \hat{w}_{k,j} = \sum_{k=0}^{n} \hat{w}_k,$$

$$f = \sum_{k=0}^{n} \sum_{j=1}^{n_k} \hat{f}_{k,j} = \sum_{k=0}^{n} \hat{f}_k,$$

and

$$e = \sum_{k=0}^{n} \sum_{j=1}^{n_k} \hat{e}_{k,j} = \sum_{k=0}^{n} \hat{e}_k.$$

The above equations simply state that any calendar period amount of interest is the sum of all policy period amounts of interest.

Fix a period $[s_1, s_2)$ and consider the task of on-leveling the earned premium in it. Let w_p and e_p be the written premium and earned premium functions and let $\tilde{w}_p$ and $\tilde{e}_p$ be the written premium and earned premium as though the current rates had been in effect since time $t = 0$. Let $\tilde{w}_{p,k}$ and $\tilde{e}_{p,k}$ be the policy period written premium and earned premium for the period $[t_k, t_{k+1})$ for $0 \leq k \leq n$ as though the current rates had been in effect during each of those terms. Hence, while $e_p(s_2) - e_p(s_1)$ is what is recorded in the company's data, the desired, on-leveled quantity is $\tilde{e}_p(s_2) - \tilde{e}_p(s_1)$. Equivalently, one desires the quantity

$$\frac{\tilde{e}_p(s_2) - \tilde{e}_p(s_1)}{e_p(s_2) - e_p(s_1)}.$$

This quantity is often referred to as the on-level factor. If one can re-rate all of its old policies using current rating rules then this number may be computed exactly. If one is unable to do this, then one must make additional assumptions. The first two assumptions, while not possessing names outside of this paper, shall henceforth be referred to as the Classical On-Leveling Assumptions within this paper.

5 The Classical On-Leveling Assumptions

Knowledge of the relationship between $\hat{e}_{p,k}$ and $\tilde{e}_{p,k}$, and thus, the relationship between $\hat{w}_{p,k}$ and $\tilde{w}_{p,k}$, is essential to calculating on-leveled premium. However, it is difficult in practice to know that relationship. This is due to the fact that $\tilde{w}_{p,k}$ may be the result of numerous changes involving reclassifications, rate capping, and other potentially complicated changes. If one is to on-level one's premium while avoiding the use of extension of exposures, one must make some assumption about the relationship between $\hat{w}_{p,k}$ and $\tilde{w}_{p,k}$ with tractable consequences. To proceed, let w_e be the written exposure function, recognize that the mathematical manifestation of the rate is $\frac{dw_p}{dw_e}$, and that implementing a rate change is making a change to $\frac{dw_p}{dw_e}$. For example, suppose that premium is written at a constant rate of \$1000 / exposure. Mathematically, that is expressed as $\frac{dw_p}{dw_e} = 1000$. If the company's first rate change takes place at time $t = 1$ and is a 3% rate change applied to all policies, this would mean that premium is written at a constant rate of \$1030 / exposure after $t = 1$. Mathematically, this would mean that $\frac{dw_p}{dw_e} = 1030$. Since all premium functions are merely sums of atomic transaction functions, we must express the assumption in terms of the atomic transaction written premium functions. To this end, for $0 \leq k \leq n$ and $1 \leq j \leq n_k$, let $\hat{w}_{p,k,j}$ and $\hat{w}_{e,k,j}$ be the written premium and written exposure functions corresponding to the jth atomic transaction associated to a policy originating in the kth policy period. Similarly, let $\tilde{w}_{p,k,j}$ be the written premium function for the jth atomic transaction associated to a policy originating in the kth policy period assuming the current rating structure was in effect. The first classical on-leveling assumption may now be stated in the following way. For $0 \leq k \leq n$ there is some c_k such that

$$\frac{d\tilde{w}_{p,k,j}}{d\hat{w}_{e,k,j}} = c_k \frac{d\hat{w}_{p,k,j}}{d\hat{w}_{e,k,j}} \tag{1}$$

or, more concisely using the chain rule,

$$\frac{d\tilde{w}_{p,k,j}}{d\hat{w}_{p,k,j}} = \frac{d\tilde{w}_{p,k,j}}{d\hat{w}_{e,k,j}} \cdot \frac{d\hat{w}_{e,k,j}}{d\hat{w}_{p,k,j}} = c_k \frac{d\hat{w}_{p,k,j}}{d\hat{w}_{e,k,j}} \cdot \frac{1}{\frac{d\hat{w}_{p,k,j}}{d\hat{w}_{e,k,j}}} = c_k$$

Note that, by definition, $c_n = 1$. Since $\tilde{w}_{p,k,j}(0) = \hat{w}_{p,k,j}(0) = 0$, the first classical on-leveling assumption implies that $\tilde{w}_{p,k,j} = c_k \hat{w}_{p,k,j}$. Observe that this equation holds whether one considers these functions as depending on time or written exposure. Hence, for $0 \leq k \leq n$, $\tilde{e}_k = c_k \hat{e}_k$,

$$\tilde{e}_p(s_2) - \tilde{e}_p(s_1) = \sum_{k=0}^{n} c_k(\hat{e}_{p,k}(s_2) - \hat{e}_{p,k}(s_1)),$$

and the on-level factor is

$$\frac{1}{e_p(s_2) - e_p(s_1)} \sum_{k=0}^{n} c_k(\hat{e}_{p,k}(s_2) - \hat{e}_{p,k}(s_1)) = \sum_{k=0}^{n} c_k \frac{\hat{e}_{p,k}(s_2) - \hat{e}_{p,k}(s_1)}{e_p(s_2) - e_p(s_1)}.$$

The on-level factor is a weighted sum of the average of the policy period % increases (i.e. c_k) weighted by the amount of the $[s_1, s_2)$ calendar period earned premium which came from policy period $[t_k, t_{k+1})$. Thus, if for all calendar period earned premium, one is able to determine the policy period with which that premium is associated, one can calculate the above on-level factor. If one's transaction processing database is sufficiently granular, then a simple database query should be able to accomplish this. This observation is made in [6] (p. 80). This formula and the next is all that is needed to on-level by rate book.

A natural question to arise is how to compute the c_ks. To provide a simple answer to that, another definition must be given. For $0 \leq i \leq n$, let ${}_i\tilde{w}_{p,k,j}$ be the written premium function for the jth atomic transaction associated to a policy originating in the kth policy period assuming the rating structure for the time $[t_i, t_{i+1})$ was in effect. Observe that a consequence of this definition is that ${}_n\tilde{w}_{p,k,j} = \tilde{w}_{p,k,j}$ and ${}_k\tilde{w}_{p,k,j} = \hat{w}_{p,k,j}$. The second classical on-leveling assumption may now be made to help calculate the c_ks. For $0 \leq k \leq n-1$

$$\frac{d\ {}_{k+1}\tilde{w}_{p,k,j}}{d\hat{w}_{e,k,j}} = \frac{c_k}{c_{k+1}} \frac{d\hat{w}_{p,k,j}}{d\hat{w}_{e,k,j}}. \tag{2}$$

That is, ${}_{k+1}\tilde{w}_{p,k,j} = (c_k/c_{k+1})\hat{w}_{p,k,j}$. This is assumption is helpful in the following way. Often when implementing a rate change, a rate impact is calculated on the present book of business. That is, the rates that are going to be in effect in the future are applied to the current book of business and compared to current inforce premium. The above assumption is that that rate impact is equal to c_k/c_{k+1}. For example, suppose that there have been two rate changes, one at time $t = 1$ and another at time $t = 2$, and another one planned for time $t = 3$. Suppose that the rate impact at time $t = 1$ was found to be 3% and at time $t = 2$ is 5%. Then assumption (2) and the fact that $c_3 = 1$ by definition yields

$$c_2 = \frac{c_2}{c_3} c_3 = 1.05 \cdot 1 = 1.05 \text{ and } c_1 = \frac{c_1}{c_2} c_2 = 1.03 \cdot 1.05.$$

This assumption is merely a formalization of the usual reasoning used to calculate the cumulative rate level index ([6], p. 76).

6 Parallelogram Method

If one is unable to on-level one's premium using only the classical on-leveling assumptions, then one must make additional simplifying assumptions. One such set of assumptions forms the basis of The Parallelogram Method. Werner and Modlin in [6] on pages 73 and page 75 respectively state these assumptions in words as

1. "that premium is written evenly throughout the time period" and
2. "the distribution of policies written is uniform over time".

Using our previous notation and the assumption that the number of policies written is proportionate to the number of exposures written, the above may be stated mathematically as

$$\text{for } 0 \leq k \leq n \text{ there is some } r_k \text{ such that } \frac{d\hat{w}_{p,k}}{dt} = r_k \text{ on } [t_k, t_{k+1}) \text{ and} \tag{3}$$

$$\text{there is some } r \text{ such that for } 0 \leq k \leq n \ \frac{d\tilde{w}_{p,k}}{d\hat{w}_{e,k}} = r \text{ on } [t_k, t_{k+1}) \text{ and} \tag{4}$$

$$\text{there is some } c \text{ such that for } 0 \leq k \leq n \ \frac{d\hat{w}_{e,k}}{dt} = c \text{ on } [t_k, t_{k+1}). \tag{5}$$

Assumption (4) states that the on-leveled rate per exposure for each of the time periods is the same. This is the mathematical expression of the idea of a "steady mix of business". While not explicity mentioned in [6], this additional assumption is necessary for the Parallelogram Method to work. An example illustrating the insufficiency of only assumptions (3) and (5) is given later. From assumptions (4) and (5) it follows for $0 \leq k \leq n$ that

$$\frac{d\tilde{w}_{p,k}}{dt} = \frac{d\tilde{w}_{p,k}}{d\hat{w}_{e,k}} \frac{d\hat{w}_{e,k}}{dt} = rc$$

on $[t_k, t_{k+1})$ and 0 elsewhere. Thus, $\frac{d\tilde{w}_p}{dt} = \sum_{k=0}^{n} \frac{d\tilde{w}_{p,k}}{dt} = rc$. Hence, the phrase "that premium is written evenly throughout the time period", if applied to on-leveled premium as well as recorded premium, is sufficient to obtain the results of the parallolgram method and no assumption on the writing of exposures need be made.

It shall be now be established that the assumptions of the Parallelogram Method can be used to derive the classical on-leveling assumptions. First, note that

$$\frac{d\tilde{w}_{p,k}}{d\hat{w}_{e,k}} = r = \frac{rc}{r_k} r_k \frac{1}{c} = \frac{rc}{r_k} \frac{d\hat{w}_{p,k}}{dt} \frac{dt}{d\hat{w}_{e,k}} = \frac{rc}{r_k} \frac{d\hat{w}_{p,k}}{d\hat{w}_{e,k}}.$$

This establishes (1). Second, since $c_k = \frac{rc}{r_k}$ as shown above, and

$$\frac{d\hat{w}_{p,k}}{d\hat{w}_{e,k}} = \frac{d\hat{w}_{p,k}}{dt}\frac{dt}{d\hat{w}_{e,k}} = \frac{r_k}{c},$$

then $r_k c_k = rc$ and, by definition, for $0 \leq k \leq n-1$

$$\frac{d\ {}_{k+1}\tilde{w}_{p,k}}{d\hat{e}_{p,k}} = \frac{r_{k+1}}{c} = \frac{r_{k+1}}{r_k}\frac{r_k}{c} = \frac{c_k}{c_{k+1}}\frac{d\hat{w}_{p,k}}{d\hat{w}_{e,k}}.$$

This establishes (2).

For the purposes of computation, observe that the above also implies $c_k = c_{k+1}\left(\frac{r_{k+1}}{c} \div \frac{r_k}{c}\right)$ for $0 \leq k \leq n-1$. This means simply that c_k is proportional to the rate impact of the $(k+1)$st rate change. This, along with the fact that $c_n = 1$, allows one to compute all c_ks.

It should be mentioned that assumptions (3) and (5) present two theoretical difficulties. As mentioned in the section on atomic transactions, each of those functions has a zero derivative almost everywhere, and thus the sum of those functions has a zero derivative almost everywhere. Hence, if r_k or c is anything other 0, then the functions mentioned in the above assumptions are truly model written quantity functions, not actual written quantity functions. This leads to the second theoretical difficulty. If the functions in the assumptions above are model quantity functions and not tied to actual observed quantities or atomic transactions, then an additional assumption must be made in order to derive an earned quantity from a written quantity. This leads to the fourth Parallelogram Method assumption.

There is some τ such that all atomic transactions have a term length of τ. (6)

Functionally, this means that no policies will be canceled and no endorsements will be added or canceled. Using Assumption (6) and the notation from the first section, an equation linking written quantity and earned quantity functions may be derived by observing

$$f(t) = \sum_{k=1}^{n} f_k(t) = \sum_{k=1}^{n} w_k(t) - w_k(t-\tau_k) = \sum_{k=1}^{n} w_k(t) - w_k(t-\tau) = w(t) - w(t-\tau)$$

and

$$\begin{aligned} e(t) &= \sum_{k=1}^{n} e_k(t) = \sum_{k=1}^{n} \frac{a_k}{\tau_k}(\max(t-t_k,0) - \max(t-(t_k+\tau_k),0)) = \sum_{k=1}^{n}\int_0^t \frac{1}{\tau_k}(w_k(s) - w_k(s-\tau_k))ds \\ &= \int_0^t \frac{1}{\tau}\sum_{k=1}^{n} w_k(s) - w_k(s-\tau)ds = \int_0^t \frac{1}{\tau}(w(s) - w(s-\tau))ds = \int_0^t \frac{1}{\tau}f(s)ds. \end{aligned}$$

Based on the above equations, if w is a model written quantity function and τ is the term length for all transactions, then define f by $f(t) = w(t) - w(t-\tau)$ and e by $e(t) = \int_0^t \frac{1}{\tau}(w(t) - w(t-\tau))dt$.

Using assumption (3) and the fact that $\hat{w}_{p,k}(t_k) = 0$, formulae for policy term earned premium may be developed. In particular,

$$\hat{w}_{p,k}(t) = \begin{array}{ll} 0 & t \le t_k \\ r_k(t - t_k) & t_k \le t \le t_{k+1}. \\ r_k(t_{k+1} - t_k) & t_{k+1} \le t \end{array}$$

The premium in force and earned premium functions for the cases $t_{k+1} - t_k \le \tau$ and $\tau \le t_{k+1} - t_k$ are slightly different but may be compactly stated as below. Their development is given in Appendix A. The general form of the premium in force is

$$\hat{f}_{p,k}(t) = \begin{array}{ll} 0 & t \le t_k \\ r_k(t - t_k) & t_k \le t \le \min(t_k + \tau, t_{k+1}) \\ r_k(\min(t_{k+1}, t) - \max(t_k, t - \tau)) & \min(t_k + \tau, t_{k+1}) \le t \le \max(t_k + \tau, t_{k+1}), \\ r_k(t_{k+1} + \tau - t) & \max(t_k + \tau, t_{k+1}) \le t \le t_{k+1} + \tau \\ 0 & t_{k+1} + \tau \le t \end{array}$$

and the $[t_k, t_{k+1}]$ policy term earned premium through time t is

$$\hat{e}_{p,k}(t) = \begin{array}{ll} 0 & t \le t_k \\ .5\frac{r_k}{\tau}(t - t_k)^2 & t_k \le t \le \min(t_k + \tau, t_{k+1}) \\ .5\frac{r_k}{\tau}\min(t_{k+1} - t_k, \tau)^2 & \min(t_k + \tau, t_{k+1}) \le t \le \max(t_k + \tau, t_{k+1}) \\ +\frac{r_k}{\tau}\min(t_{k+1} - t_k, \tau) & \\ \cdot(t - (\min(t_{k+1} - t_k, \tau) + t_k)) & \\ \frac{r_k}{\tau}\min(t_{k+1} - t_k, \tau)^2 & \max(t_k + \tau, t_{k+1}) \le t \le t_{k+1} + \tau \\ +\frac{r_k}{\tau}\min(t_{k+1} - t_k, \tau) & \\ \cdot(\max(t_{k+1} - t_k, \tau) - \min(t_{k+1} - t_k, \tau)) & \\ -.5\frac{r_k}{\tau}(t_{k+1} + \tau - t)^2 & \\ \frac{r_k}{\tau}\min(t_{k+1} - t_k, \tau)^2 & t_{k+1} + \tau \le t. \\ +\frac{r_k}{\tau}\min(t_{k+1} - t_k, \tau) & \\ \cdot(\max(t_{k+1} - t_k, \tau) - \min(t_{k+1} - t_k, \tau)) & \end{array}$$

7 Parallelogram Examples

7.1 Example 1

For an example, consider the "simple Example" given in [6] starting on p.72. In this example the policy term is annual (i.e. that is $\tau = 1$).

Rate Level Group	Effective Date	Overall Average Rate Change
1	Initial	$--$
2	07/01/2010	5%
3	01/01/2011	10%
4	04/01/2012	-1%

The task is to on-level calendar year 2011 earned premium. Suppose that time $t = 0$ corresponds to 01/01/2010. Let $\hat{w}_{p,0}$, $\hat{w}_{p,1}$, $\hat{w}_{p,2}$, and $\hat{w}_{p,3}$ be the policy term written premium functions corresponding to the terms $[0, .5]$, $[.5, 1]$, $[1, 2.25]$, and $[2.25, \infty)$ respectively and let $\hat{e}_{p,0}$, $\hat{e}_{p,1}$, $\hat{e}_{p,2}$, and $\hat{e}_{p,3}$ be the corresponding policy term earned premium functions. Based on the table of rate changes, we have that

$$c_0 = 1.05 \cdot 1.1 \cdot .99$$
$$c_1 = 1.1 \cdot .99$$
$$c_2 = .99$$
$$c_3 = 1.$$

Since we wish to on-level the 2011 calendar year, we calculate

$$\hat{e}_{p,0}(2) - \hat{e}_{p,0}(1) = \frac{1}{1}(r_0 \cdot .5^2 + r_0 \cdot .5(1-.5)) - \frac{1}{1}(.5r_0 \cdot .5^2 + r_0 \cdot .5(1-.5)) = .5r_0 - .375r_0 = .125r_0$$
$$\hat{e}_{p,1}(2) - \hat{e}_{p,1}(1) = \frac{1}{1}(r_1 \cdot .5^2 + r_1 \cdot .5(1-.5)) - \frac{1}{1}.5r_1(1-.5)^2 = .5r_1 - .125r_1 = .375r_1 = .375 \cdot 1.05r_0$$
$$\hat{e}_{p,2}(2) - \hat{e}_{p,2}(1) = \frac{1}{1}.5r_2 \cdot 1^2 - \frac{1}{1} \cdot 0 = .5r_2 = .5 \cdot 1.05 \cdot 1.1r_0$$
$$\hat{e}_{p,3}(2) - \hat{e}_{p,3}(1) = 0 - 0 = 0.$$

Let $\tilde{e}_p(t)$ be the earned premium function assuming that all premium has been written at current rates. Then

$$\tilde{e}_p(2) - \tilde{e}_p(1) = \frac{1}{1}(.5r_3 \cdot 1 + r_3 \cdot 1(2-1)) - \frac{1}{1} \cdot .5r_3 \cdot 1^2 = r_3 = 1.05 \cdot 1.1 \cdot .99r_0.$$

Hence, the on-level factor for calendar year 2011 is

$$\frac{1.05 \cdot 1.1 \cdot .99r_0}{.125r_0 + .375 \cdot 1.05r_0 + .5 \cdot 1.05 \cdot 1.1r_0} \approx 1.0431.$$

7.2 Example 2

The second example is like the first example except that the policy term is 6 months. (i.e. that is $\tau = .5$). Using the same notation as before, we calculate

$$\hat{e}_{p,0}(2) - \hat{e}_{p,0}(1) = \frac{1}{.5}(r_0 \cdot .5^2 + r_0 \cdot .5 \cdot 0) - \frac{1}{.5}(r_0 \cdot .5^2 + r_0 \cdot .5 \cdot 0) = 0$$
$$\hat{e}_{p,1}(2) - \hat{e}_{p,1}(1) = \frac{1}{.5}(r_1 \cdot .5^2 + r_1 \cdot .5 \cdot 0) - \frac{1}{.5}(.5r_1 \cdot .5^2) = .25r_1 = .25 \cdot 1.05r_0$$
$$\hat{e}_{p,2}(2) - \hat{e}_{p,2}(1) = \frac{1}{.5}(.5r_2 \cdot .5^2 + r_2 \cdot .5 \cdot (2 - (.5+1))) - \frac{1}{.5}0 = .75r_2 = .75 \cdot 1.05 \cdot 1.1r_0$$
$$\hat{e}_{p,3}(2) - \hat{e}_{p,3}(1) = \frac{1}{.5}0 - \frac{1}{.5}0 = 0.$$

Let $\tilde{e}_p(t)$ be the earned premium function assuming that all premium has been written at current rates. Then

$$\tilde{e}_p(2) - \tilde{e}_p(1) = \frac{1}{.5}(.5r_3 \cdot .5^2 + r_3 \cdot .5(2-.5)) - \frac{1}{.5}(.5r_3 \cdot .5^2 + r_3 \cdot .5(1-.5)) = r_3 = 1.05 \cdot 1.1 \cdot .99r_0.$$

Hence, the on-level factor for calendar year 2011 is

$$\frac{1.05 \cdot 1.1 \cdot .99 r_0}{.25 \cdot 1.05 r_0 + .75 \cdot 1.05 \cdot 1.1 r_0} \approx 1.0130.$$

7.3 Example 3

In this example, we demonstrate the necessity of assumption (4) in the Parallelogram Method. To this end, suppose that policies are annual (e.g. $\tau = 1$), there was a single rate change on 01/01/2011. Suppose that time $t = 0$ corresponds to 01/01/2010. Let $\hat{w}_{p,0}$ and $\hat{w}_{p,1}$ be the policy term written premium functions corresponding to the terms $[0, 1]$ and $[1, 2]$ respectively and let $\hat{e}_{p,0}$ and $\hat{e}_{p,1}$ be the corresponding policy term earned premium functions. Assume further that

$$\begin{aligned}
\frac{d\hat{w}_{e,1}}{dt} &= \frac{d\hat{w}_{e,2}}{dt} = 1000 \\
\frac{d\hat{w}_{p,1}}{d\hat{w}_{e,1}} &= 100 \\
\frac{d\tilde{w}_{p,1}}{d\hat{w}_{e,1}} &= 105 \\
\frac{d\tilde{w}_{p,2}}{d\hat{w}_{e,2}} &= \frac{d\hat{w}_{p,2}}{d\hat{w}_{e,1}} = 110.
\end{aligned}$$

That is, 1000 exposures are being written each year. In 2010, the written premium per exposure is \$100 and is \$105 at current rate level. In 2011, the written premium per exposure is \$110. Hence,

$$\begin{aligned}
\hat{e}_{p,1}(2) - \hat{e}_{p,1}(1) &= 100 \cdot 1000 \cdot 1^2 + 100 \cdot 1000 \cdot 1 \cdot (1-1) - .5 \cdot 100 \cdot 1000 \cdot (1+1-2) - .5 \cdot 100 \cdot 1000 \cdot 1^2 \\
&= 50000 \\
\hat{e}_{p,2}(2) - \hat{e}_{p,2}(1) &= .5 \cdot 110 \cdot 1000 \cdot (2-1)^2 - 0 = 55000.
\end{aligned}$$

Let $\tilde{e}_p(t)$ be the earned premium function assuming that all premium has been written at current rates. Then

$$\begin{aligned}
\tilde{e}_p(2) - \tilde{e}_p(1) &= (\tilde{e}_{p,1}(2) - \tilde{e}_{p,1}(1)) + (\tilde{e}_{p,2}(2) - \tilde{e}_{p,2}(1)) \\
&= (105 \cdot 1000 \cdot 1^2 + 105 \cdot 1000 \cdot 1 \cdot (1-1) - .5 \cdot 105 \cdot 1000 \cdot (1+1-2) - .5 \cdot 105 \cdot 1000 \cdot 1^2) \\
&+ 55000 = 52500 + 55000 = 107,500.
\end{aligned}$$

Hence, the on-level factor for calendar year 2011 is

$$\frac{107500}{105000} \approx 1.0238 \neq 1.0244 \approx= \frac{1.05}{.5 + .5 \cdot 1.05}.$$

Observe that the factor does not match the form typically prescribed by the Parallelogram Method.

8 Conclusion

On-leveling premium is merely an arithmetic problem. The impediment to the extension of exposures is technology and data driven. Thus, the pace of technology and availability of data will soon make papers like this obselete. Until then, if one is unable to apply the extension of exposures technique, then one must make some assumptions. The most popular set of assumptions, as indicated by a review of the literature, are those which give rise to the Parallelogram Method. The popularity of this method is unsurprising given that it originated in a time when there was not access to fast databases and that it can be applied using periodic reports of earned premium and exposure. However, this is a different time. Today, even if one does not have all of the data or the rating engines to use the extension of exposures, one can certainly, with a few database queries determine under what rate book every piece of earned premium was written. Hence, one may, with only a modicum of effort, on-level premium by rate book. Moreover, as shown by this paper, the assumptions governing the Parallelogram Method are much stronger than those needed to on-level by rate book. Therefore, it is the opinion of this author that the reader experiment by comparing premium on-leveled by the three different methods and determine whether on-leveling by the Parallelogram Method or by rate book is closer to the correct answer of on-leveling by extension of exposures. The importance of this answer lies in its ability to allow one to price more accurately. The more accurate one's on-leveling procedure, the less one's premium trend calculations must account for one's on-leveling inaccuracies, and the more precise one's projection of premium will be.

9 Appendix - Derivation Of Parallelgram Method

In order to obtain the general expression of the premium in force function for the Parallelogram Method, first consider the case $t_{k+1} - t_k \leq \tau$. In this case,

$$\hat{f}_{p,k}(t) = \begin{array}{ll} 0 & t \leq t_k \\ r_k(t - t_k) & t_k \leq t \leq t_{k+1} \\ r_k(t_{k+1} - t_k) & t_{k+1} \leq t \leq t_k + \tau \\ r_k(t_{k+1} + \tau - t) & t_k + \tau \leq t \leq t_{k+1} + \tau \\ 0 & t_{k+1} + \tau \leq t \end{array}$$

Next, consider the case $\tau \leq t_{k+1} - t_k$. In this case,

$$\hat{f}_{p,k}(t) = \begin{array}{ll} 0 & t \leq t_k \\ r_k(t - t_k) & t_k \leq t \leq t_k + \tau \\ r_k\tau & t_k + \tau \leq t \leq t_{k+1} \\ r_k(t_{k+1} + \tau - t) & t_{k+1} \leq t \leq t_{k+1} + \tau \\ 0 & t_{k+1} + \tau \leq t \end{array}$$

Combining the above results yields the general form of the premium in force function as

$$\hat{f}_{p,k}(t) = \begin{array}{ll} 0 & t \leq t_k \\ r_k(t - t_k) & t_k \leq t \leq \min(t_k + \tau, t_{k+1}) \\ r_k(\min(t_{k+1}, t) - \max(t_k, t - \tau)) & \min(t_k + \tau, t_{k+1}) \leq t \leq \max(t_k + \tau, t_{k+1}). \\ r_k(t_{k+1} + \tau - t) & \max(t_k + \tau, t_{k+1}) \leq t \leq t_{k+1} + \tau \\ 0 & t_{k+1} + \tau \leq t \end{array}$$

Similarly, the derivation of the general form of the earned premium function must be split

into two cases. In the first case, $t_{k+1} - t_k \leq \tau$ and $\min(t_k + \tau, t_{k+1}) = t_{k+1}$. Hence,

$$\int_0^t \frac{1}{\tau}\hat{f}_{p,k}(s)ds = \int_0^t 0 = 0 \text{ for } t \leq t_k,$$

$$\int_0^t \frac{1}{\tau}\hat{f}_{p,k}(s)ds = \int_0^{t_k} 0 + \int_{t_k}^t \frac{r_k}{\tau}(t - t_k) = .5\frac{r_k}{\tau}(t - t_k)^2 \text{ for } t_k \leq t \leq \min(t_k + \tau, t_{k+1}),$$

$$\int_0^t \frac{1}{\tau}\hat{f}_{p,k}(s)ds = \int_0^{\min(t_k+\tau,t_{k+1})} \frac{1}{\tau}\hat{f}_{p,k}(s)ds + \int_{\min(t_k+\tau,t_{k+1})}^t \frac{r_k}{\tau}(\min(t_{k+1}, s) - \max(t_k, s - \tau))ds$$

$$= .5\frac{r_k}{\tau}(t_{k+1} - t_k)^2 + \int_{t_{k+1}}^t \frac{r_k}{\tau}(t_{k+1} - t_k)ds = .5\frac{r_k}{\tau}(t_{k+1} - t_k)^2 + \frac{r_k}{\tau}(t_{k+1} - t_k)(t - t_{k+1})$$

$$= .5\frac{r_k}{\tau}\min(t_{k+1} - t_k, \tau)^2 + \frac{r_k}{\tau}\min(t_{k+1} - t_k, \tau)(t - (\min(t_{k+1} - t_k, \tau) + t_k))$$

for $\min(t_k + \tau, t_{k+1}) \leq t \leq \max(t_k + \tau, t_{k+1})$,

$$\int_0^t \frac{1}{\tau}\hat{f}_{p,k}(s)ds = \int_0^{\max(t_k+\tau,t_{k+1})} \frac{1}{\tau}\hat{f}_{p,k}(s)ds + \int_{\max(t_k+\tau,t_{k+1})}^t \frac{1}{\tau}\hat{f}_{p,k}(s)ds$$

$$= .5\frac{r_k}{\tau}(t_{k+1} - t_k)^2 + \frac{r_k}{\tau}(t_{k+1} - t_k)(t_k + \tau - t_{k+1}) + \int_{t_k+\tau}^t \frac{r_k}{\tau}(t_{k+1} + \tau - s)ds$$

$$= .5\frac{r_k}{\tau}(t_{k+1} - t_k)^2 + \frac{r_k}{\tau}(t_{k+1} - t_k)(t_k + \tau - t_{k+1}) + .5\frac{r_k}{\tau}((t_{k+1} - t_k)^2 - (t_{k+1} + \tau - t)^2)$$

$$= \frac{r_k}{\tau}(t_{k+1} - t_k)^2 + \frac{r_k}{\tau}(t_{k+1} - t_k)(t_k + \tau - t_{k+1}) - .5\frac{r_k}{\tau}(t_{k+1} + \tau - t)^2$$

$$= \frac{r_k}{\tau}\min(t_{k+1} - t_k, \tau)^2 + \frac{r_k}{\tau}\min(t_{k+1} - t_k, \tau)(\max(t_{k+1} - t_k, \tau) - \min(t_{k+1} - t_k, \tau)) - .5\frac{r_k}{\tau}(t_{k+1} + \tau - t)^2$$

for $\max(t_k + \tau, t_{k+1}) \leq t \leq t_{k+1} + \tau$, and

$$\int_0^t \frac{1}{\tau}\hat{f}_{p,k}(s)ds = \int_0^{t_{k+1}+\tau} \frac{1}{\tau}\hat{f}_{p,k}(s)ds + \int_{t_{k+1}+\tau}^t \frac{1}{\tau}\hat{f}_{p,k}(s)ds$$

$$= \frac{r_k}{\tau}(t_{k+1} - t_k)^2 + \frac{r_k}{\tau}(t_{k+1} - t_k)(\max(t_{k+1} - t_k, \tau) - \min(t_{k+1} - t_k, \tau)) + \int_{t_{k+1}+\tau}^t 0ds$$

$$= \frac{r_k}{\tau}\min(t_{k+1} - t_k, \tau)^2 + \frac{r_k}{\tau}\min(t_{k+1} - t_k, \tau)(\max(t_{k+1} - t_k, \tau) - \min(t_{k+1} - t_k, \tau)) \text{ for } t_{k+1} \leq t.$$

In the second case, $\tau \leq t_{k+1} - t_k$ and $\min(t_k + \tau, t_{k+1}) = t_k + \tau$. Hence,

$$\int_0^t \frac{1}{\tau}\hat{f}_{p,k}(s)ds = \int_0^t 0 = 0 \text{ for } t \le t_k,$$

$$\int_0^t \frac{1}{\tau}\hat{f}_{p,k}(s)ds = \int_0^{t_k} 0 + \int_{t_k}^t \frac{r_k}{\tau}(t-t_k) = .5\frac{r_k}{\tau}(t-t_k)^2 \text{ for } t_k \le t \le \min(t_k+\tau, t_{k+1}),$$

$$\int_0^t \frac{1}{\tau}\hat{f}_{p,k}(s)ds = \int_0^{\min(t_k+\tau,t_{k+1})} \frac{1}{\tau}\hat{f}_{p,k}(s)ds + \int_{\min(t_k+\tau,t_{k+1})}^t \frac{r_k}{\tau}(\min(t_{k+1},s) - \max(t_k, s-\tau))ds$$

$$= .5\frac{r_k}{\tau}\tau^2 + \int_{t_k+\tau}^t \frac{r_k}{\tau}\tau ds = .5\frac{r_k}{\tau}\tau^2 + \frac{r_k}{\tau}\tau(t-(t_k+\tau))$$

$$= .5\frac{r_k}{\tau}\min(t_{k+1}-t_k,\tau)^2 + \frac{r_k}{\tau}\min(t_{k+1}-t_k,\tau)(t-(\min(t_{k+1}-t_k,\tau)+t_k))$$

for $\min(t_k+\tau, t_{k+1}) \le t \le \max(t_k+\tau, t_{k+1})$,

$$\int_0^t \frac{1}{\tau}\hat{f}_{p,k}(s)ds = \int_0^{\max(t_k+\tau,t_{k+1})} \frac{1}{\tau}\hat{f}_{p,k}(s)ds + \int_{\max(t_k+\tau,t_{k+1})}^t \frac{1}{\tau}\hat{f}_{p,k}(s)ds$$

$$= .5\frac{r_k}{\tau}\tau^2 + \frac{r_k}{\tau}\tau(t_{k+1}-(t_k+\tau)) + \int_{t_{k+1}}^t \frac{r_k}{\tau}(t_{k+1}+\tau-s)ds$$

$$= .5\frac{r_k}{\tau}\tau^2 + \frac{r_k}{\tau}\tau(t_{k+1}-(t_k+\tau)) + .5\frac{r_k}{\tau}(\tau^2-(t_{k+1}+\tau-t)^2)$$

$$= \frac{r_k}{\tau}\tau^2 + \frac{r_k}{\tau}\tau(t_{k+1}-(t_k+\tau)) - .5\frac{r_k}{\tau}(t_{k+1}+\tau-t)^2$$

$$= \frac{r_k}{\tau}\min(t_{k+1}-t_k,\tau)^2 + \frac{r_k}{\tau}\min(t_{k+1}-t_k,\tau)(\max(t_{k+1}-t_k,\tau) - \min(t_{k+1}-t_k,\tau)) - .5\frac{r_k}{\tau}(t_{k+1}+\tau-t)^2$$

for $\max(t_k+\tau, t_{k+1}) \le t \le t_{k+1}+\tau$, and

$$\int_0^t \frac{1}{\tau}\hat{f}_{p,k}(s)ds = \int_0^{t_{k+1}+\tau} \frac{1}{\tau}\hat{f}_{p,k}(s)ds + \int_{t_{k+1}+\tau}^t \frac{1}{\tau}\hat{f}_{p,k}(s)ds$$

$$= \frac{r_k}{\tau}\tau^2 + \frac{r_k}{\tau}\tau(\max(t_{k+1}-t_k,\tau) - \min(t_{k+1}-t_k,\tau)) + \int_{t_{k+1}+\tau}^t 0 ds$$

$$= \frac{r_k}{\tau}\min(t_{k+1}-t_k,\tau)^2 + \frac{r_k}{\tau}\min(t_{k+1}-t_k,\tau)(\max(t_{k+1}-t_k,\tau) - \min(t_{k+1}-t_k,\tau)) \text{ for } t_{k+1} \le t.$$

10 Appendix - Reconciliation To Previous Papers

There are three main papers which seek to give a systematic treatment to on-leveling premium. It should be noted that the expressions given in this paper, under a change of notation, yield the same results as those papers. Also, these papers are considering model written exposure functions instead of those related to actual atomic transactions. The first paper, [5], is that of Ross from 1975. On p. 52 in [5], Ross presents a formula, denoted as $EE(x_0, x_1)$, for the amount of exposures earned between time x_0 and x_1, denotes time and policy term length using the letters x and t respectively, and writes "Let the function $f(x)$ stand for the rate of exposure writing at time x". In order to see the equivalence of formulae, note that the set of points in the plane which describe the region $\{(x,s) : x_0 \le x \le x_1, x-\tau \le s \le x\}$ can also be described by

$$\begin{array}{ll} x_0 \le s \le x_0 & x_0 \le x \le s+\tau \\ x_0 \le s \le x_1-\tau & s \le x \le s+\tau \\ x_1-\tau \le s \le x_1 & s \le x \le x_1. \end{array}$$

Using his notation, the equivalence between his formula the one developed in this paper is as follows:

$$\begin{aligned} e_e(x_1)-e_e(x_0) &= \frac{1}{\tau}\int_{x_0}^{x_1} f_e(x)dx = \frac{1}{\tau}\int_{x_0}^{x_1} w_e(x)-w_e(x-\tau)dx = \frac{1}{\tau}\int_{x_0}^{x_1}\int_{x-\tau}^{x} f(s)dsdx \\ &= \frac{1}{\tau}\left(\int_{x_0-\tau}^{x_0}\int_{x_0}^{s+t} f(s)dxds + \int_{x_0}^{x_1-\tau}\int_{s}^{s+t} f(s)dxds + \int_{x_1-\tau}^{x_1}\int_{s}^{x_1} f(s)dxds\right) \\ &= \int_{x_0-\tau}^{x_0}\frac{s+\tau-x_0}{\tau}f(s)ds + \int_{x_0}^{x_1-\tau} f(s)ds + \int_{x_1-\tau}^{x_1}\frac{x_1-s}{\tau}f(s)ds = EE(x_0,x_1). \end{aligned}$$

The second paper, [4], is that of Miller and Davis from 1976. On p. 121 in [4], Miller and Davis, giving a geometric interpretation to the work of Ross, derived a formula equivalent to his. They used notation similar to his except that they denoted the term length as k. Using their notation, equivalence between their formula and the one developed in this paper is as follows:

$$\begin{aligned} e_e(x_1)-e_e(x_0) &= \frac{1}{\tau}\int_{x_0}^{x_1} f_e(x)dx = \frac{1}{\tau}\int_{x_0}^{x_1} w_e(x)-w_e(x-\tau)dx = \frac{1}{\tau}\int_{x_0}^{x_1}\int_{x-\tau}^{x} f(s)dsdx \\ &= \int_{x_0}^{x_1}\int_{x}^{x-\tau} -\frac{1}{\tau}f(s)dsdx = \int_{x_0}^{x_1}\int_{0}^{1} f(x-\tau s)dsdx = EE(x_0,x_1). \end{aligned}$$

The third paper, [1], is that of Bill from 1989. Bill's work is an application of the formulae of Ross (p. 207). Hence, the work of Bill can also be derived from the formulae presented in this paper.

Finally, since the formulae implicitly referenced in [2], [3], and [6] are stated explicitly and developed from a more general set of equations, the claim set forth in the abstract, that this paper subsumes all works in the bibliography, is established.

References

[1] Bill, Richard (1989). Generalized Earned Premium Rate Adjustment Factors. Casualty Actuarial Society Forum, Fall 1989 Edition, pp. 201 - 218.

[2] Friedland, Jacqueline. Fundamentals Of General Insurance Actuarial Analysis, Society Of Actuaries, 2014.

[3] Kallop, Roy H. (1975). A Current Look At Workers' Compensation Ratemaking. In Proceedings of the Casualty Actuarial Society, Volume LXII, Numbers 117 & 118, pp. 62 - 133.

[4] Miller, David L. and Davis, George E. (1976). A Refined Model For Premium Adjustment. In Proceedings of the Casualty Actuarial Society, Volume LXIII, Numbers 119 & 120, pp. 117 - 124.

[5] Ross, James P. (1975). Generalized Premium Formulae. In Proceedings of the Casualty Actuarial Society, Volume LXII, Numbers 117 & 118, pp. 50 - 61.

[6] Werner, Geoff and Claudine Modlin, Basic Ratemaking, Casualty Actuarial Society, 2010.

An Alternative Approach to Credibility for Large Account and Excess of Loss Treaty Pricing

Uri Korn, FCAS, MAAA

Abstract: This paper illustrates a comprehensive approach to utilizing and credibility weighting all available information for large account and excess of loss treaty pricing. The typical approach to considering the loss experience above the basic limit is to analyze the burn costs in these excess layers directly (see Clark 2011, for example). Burn costs are extremely volatile in addition to being highly right skewed, which does not perform well with linear credibility methods, such as Buhlmann-Straub or similar methods (Venter 2003). Using burn costs also involves developing and making a selection for each excess layer, which can be cumbersome. Also the formulas for calculating all of the correlations needed for determining the credibilities are complicated.

An alternative approach is shown that uses all of the available data in a more robust and seamless manner. Credibility weighting of the account's experience with the exposure cost for the basic limit is performed using Buhlmann-Straub credibility. Modified formulae are shown that are more suitable for this scenario. For the excess layers, the excess losses themselves are utilized to modify the severity distribution that is used to calculate the increased limit factors. This is done via a simple Bayesian credibility technique that does not require any specialized software to run. Such an approach considers all available information in the same way as analyzing burn costs, but does not suffer from the same pitfalls. Certain modifications are also illustrated to produce a method that does not differentiate between basic limit and the excess losses. Lastly, it is shown how the method can be improved for higher layers by leveraging Extreme Value Theory.

Keywords. Buhlmann-Straub Credibility, Bayesian Credibility, Loss Rating, Exposure Rating, Burn Cost, Extreme Value Theory

1. INTRODUCTION

This paper illustrates a comprehensive approach to utilizing and credibility weighting all available information for large account and excess of loss treaty pricing. The typical approach to considering the loss experience above the basic limit is to analyze the burn costs in these excess layers directly (see Clark 2011, for example). Burn costs are extremely volatile in addition to being highly right skewed, which does not perform well with linear credibility methods, such as Buhlmann-Straub or similar methods (Venter 2003). Using burn costs also involves developing and making a selection for each excess layer, which can be cumbersome. Also, the formulas for calculating all of the correlations needed for determining the credibilities are complicated.

An alternative approach is shown that uses all of the available data in a more robust and seamless manner. Credibility weighting of the account's experience with the exposure cost[1] for the basic limit

[1] Throughout this paper, the following definitions will be used:
Exposure cost: Pricing of an account based off of the insured characteristics and size using predetermined rates
Experience cost: Pricing of an account based off of the insured's actual losses. An increased limits factor is then usually applied to this loss pick to make the estimate relevant for a higher limit or layer.

is performed using Buhlmann-Straub credibility. Modified formulae are shown that are more suitable for this scenario. For the excess layers, the excess losses themselves are utilized to modify the severity distribution that is used to calculate the increased limit factors. This is done via a simple Bayesian credibility technique that does not require any specialized software to run. Such an approach considers all available information in the same way as analyzing burn costs, but does not suffer from the same pitfalls. Certain modifications are also illustrated to produce a method that does not differentiate between basic limit and the excess losses. Lastly, it is shown how the method can be improved for higher layers by leveraging Extreme Value Theory.

1.1 Research Context

Clark (2011) as well as Marcus (2010) and many others develop an approach for credibility weighting all of the available account information up an excess tower. The information considered is in the form of the exposure cost for each layer, the capped loss cost estimate for the chosen basic limit, and the burn costs associated with all of the layers above the basic limit up to the policy layer. Formulae are shown for calculating all of the relevant variances and covariances between the different methods and between the various layers, which are needed for calculating all of the credibilities.

This paper takes a different approach and uses the excess losses to modify the severity distribution that is used to calculate the ILF; this is another way of utilizing all of the available account information. This technique does not require the development and selection of a burn cost estimate for every excess layer. It also does not suffer from the problem of applying linear credibility methods, such as Buhlmann-Straub or similar methods, to highly skewed values, which can result in large errors (Venter 2003). Excess burn costs are definitely highly skewed.

1.2 Objective

The goal of this paper is to show how all available information pertaining to an account in terms of the exposure cost estimate and the loss information can be incorporated to produce an optimal estimate of the prospective cost.

1.3 Outline

Section 2 provides a review of account rating and gives a quick overview of the current approaches. Section 3 discusses credibility weighting of the basic layer loss cost, and section 4 shows strategies for credibility weighting the excess losses with the portfolio severity distribution. The end of this section shows simulation results to illustrate the relative benefit that can be achieved from this alternative

Burn Cost: Pricing of an excess account based off of the insured's actual losses in a non-ground up layer.

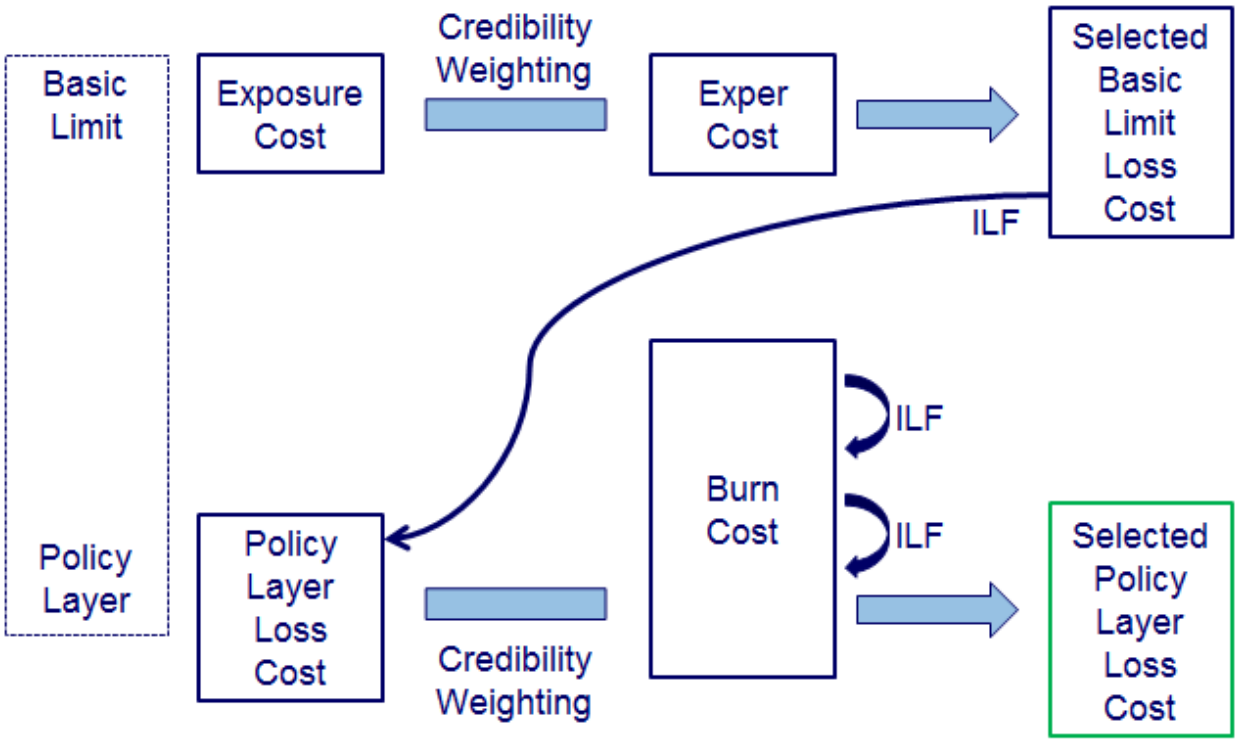
method, even with only a small number of claims. Section 5 shows how Extreme Value Theory can be leveraged to improve the method for high layers.

2. A BRIEF OVERVIEW OF ACCOUNT RATING AND THE CURRENT APPROACH

When an account is priced, certain characteristics about the account may be available, such as the industry or the state of operation. This information can be used to select the best exposure loss cost for the account, which is used as the a priori estimate for the account before considering the loss experience. The exposure loss cost can come from company data by analyzing the entire portfolio of accounts, from a large, external insurance services source, such as ISO or NCCI, from public rate filing information, from publicly available or purchased relevant data, or from judgment.

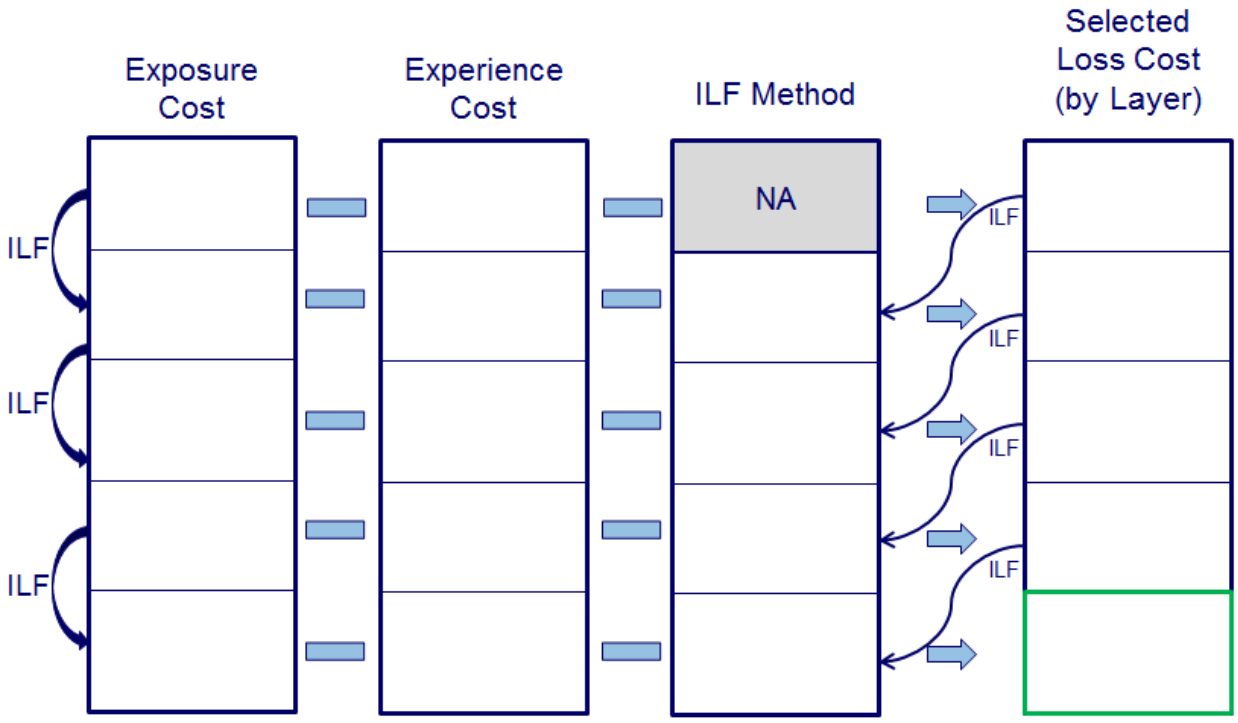
Very often, individual loss information is only available above a certain large loss threshold. Below this threshold, information is given in aggregate, which usually includes the sum of the total capped loss amount and the number of claims. More or less information may be available depending on the account. A basic limit is chosen, usually greater than the large loss threshold, as a relatively stable point in which to develop and analyze the account's losses. Once this is done, if the policy is excess or if the policy limit is greater than the basic limit, an ILF is applied to the basic limit losses to produce the loss estimate for the policy layer. It is also possible to look at the account's actual losses in the policy layer, or even below it but above the basic limit, which are known as the burn costs, as an another alternative estimate. The exposure cost is the most stable, but may be less relevant to a particular account. The loss experience is more relevant, but is usually more volatile, depending on the size of the account. The burn costs are the most relevant, but also the most volatile. Determining the amount of credibility to assign to each estimate can be difficult. Such an approach is illustrated in Figure 1 (where "Exper Cost" stands for the Experience Cost). The exact details pertaining to how the credibilities are calculated vary by practitioner.

Figure 1: Current Approach



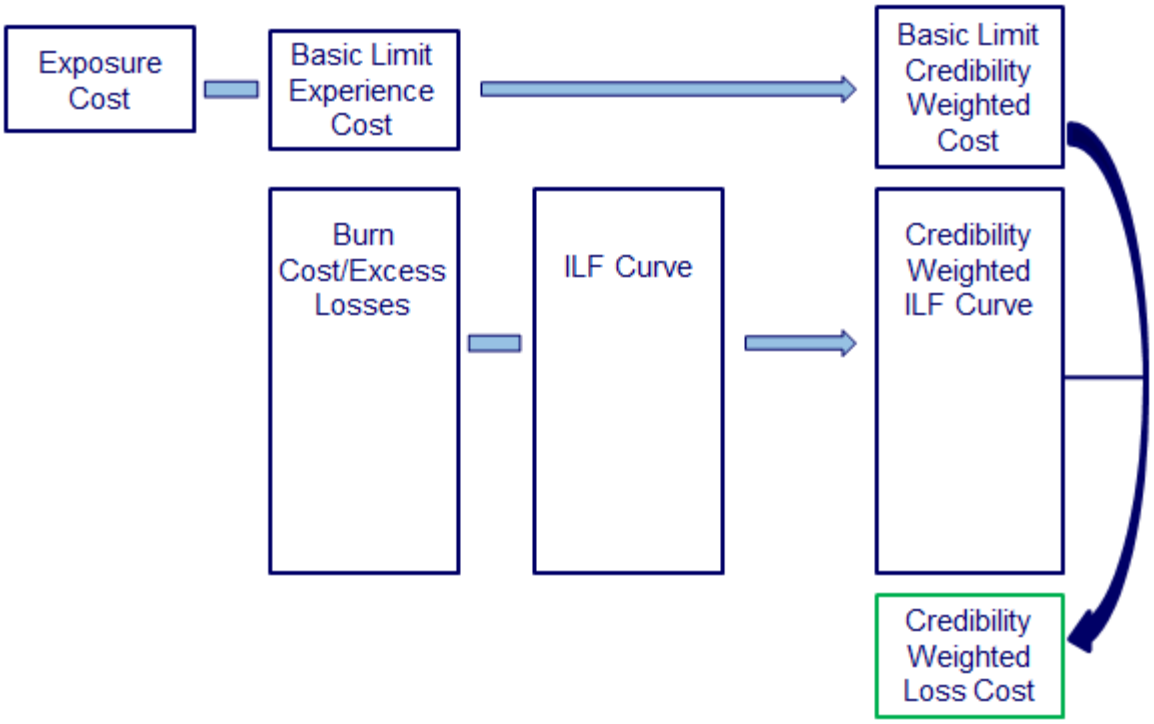
Clark (2011) developed a comprehensive approach to utilizing all of the data. For the basic limit, a selection is made based off of a credibility weighting between the exposure cost and the loss rating cost. For each excess layer, a credibility weighting is performed between the exposure cost multiplied by the appropriate ILF, the actual loss cost in the layer (i.e., the burn cost), and the previous layer's selection multiplied by the appropriate ILF. Formulas are shown for calculating all relevant variances and covariances, which are needed for estimating the optimal credibilities for each method in each layer. For further details on this method, refer to the paper. This approach is illustrated in Figure 2.

Figure 2: Clark's Method



The approach discussed in this paper is illustrated in Figure 3. It can be seen that all of the data that is used in Clark's approach is used here as well. Credibility weighting of the basic limit is expounded upon in the next section. Credibility weighting of the excess layers is discussed in section 4.

Figure 3: Proposed Method



3. CREDIBILITY WEIGHTING THE BASIC LAYER

Buhlmann-Straub credibility can be used to perform credibility weighting between the exposure cost and the account's actual losses in the basic layer. But account pricing is a bit different from the typical scenario of credibility weighting various segmentations in three ways:

1. Each item being credibility weighted has a different a priori loss cost (since the exposure costs can differ based on the class, etc.), that is, the complements are not the same. This also puts each account on a different scale. A difference of $1000 may be relatively large for one account, but not as large for another.
2. The expected variances differ between accounts since their losses may be capped at different amounts. The standard Buhlmann-Straub formulae assume that there is a fixed relationship between the variance and the exposures.
3. Additional information is available that can be used to improve the estimates in the form of exposure costs and ILF distributions, which can be used to calculate some of the expected values and variances.

Credibility can be performed either on the frequency and severity separately, or on the combined aggregate losses. The former will be discussed first. Accounting for trend and development is discussed at the end of the section. A related but off topic question of choosing the optimal capping point for the basic limit is discussed in Appendix D.

3.1 Separate Frequency and Severity

Splitting up frequency and severity often results in more robust estimates, although requires slightly more work than combining them. This section assumes that separate frequency and severity exposure estimates are available. If only a loss cost is available, the frequency can be calculated by dividing out the average capped severity using the appropriate ILF distribution (after removing legal expenses, if relevant). If various rating factors are applied to the initial loss cost estimate, for each one, it will need to be determined what percentage applies to the frequency versus the severity. Most factors are usually more related to frequency and so this can serve as the default assumption.

3.1.1 Frequency

It is usually assumed that the variance of the frequency is proportional to the mean (such as in Generalized Linear Models). Since the complements of credibility are different for each account, the expected variances are expected to differ as well, even for the same number of exposures. The Buhlmann-Straub within and between variance formulae assume a constant variance per number of exposures, but they can be modified to take this situation into account by dividing the variance component (that is, the square of the differences) by the frequency mean[2]. Doing this, the variances are calculated as a percentage of the expected frequency. The formulae are as follows:

$$\widehat{EPV} = \frac{\sum_{g=1}^{G} \sum_{n=1}^{N_g} e_{gn} \left(f_{gn} - \bar{f}_g \right)^2 / \bar{F}_g}{\sum_{g=1}^{G} \left(N_g - 1 \right)} \tag{3.1}$$

$$\widehat{VHM} = \frac{\sum_{g=1}^{G} e_g \left(\bar{f}_g - \bar{F}_g \right)^2 / \bar{F}_g - (G - 1) \widehat{EPV}}{e - \frac{\sum_{g=1}^{G} e_g^2}{e}} \tag{3.2}$$

[2] The proof of this formula is that it is similar to treating the number of exposures divided by the frequency as the weight in the formula, which is expected to be inversely proportional to the variance

Where *EPV* is the expected value of the process variance, or the "within variance", and *VHM* is the variance of the hypothetical means, or the "between variance". *G* is the number of groups, *N* is the number of periods, *e* is the number of exposures, f_{gn} is the frequency (per exposure) for group *g* and period *n*, $\bar{f}_g$ is the average frequency for group *g*, and $\bar{F}_g$ is the expected frequency for group *g* using the exposure costs. If the exposure frequency used comes from an external source, it can be seen that any overall error between it and the actual loss experience will increase the between variance and will thus raise the credibility given to the losses, which is reasonable. If this is not desired, the actual average frequency from the internal experience can be used instead in the formulae even if it is not used during the actual pricing. It can be seen that if the exposure frequency, $\bar{F}_g$, is the same for every account, these terms will cancel out in the resulting credibility calculations and the formulae will be identical to the original.

Equivalent formulae can also be used that utilize the actual claims counts instead of the frequency. These can be obtained by multiplying the numerator and denominator by the exposures.

$$\widehat{EPV} = \frac{\sum_{g=1}^{G} \sum_{n=1}^{N_g} (c_{gn} - \bar{c}_g)^2 / \bar{C}_g}{\sum_{g=1}^{G} (N_g - 1)} \tag{3.3}$$

$$\widehat{VHM} = \frac{\sum_{g=1}^{G} (\bar{c}_g - \bar{C}_g)^2 / \bar{C}_g - (G - 1)\,\widehat{EPV}}{e - \frac{\sum_{g=1}^{G} e_g^2}{e}} \tag{3.4}$$

Where *c* are the actual claim counts for the account and *C* are the claim counts using the exposure information for the account. This between variance can be calculated using a sample of actual accounts. The formula assumes that the between variance of accounts is proportional to the expected mean as well, which is a reasonable assumption. If the exposure costs used in the between variance formula (that is $\bar{F}$ or $\bar{C}$) do not utilize the same data being used to calculate the between variance (that is *f* or *c*), the bias correction component in the denominator can be removed and the formula becomes[3]:

[3] For a proof of this, refer to the appendix in Dean 2005

$$\widehat{VHM} = \frac{\sum_{g=1}^{G} e_g \, (\bar{f}_g - \bar{F}_g)^2 / \bar{F}_g - (G - 1) \, \widehat{EPV}}{e} \tag{3.5}$$

Once the within and between variances are calculated, the credibility assigned to an account can be calculated as normal:

$$k = \frac{\widehat{EPV}}{\widehat{VHM}} \tag{3.6}$$

$$Z = \frac{e}{e + k} \tag{3.7}$$

If only claims above a certain threshold are being considered, and this threshold can differ by account, then different formulae are needed and are shown below. These formulae work by calculating the within and between variances on the excess frequencies, but then converting them to ground up variances before combining them so that all variances are at the same level. A full explanation is shown in Appendix C. For these formulae to work, the frequencies should be expressed relative to one unit of exposure. So if, for example, the exposure unit is $100,000 of revenue, an account/year with $500,000 of revenue should be counted as 5 exposures. The survival probability at the threshold is shown as *p*.

$$\widehat{EPV} = \frac{\sum_{g=1}^{G} \sum_{n=1}^{N_g} \{ [e_{gn} (f_{gn} - \bar{f}_g \times p_{gn})^2 / (\bar{F}_g \times p_{gn}) - 1] / p_{gn} + 1 \}}{\sum_{g=1}^{G} (N_g - 1)} \tag{3.8}$$

$$\widehat{VHM} = \frac{\sum_{g=1}^{G} e_g \{ (\bar{f}_g - \bar{F}_g \times p_g)^2 / (\bar{F}_g \times p_g) - [(\frac{G - 1}{G} EPV - 1) \widehat{\times p_g} + 1] / e_g \} / p_g}{e - \frac{\sum_{g=1}^{G} e_g^2}{e}} \tag{3.9}$$

To calculate the credibility for an account, *k* can be calculated as below. Higher retentions (which have

lower *p* values) will result in higher values of *k*, and thus lower credibility values.

$$\mathrm{k} = \frac{(EPV - 1) \times p + 1}{VHM \times p} \tag{3.10}$$

3.1.2 Severity

Buhlmann-Straub credibility can be used to perform credibility weighting on the account's basic limit average severity as well. The common assumption for severity is that the standard deviation is proportional to the mean, or equivalently, that the variance is proportional to the mean squared (such as in Generalized Linear Models, for example). The formulae can be modified to calculate the variances as a percentage of the square of the expected capped severity. The Buhlmann-Straub formulae that account for this are below:

$$\widehat{EPV} = \frac{\sum_{g=1}^{G} \sum_{n=1}^{N_g} c_{gn} (s_{gn} - \bar{s}_g)^2 / \bar{S}_g^2}{\sum_{g=1}^{G} (N_g - 1)} \tag{3.11}$$

$$\widehat{VHM} = \frac{\sum_{g=1}^{G} c_g (\bar{s}_g - \bar{S}_g)^2 / \bar{S}_g - (G - 1)\widehat{EPV}}{c - \frac{\sum_{g=1}^{G} c_g^2}{c}} \tag{3.12}$$

Where c is the claim count, s_{gn} is the average severity for group g and period n, $\bar{s}_g$ is the average severity for group g across all years, and $\bar{S}_g$ is the expected average severity for group g using the ILF distribution. These formulae use the actual (i.e., undeveloped) claim count as the weights, which is appropriate as the variance of the average severity equals the variance of the severity divided by the claim count.

For these formulae to work, however, the capping point would need to be the same for every account. They also do not take advantage of all available information, as the ILF distribution can be used to estimate the expected volatility. Modified formulae are shown below. The derivation of these formulae is shown in Appendix A.

$$\widehat{EPV}_{g,cap} = \frac{LEV2(cap) - LEV(cap)^2}{\bar{S}^2} \tag{3.13}$$

$$\widehat{VHM} = \frac{\sum_{g=1}^{G} c_g \left[(\bar{s}_g - \bar{S}_g)^2 / \bar{S}_g^{\,2} - \frac{(G-1)\,\widehat{EPV}_{g,cap}}{G\,c_g} \right]}{c - \frac{\sum_{g=1}^{G} c_g^2}{c}} \tag{3.14}$$

Where *LEV(x)* is the limited expected value at *x*. A separate EPV is calculated for each account, while a common VHM is used across all accounts regardless of the expected severity or capping point. The credibility for each account can then be calculated as normal using the account's calculated EPV and the portfolio calculated VHM.

$$k = \frac{\widehat{EPV}_{g,cap}}{\widehat{VHM}} \tag{3.15}$$

$$Z = \frac{c}{c + k} \tag{3.16}$$

Note that the final credibility depends on the number of reported claims and the EPV, which depends on the capping point. Higher capping points will produce higher EPV values and thus will be assigned lower credibility and vice versa.

3.2 A Combined Loss Cost Approach

Sometimes, it may be more desirable to develop and perform credibility on the aggregate losses with the frequency and severity combined. The common assumption for aggregate losses is that the variance is proportional to the average taken to some power between one and two (as in the Tweedie distribution), although these equations are far less sensitive to the power used than in GLM modeling. A common assumption is to set this power to 1.67 (Klinker 2011). Alternatively, it may be an

acceptable simplification to assume that the standard deviation is roughly proportional to the mean and to use a power of two. The formulae for aggregate losses are as follows:

$$\widehat{EPV} = \frac{\sum_{g=1}^{G} \sum_{n=1}^{N_g} e_{gn} (l_{gn} - \bar{l}_g)^2 / \bar{L}_g^p}{\sum_{g=1}^{G} (N_g - 1)} \tag{3.17}$$

$$\widehat{VHM} = \frac{\sum_{g=1}^{G} e_g (\bar{l}_g - \bar{L}_g)^2 / \bar{L}_g^p - (G - 1) \widehat{EPV}}{e - \frac{\sum_{g=1}^{G} e_g^2}{e}} \tag{3.18}$$

Where l_{gn} is the loss cost (per exposure) for group g and period n, $\bar{l}_g$is the average loss cost for group g, $\bar{L}_g$ is the expected exposure loss cost for group g, and p is the Tweedie power used.

Similar to severity, this within variance formula does not handle different capping points and also does not leverage information from the severity distribution. Modified formulae are shown below. The derivation is shown in Appendix B.

$$\widehat{EPV}_{g,cap} = \frac{[LEV2(cap) - LEV(cap)^2] \times \bar{F} + \widehat{EPV}_f \times \bar{F} \times \bar{S}^2}{\bar{L}^P} \tag{3.19}$$

$$\widehat{VHM} = \frac{\sum_{g=1}^{G} e \left[(\bar{l}_g - \bar{L}_g)^2 / \bar{L}_g^{\ P} - \frac{(G - 1) \widehat{EPV}_{g,cap}}{G\, e_g} \right]}{e - \frac{\sum_{g=1}^{G} e_g^2}{e}} \tag{3.20}$$

Similar to the above, if the loss costs come from an external source, the bias correction in the denominator can be removed. The credibility can now be calculated as normal. Similar to severity, higher loss caps will result in less credibility being assigned and vice versa.

3.3 Accounting for Trend and Development

Accounting for trend is relatively straightforward. All losses should be trended to the prospective year before all of the calculations mentioned above. The basic limit as well as the large loss threshold are trended as well, with no changes to procedure due to credibility weighting.

To account for development, a Bornhuetter-Ferguson method should not be used since it pushes each year towards the mean and thus artificially lowers the volatility inherent in the experience. Instead, a Cape Cod-like approach can be used, which allows for a more direct analysis of the experience itself. This method compares the reported losses against the "used" exposures, which results in the chain ladder estimates for each year, but the final result is weighted by the used exposures, which accounts for the fact that more volatility is expected in the greener years (Korn 2015a).

For frequency, the development factor to apply to the claim counts and the exposures is the claim count development factor. For severity, the actual claim count should be used since these are the exposures for the current estimate of the average severity. The actual average severity still needs to be developed though, since it has a tendency to increase with age. Severity development factors can be calculated by dividing the loss development factors by the claim count development factors (Siewert 1996), or the severity development can be analyzed directly to produce factors. The total exposures for each group should be the sum of the used exposures across all years.

4. CONSIDERING THE EXCESS LOSSES

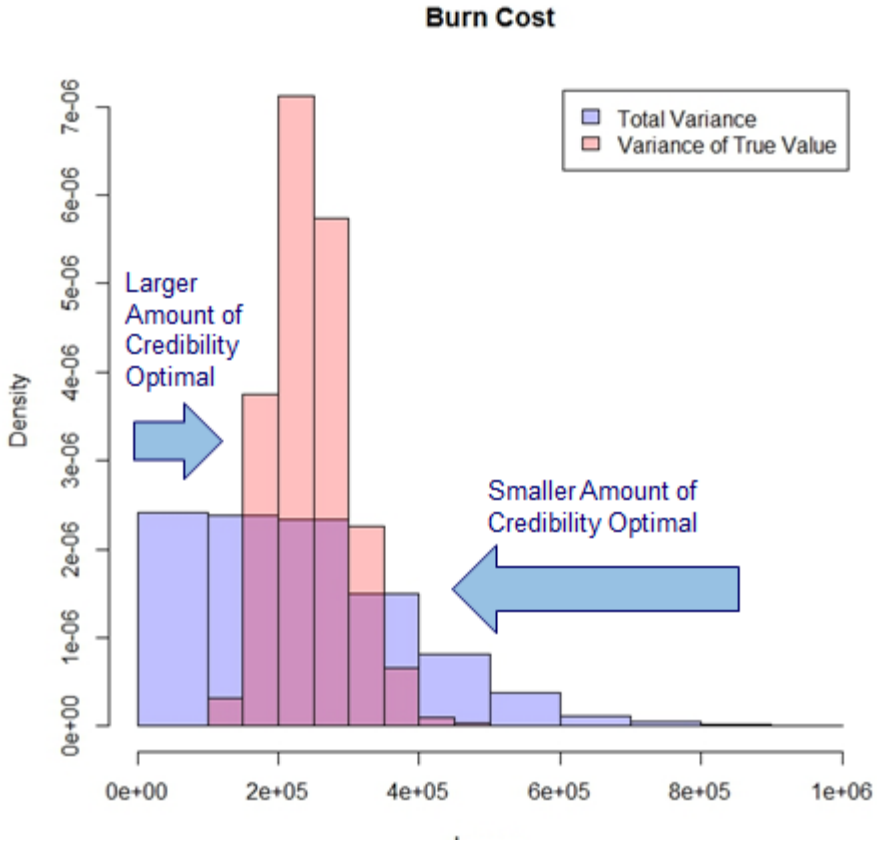
4.1 Introduction

Another source of information not considered in the basic layer losses are the excess losses, that is, the losses greater than the basic limit. These losses can be used to calculate the burn cost in each excess layer above the basic limit. After applying the appropriate ILF to each, if relevant, these values can serve as alternative loss cost estimates as well. In this type of approach, each of these excess layers needs to be developed separately, and credibility needs to be determined for each, which can be cumbersome. Calculating the credibility of each method in each layer requires the calculation of each variance as well as all of the correlations between them.

Burn costs are also right skewed, which do not perform well with linear credibility methods, as mentioned. To get a sense of why this is so, consider Figure 4, which shows the distribution of the burn cost in a higher layer produced via simulation. The majority of the time, the burn cost is only slightly lower than the true value (the left side of the figure). A smaller portion of the time, such as when there has been a large loss, the burn cost is much greater than the true value (the right side of

the figure). For cases where the burn cost is lower than the true value and not that far off, a larger amount of credibility can be assigned to the estimate on average than when it is greater that the true value and is very far off. That is why linear credibility methods that assign a single weight to an estimate do not work well in this case.

Figure 4: Example of a Burn Cost Distribution



As an alternative, instead of examining the burn costs directly, the excess losses can be leveraged to modify the severity distribution that is used to calculate the increased limit factor. Such an approach considers all available information just as the direct burn cost approach does. It is also more robust as mentioned.

This remainder of this section discusses an implementation of this method and addresses various potential hurdles.

4.2 Method of Fitting

The first question to consider is what is the best fitting method when only a small number of claims, often only in summarized form, are available To answer this question a simulation was performed with only 25 claims and a large loss threshold of 200 thousand. See the following footnote for more details on the simulation[4]. For the maximum likelihood method, the full formula from section 4.8 was used but without the credibility component, which is discussed later. The bias and root mean square error (RMSE) was calculated by comparing the fitted limited expected values against the actual. The results are shown below in Table 1.

Table 1: Performance of Different Fitting Techniques

Method	Bias	RMSE (Thousands)
MLE	4.7%	194
CSP Error Squared	16.5%	239
CSP Error Percent Squared	13.5%	243
CSP Binomial	8.9%	209
LEV Error Percent Squared	55.2%	282
Counts Chi-Square	41.4%	256

CSP stands for conditional survival probability. The methods that utilized this sought to minimize either the squared error or the square of the percentage error, or performed MLE on these probabilities using a binomial distribution. Another method sought to minimize the squared percentage errors of the fitted and actual LEVs. The final method shown looked at the number of excess claims in each layer and sought to minimize the chi-squared statistic. It can be seen that the maximum likelihood has both the lowest bias and the lowest root mean square error (RMSE). (Using credibility will reduce the bias further to small amounts, as is shown with the simulated results in

[4] A lognormal was simulated with mean mu and sigma parameters of 11 and 2.5, respectively. The standard deviation of the parameters was 10% of the mean values. The policy attachment point and limit was both 10 million.

section 4.10.) It is also the most theoretically sound and the best for incorporating credibility, as is explained in the following section. For all of these reasons, maximum likelihood is used as the fitting method for the remainder of this paper.

Before deriving the likelihood formula for aggregate losses, first note that instead of applying an ILF to the basic limit losses, it is also possible to simply multiply an account's estimated ultimate claim count by the limited average severity calculated from the same severity distribution. The advantage of using an ILF is that it gives credibility to the basic limit losses, as shown below, where N is the estimated claim count for the account and $LEV(x)$ is the limited expected value calculated at x:

$$\begin{aligned} & Capped\ Losses\ \times ILF(Policy\ Layer) \\ =\ & N\ \times LEV_{Account}\ (Loss\ Cap)\ \times \frac{LEV_{Portfolio}(Policy\ Layer)}{LEV_{Portfolio}(Loss\ Cap)} \\ =\ & N\ \times LEV_{Portfolio}\ (Policy\ Layer)\ \times \frac{LEV_{Account}\ (Loss\ Cap)}{LEV_{Portfolio}\ (Loss\ Cap)} \end{aligned} \tag{4.1}$$

So applying an ILF is the same as multiplying an account's claim count by the portfolio estimated limited expected value at the policy layer, multiplied by an experience factor of the account's actual capped losses divided by the expected capped loss. This last component gives (full) credibility to the account's experience.

If individual claim data is only available above a certain threshold, which is often the case, there are three pieces of information relevant to an account's severity: the sum of the capped losses, the number of losses below the large loss threshold, and the number and amounts of the losses above the threshold. If the ILF method is used, the first component is already accounted for, and so only the two latter items should be considered[5]. (Including the first component when using an ILF actually produces slightly worse estimates, as is shown in the simulation results in section 4.10.) The claims below the threshold are left censored (as opposed to left truncated or right censored, which actuaries are more used to), since we are aware of the presence of each claim but do not know its exact value, similar to the effect of a policy limit. Maximum likelihood estimation can handle left censoring similar to how it handles right censoring. For right censored data, the logarithm of the survival function at

[5] Note that even though there may be some slight correlation between the sum of the capped losses and the number of claims that do not exceed the cap, as mention by Clark (2011), these are still different pieces of information and need to be accounted for separately.

the censoring point is added to log-likelihood. Similarly, for a left censored point, the logarithm of the cumulative distribution function at the large loss threshold is added to the log-likelihood. This should be done for every claim below the large loss threshold and so the logarithm of the CDF at the threshold should be multiplied by the number of claims below the threshold. Expressed algebraically, the formula for the log-likelihood is:

$$\sum_{x=Claims>LLT} PDF(x) + n \times CDF(LLT) \tag{4.2}$$

Where *LLT* is the large loss threshold, *PDF* is the logarithm of the probability density function, *CDF* is the logarithm of the cumulative density function, and n is the number of claims below the large loss threshold. The number of claims used in this calculation should be on a loss-only basis and claims with only legal payments should be excluded from the claim counts, unless legal payments are included in the limit and are accounted for in the ILF distribution. If this claim count cannot be obtained directly, factors to estimate the loss-only claim count will need to be derived for each duration.

As an example, assume that an account had a total of ten claims, three of which were above the large loss threshold of $100 thousand with the following values: $200 thousand, $500 thousand, and one million. The log-likelihood using a lognormal distribution for mu and sigma parameters of 10 and 2 would equal the following:

log(lognormal-pdf(200000, 10, 2)) + log(lognormal-pdf(500000, 10, 2))

+ log(lognormal-pdf(1000000, 10, 2)) + 7 x log(lognormal-cdf(100000, 10, 2))

Where *lognormal-pdf(a, b, c)* is the lognormal probability density function at *a* with mu and sigma parameters of *b* and *c,* respectively, and *lognormal-cdf(a, b, c)* is the lognormal cumulative density function at *a* with mu and sigma parameters of *b* and *c,* respectively.

4.3 Method of Credibility Weighting

Bayesian credibility will be used to incorporate credibility into the severity fit. This method performs credibility on each of the distribution parameters simultaneously while fitting the distribution and so is optimal to another approach that may attempt to credibility weight already fitted parameters. This method can be implemented without the use of specialized software. The distribution of

maximum likelihood parameters is assumed to be approximately normally distributed. A normally distributed prior distribution will be used (which is the complement of credibility, in Bayesian terms), which is the common assumption. This is a conjugate prior and the resulting posterior distribution (the credibility weighted result, in Bayesian terms) is normally distributed as well. Maximum likelihood estimation (MLE) returns the mode of the distribution, which will also return the mean in the case, since the mode equals the mean for a normal distribution. So, this simple Bayesian credibility model can be solved using just MLE (Korn 2015b). It can also be confirmed that the resulting parameter values are almost identical whether MLE or specialized software is used.

To recap, the formula for Bayesian credibility is f(Posterior) ~ f(Likelihood) x f(Prior), or f(Parameters | Data) ~ f(Data | Parameters) x f(Parameters). When using regular MLE, only the first component, the likelihood, is used. Bayesian credibility adds the second component, the prior distribution of the parameters, which is what performs the credibility weighting. The prior used for each parameter will be a normal distribution with a mean of the portfolio parameter. The equivalent of the within variances needed for the credibility calculation to take place are implied automatically based on the shape of the likelihood function and do not need to be calculated, but the between variances do, which is discussed in section 4.4. This prior log-likelihood should be added to the regular log-likelihood. The final log-likelihood formula for a two parameter distribution that incorporates credibility is as follows:

$$\sum_{x=Claims>LLT} PDF(x,p1,p2) \; + \; n \times CDF(LLT,p1,p2) \; + \; Norm(p1, Porfolio\; p1, Between\; Var1) \; + \; Norm(p2, Porfolio\; p2, Between\; Var2) \tag{4.3}$$

Where *PDF(x, p1, p2)* is the logarithm of the probability density function evaluated at *x* and with parameters, *p1* and *p2*; *CDF(x, p1, p2)* is the logarithm of the cumulative density function evaluated at *x* and with parameters, *p1* and *p2*; and *Norm(x, p, v)* is the logarithm of the normal probability distribution function evaluated at *x*, with a mean of *p*, and a variance of *v*. *Portfolio p1* and *Portfolio p2* are the portfolio parameters for the distribution and *Between Var 1* and *Between Var 2* are the between variances for each of the portfolio parameters.

Using the same example from the previous section and assuming that the mu and sigma parameters for the portfolio are 11 and 3 with between variances of 1 and 0.5, respectively, the log-likelihood at mu and sigma parameters of 10 and 2 would be equal to the following:

$$\log(\text{lognormal-pdf}(200000, 10, 2)) + \log(\text{lognormal-pdf}(500000, 10, 2))$$
$$+ \log(\text{lognormal-pdf}(1000000, 10, 2)) + 7 \times \log(\text{lognormal-cdf}(100000, 10, 2))$$
$$+ \log(\text{normal-pdf}(10, 11, 1)) + \log(\text{normal-pdf}(2, 3, 0.5))$$

Where *normal-pdf(a, b, c)* is the probability density function of a normal distribution at *a* with a mean of *b* and variance of *c*.

4.4 Calculating the Between Variance of the Parameters

Calculation of the variances used for the prior distributions can be difficult. The Buhlmann-Straub formulae do not work well with interrelated values such as distribution parameters. MLE cannot be used either as the distributions of the between variances are usually not symmetric and so the mode that MLE returns is usually incorrect and is very often at zero. A Bayesian model can be built, but this requires a specialized expertise that not everyone has and will not be discussed here. Another technique is to use a method similar to ridge regression which estimates the between variances using cross validation.

This method is relatively straightforward to explain and is quite powerful as well[6]. Possible candidate values for the between variance parameters are tested and are used to fit the severity distribution for each risk on a fraction of the data, and then the remainder of the data is used to evaluate the resulting fitted distributions. The between variance combination with the highest out-of-sample total likelihood is chosen. The calculation of the likelihood on the test data should not include the prior/credibility component. The fitting and testing for each set of parameters should be run multiple times until stability is reached, which can be verified by graphing the results. The same training and testing samples should be used for each set of parameters as this greatly adds to the stability of this approach. Simulation tests using this method (with two thirds of the data used to fit and the remaining one third to test) on a variety of different distributions are able to reproduce the actual between variances on average, which shows that the method is working as expected. Repeated n-fold cross validation can be used as well, but will not be discussed here.

4.5 Distributions with More Than Two Parameters

If the portfolio distribution has more than two (or perhaps three) parameters, it may be difficult to apply Bayesian credibility in this fashion. The method can still be performed as long as two

[6] One advantage of this approach over using a Bayesian model is that this method works well even with only two or three groups, whereas a Bayesian model tends to overestimate the prior variances in these cases. Though not relevant to this topic, as many accounts should be available to calculate the between variances, this is still a very useful method in general for building portfolio ILF distributions.

"adjustment parameters" can be added that adjust the original parameters of the severity distribution. For a mixed distribution, such as a mixed exponential or a mixed lognormal, one approach is to have the first adjustment parameter apply a scale adjustment, that is, to modify all claims by the same factor. The second adjustment parameter can be used to shift the weights forwards and backwards, which will affect the tail of the distribution if the individual distributions are arranged in order of their scale parameter. To explain the scale adjustment, most distributions have what is known as a scale parameter which can be used to adjust all claims by the same factor. For the exponential distribution, the theta parameter is a scale parameter, and so multiplying this parameter by 1.1, for example, will increase all claim values by 10%. For the lognormal distribution, the mu parameter is a log-scale parameter, and so to increase all claims by 10%, for example, the logarithm of 1.1 can be added to this parameter. For a mixed distribution, the scale parameter of each of the individual distributions should be adjusted.

One way to implement this is as follows, using the mixed exponential distribution as the example:

$$\theta_i' = \theta_i \times exp(Adj1) \tag{4.4}$$

$$R_i = W_i \times exp(i \times Adj2) \tag{4.5}$$

$$W_i' = R_i \,/\, \sum R \tag{4.6}$$

Where *Adj1* and *Adj2* are the two adjustment parameters, *i* represents each individual distribution within the mixed exponential ordered by the theta parameters, *R* is a temporary variable, and *W* are the weights for the mixed distribution. Adjustment parameters of zero will cause no change, positive adjustment parameters will increase the severity, and negative adjustment parameters will decrease the severity.

4.6 Separate Primary and Excess Distributions

Sometimes a separate severity distribution is used for the lower and upper layers and they are then joined together in some fashion to calculate all relevant values. One way to join the distributions is to use the survival function of the upper distribution to calculate all values conditional on the switching point (that is, the point at which the first distribution ends and the second one begins), and then use the survival function of the lower distribution to convert the value to be unconditional again from ground up. The formulae for the survival function and for the LEV for values in the upper layer,

assuming a switching point of p are as follows:

$$S(x) = S_U(x) / S_U(p) \times S_L(p) \tag{4.6}$$

$$LEV(x) = [LEV_U(x) - LEV_U(p)] / S_U(p) \times S_L(p) + LEV_L(p) \tag{4.7}$$

Where U indicates using the upper layer severity distribution and L indicates using the lower layer severity distribution. More than two distributions can be joined together in the same fashion as well.

Using this approach, both the lower and upper layer severity distributions can be adjusted if there is enough credible experience in each of the layers to make the task worthwhile. When adjusting the lower distribution, values should be capped at the switching point (and the survival function of the switching point should be used in the likelihood formula for claims greater than this point). When adjusting the upper distribution, only claim values above the switching point can be used and so the data should be considered to be left truncated at this point. Even if no or few claims pierce this point, modifying the lower layer severity distribution still affects the calculated ILF and LEV values in the upper layer since the upper layer lies on top of the bottom one using this or a similar approach.

4.7 An Alternative when Maximum Likelihood Cannot be Used

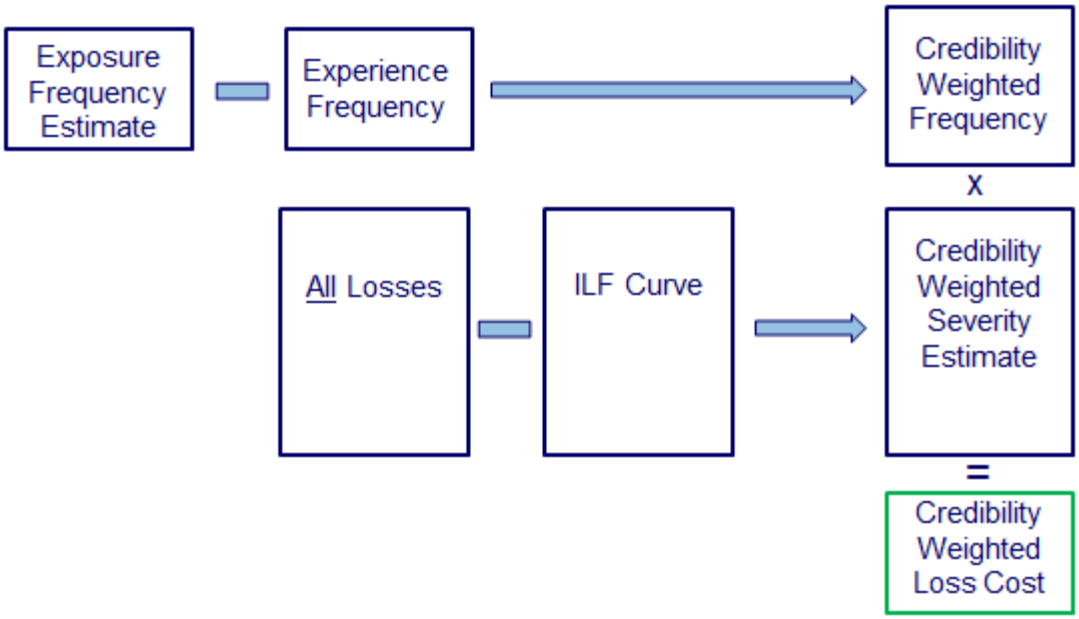
Depending on the environment a pricing system is implemented in, an optimization routine required to determine the maximum likelihood may be difficult to find. An alternative is to calculate the log-likelihood for all possible parameter values around the expected using some relatively small increment. The exponent of these log-likelihoods minus the maximum log-likelihood can then be calculated to produce likelihoods that do not round to zero. These can then be used as the relative weights, and a weighted average of the parameter values can be calculated. The result should be very close to the MLE estimate.

4.8 An Alternative that Involves Combining All Severity Information Together

Using the approach mentioned thus far, the basic limit average severity is credibility weighted using the Buhlmann-Straub method, either by itself or included with the frequency in the aggregate losses, and the excess losses are credibility weighted using the Bayesian method mentioned. It is possible to simplify this procedure and incorporate both the basic limit severity as well as the excess severity in the same step. This can be accomplished by including the average capped severity in the likelihood

formula used to modify the severity distribution. Then, instead of applying an ILF, the limited average severity in the policy layer can calculated from this credibility weighted severity distribution. Multiplying this by a (credibility weighted) frequency estimate produces the final result. (If only an exposure loss cost is available, this cost can be divided by the expected average severity in the layer to produce a frequency estimate.) This approach is illustrated below in Figure 5.

Figure 5: Proposed Approach Without a Basic Limit



Utilizing central limit theorem, it can be assumed that the average capped severity is approximately normally distributed. (Performing simulations with a small number of claims and a Box-Cox test justifies this assumption as well.) For very small number of claims, it is possible to use a Gamma distribution instead, although in simulation tests, using a Gamma does not seem to provide any benefit. The expected mean and variance of this normal or Gamma distribution can be calculated with the MLE parameters using the limited first and second moment functions of the appropriate distribution. The variance should be divided by the actual claim count to produce the variance of the average severity. For a normal distribution, these parameters can be plugged in directly; for a Gamma distribution, they can be used to solve for the two parameters of this distribution. The likelihood formula for this approach is as follows, including the credibility component:

$$\sum_{x=Claims>LLT} PDF(x,p1,p2) + n \times CDF(LLT,p1,p2) + Norm(Average\ Capped\ Severity,\mu,\sigma^2) + Norm(p1,Portfolio\ p1,Between\ Var1) + Norm(p2,Portfolio\ p2,Between\ Var2) \quad (4.8)$$

Where μ and σ^2 are calculated as:

$$\mu = LEV(Basic\ Limit,p1,p2)$$

$$\sigma^2 = [\ LEV2(Basic\ Limit,p1,p2) - LEV(Basic\ Limit,p1,p2)^2\ ] / m$$

Average Capped Severity is the average severity at the basic limit calculated from the account's losses, *n* is the number of claims below the large loss threshold, *m* is the total number of claims, and LEV2 is the second moment of the limited expected value. As above, *PDF*, *CDF*, and *Norm* are the logarithms of the probability distribution function, cumulative distribution function, and the normal probability density function respectively.

Using the example from sections 4.2 and 4.3 assuming that the average capped severity at $100 thousand is $70 thousand, the log-likelihood at mu and sigma parameters of 10 and 2 would be calculated as follows:

$$\mu = \text{lognormal-lev}(\ 100000,\ 10,\ 2\ )$$

$$\sigma^2 = [\ \text{lognormal-lev2}(\ 100000,\ 10,\ 2\ ) - \text{lognormal-lev}(\ 100000,\ 10,\ 2\ )^2\ ] / 10$$

$$\text{log-likelihood} = \log(\ \text{lognormal-pdf}(\ 200000,\ 10,\ 2\ )\ ) + \log(\ \text{lognormal-pdf}(\ 500000,\ 10,\ 2\ )\ )$$

$$+ \log(\ \text{lognormal-pdf}(\ 1000000,\ 10,\ 2\ )\ ) + 7 \text{ x } \log(\ \text{lognormal-cdf}(\ 100000,\ 10,\ 2\ )\ )$$

$$+ \log(\ \text{normal-pdf}(\ 70000,\ \mu,\ \sigma^2\ )$$

$$+ \log(\ \text{normal-pdf}(\ 10,\ 11,\ 1\ )\ ) + \log(\ \text{normal-pdf}(\ 2,\ 3,\ 0.5\ )\ )$$

Where *lognormal-lev(a, b, c)* is the lognormal limited expected value and *lognormal-lev2(a, b, c)* is the second moment of the lognormal limited expected value at *a* with mu and sigma parameters of *b* and *c,* respectively.

4.9 Accounting for Trend and Development

Both the losses and the large loss threshold should be trended to the prospective year before

performing any of the above calculations. Using the likelihood formulae above (4.3 and 4.8), it is possible to account for different years of data with different large loss thresholds by including the parts from different years separately. Or alternatively, all years can be grouped together and the highest large loss threshold can be used.

There is a tendency for the claim severity of each year to increase with time since the more severe claims often take longer to settle. The claims data needs to be adjusted to reflect this. A simple approach is to apply the same amount of adjustment that was used to adjust the portfolio data to produce the final ILF distribution, whichever methods were used. With this approach, the complement of credibility used for each account should be the severity distribution before adjustment, and then the same parameter adjustments that were used at the portfolio level can be applied to these fitted parameters.

Another simple method is to assume that severity development affects all layers by the same factor. (This is the implicit assumption if loss development factors and burn costs are used.) The severity development factor for each year can be calculated by dividing the (uncapped) LDF by the claim count development factor, or it can be calculated directly from severity triangles. Each claim above the large loss threshold as well as the threshold itself should then be multiplied by the appropriate factor per year before performing any of the credibility calculations mentioned. Many more methods are possible as well that will not be discussed here.

4.10 Simulation

A simulation was conducted to help demonstrate the benefit this method can provide even with only a small number of claims. The results are shown in Tables 2-6. Results of using aggregate claim data with the likelihood formulae discussed in this paper as well as using the individual claim data were both calculated. Both of the aggregate likelihood methods, with and without the basic limits portion, were used. The errors were calculated on the total estimated losses for the policy layer. Tables 2-4 show the results of using a lognormal severity distribution: the first shows a lower excess layer, the second shows a higher excess layer, and the last shows a primary layer with a self insured retention. Table 5 shows the results for a mixed exponential and Table 6 shows the results for a mixed lognormal. (Simulations were also conducted with Gamma and Pareto distributions as well with similar results, but are not shown here for the sake of brevity.) Refer to the following footnote for more details on how the simulation was conducted[7]. All simulations used only 25 ground up claims.

[7] For the lognormal distribution, mean mu and sigma parameters of 11 and 2.5 were used, respectively. The standard deviation as well as the prior standard deviation assumed was 10% of the mean parameter values. The large loss threshold was 200 thousand, which translated to an average of 8.1 claims above the threshold. For the mixed exponential, the

Table 2: Lognormal Distribution with Attachment Point and Limit of $2 Million

Method	Bias: LEV Method	RMSE: LEV Method (Millions)	RMSE Relative to Portfolio ILF Method	Bias: ILF Method	RMSE: ILF Method (Millions)	RMSE Relative to Portfolio ILF Method
Portfolio	0.0%	3.05	+48.2%	0.0%	2.06	0.0%
Account Only (Full Credibility)	-0.5%	1.89	-8.4%	-0.5%	1.91	-7.2%
Credibility - Individual Claims	1.3%	1.41	-31.6%	3.1%	1.48	-28.3%
Credibility - Aggregate, Including Capped Sum	-0.3%	1.43	-30.5%	2.8%	1.49	-27.7%
Credibility - Aggregate, NOT Including Capped Sum	2.1%	1.46	-29.2%	3.6%	1.47	-28.7%

following mean mu values were used: 2863.5, 22215.7, 89355.0, 266664.3, 1108333.2, 3731510.8, 9309907.8, 20249975.1, 51141863.9, 230000000.0 and the following weights were used: 0.378297, 0.327698, 0.19941, 0.080178, 0.012106, 0.001764, 0.000362, 0.000125, 0.000048, 0.000012. The large loss threshold was 30 thousand which translated to an average of 8 claims above the threshold. The standard deviation of the adjustment parameters was 1 and 0.5. For the mixed lognormal, the mu parameters were 8 and 12, the sigma parameters were 2.5 and 2.7, and the weights were 75% and 25%. The large loss threshold was 25 thousand, which translated to an average of 8.7 claims above the threshold. The standard deviation of the adjustment parameters was 1 and 0.5. Simulating with certain mean parameter values and standard deviations would result in an average policy layer LEV that differed from the LEV calculated from the mean parameters, and so using these mean parameter values as the complement of credibility would cause a bias. Using prior values that result from fitting all of the data together would also not be exact as the data from multiple draws of a certain distribution with different parameter values would not necessarily be a perfect fit to that distribution, and a bias would show up as well. Instead, the prior parameters used for credibility weighting were adjusted together so that the result from using the average LEV would be unbiased. This is only an issue for simulation and would not be an issue in practice.

Table 3: Lognormal Distribution With Attachment Point and Limit of $10 Million

Method	Bias: LEV Method	RMSE: LEV Method (Millions)	RMSE Relative to Portfolio ILF Method	Bias: ILF Method	RMSE: ILF Method (Millions)	RMSE Relative to Portfolio ILF Method
Portfolio	0.0%	5.73	+28.1%	0.0%	4.47	0.0%
Account Only (Full Credibility)	5.2%	4.70	+5.1%	5.2%	4.73	+5.8%
Credibility - Individual Claims	3.4%	3.07	-31.4%	5.8%	3.20	-28.5%
Credibility - Aggregate, Including Capped Sum	2.7%	3.09	-30.9%	6.3%	3.24	-27.6%
Credibility - Aggregate, NOT Including Capped Sum	4.9%	3.10	-30.8%	7.2%	3.19	-28.6%

Table 4: Lognormal Distribution With a Limit of $2 Million and an SIR of $50 Thousand

Method	Bias: LEV Method	RMSE: LEV Method (Millions)	RMSE Relative to Portfolio ILF Method	Bias: ILF Method	RMSE: ILF Method (Millions)	RMSE Relative to Portfolio ILF Method
Portfolio	0.0%	5.16	85.0%	0.0%	2.79	0.0%
Account Only (Full Credibility)	-0.9%	2.63	-5.7%	-0.9%	2.69	-3.4%
Credibility - Individual Claims	0.2%	2.17	-22.2%	1.2%	2.33	-16.6%
Credibility - Aggregate, Including Capped Sum	-1.7%	2.27	-18.7%	0.7%	2.35	-15.7%
Credibility - Aggregate, NOT Including Capped Sum	0.7%	2.36	-15.4%	1.3%	2.31	-17%

Table 5: Mixed Exponential Distribution With a Limit of $2 Million and an SIR of $100 Thousand

Method	Bias: LEV Method	RMSE: LEV Method (Millions)	RMSE Relative to Portfolio ILF Method	Bias: ILF Method	RMSE: ILF Method (Millions)	RMSE Relative to Portfolio ILF Method
Portfolio	0.0%	1,007	+28.0%	-0.5%	787	0.0%
Account Only (Full Credibility)	8.6%	817	+3.8%	10.2%	848	+7.8%
Credibility - Individual Claims	-0.3%	561	-28.6%	0.5%	584	-25.8%
Credibility - Aggregate, Including Capped Sum	-4.3%	582	-26.1%	-2.2%	599	-23.9%
Credibility - Aggregate, NOT Including Capped Sum	-1.6%	590	-24.9%	-1.1%	588	-25.3%

Table 6: Mixed Lognormal Distribution With an Attachment Point and Limit of $10 Million

Method	Bias: LEV Method	RMSE: LEV Method (Millions)	RMSE Relative to Portfolio ILF Method	Bias: ILF Method	RMSE: ILF Method (Millions)	RMSE Relative to Portfolio ILF Method
Portfolio	0.0%	3.53	+31.8%	-0.8%	2.68	0.0%
Account Only (Full Credibility)	-18.3%	2.73	+1.9%	-20.1%	2.70	+0.6%
Credibility - Individual Claims	3.7%	1.98	-26.2%	5.5%	2.09	-22.1%
Credibility - Aggregate, Including Capped Sum	1.5%	2.06	-23.0%	4.6%	2.17	-19.1%
Credibility - Aggregate, NOT Including Capped Sum	3.8%	2.08	-22.3%	5.2%	2.11	-21.2%

As the results show, this method is able to provide substantial benefit over the basic approach of applying an ILF to the capped loss estimate, even with only a small number of claims. The biases are very low as well. For the lognormal distributions, the sigma parameter was multiplied by $n / (n - 1)$, where n is the claim count, which is a well-known adjustment for reducing the MLE bias of the normal and lognormal distributions[8]. (The biases are slightly larger for the higher excess accounts, but this is within an acceptable range for these layers, given the high estimation volatility.) To further reduce the bias, it is possible to conduct a simulation to estimate the approximate bias factor and then divide out the bias factor from each account's loss cost estimate, although this should not be necessary most of the time.

As expected, the LEV method (with the aggregate losses) is able to perform better than the ILF method (without the aggregate losses), since it also takes into account the credibility of the basic limit losses. Also, the LEV method performs best when taking into account the basic limit losses since more information is being included. The ILF method performs better when this is not included, since this information is already captured from applying an ILF, and including it in the likelihood double counts this piece of information. Although, the difference is not drastic.

5. USING EXTREME VALUE THEORY

A method of estimating the excess severity potential was illustrated in the previous section whereby an account's losses are credibility weighted against the portfolio loss distribution. Normally, extrapolating a severity distribution past the range of available data is not recommended (unless the distribution is a distribution that allows extrapolating such as the commonly used single parameter Pareto). But in this case, it is justified even if the account's losses are well below the policy layer since the portfolio's loss data is being considered as well, which hopefully contains some losses near the layer being estimated. However, using an account's losses to predict the expected severity of a higher excess layer would not be recommended if full credibility was being assigned to the account's losses.

An alternative which may yield more accurate results is to work with a Generalized Pareto Distribution (GPD), which is the statistically recommended method of predicting severity potential in excess of the available data. Based on the Peak Over Threshold method of Extreme Value Theory, excess severity potential can be estimated by fitting a GPD to the loss data above a certain threshold.

[8] This adjustment cancels out most of the negative parameter bias. In this case, not applying this adjustment would have probably resulted in a lower overall bias. The positive part of the bias comes from the transformation involved in the LEV or ILF calculation, since even if the parameter mean value estimates are unbiased, applying a function to these estimates can create some bias, as Jensen's inequality states. As long as the parameter errors are not too great, the bias will remain small. The credibility weighting being performed reduces the parameter errors, and as a result, the bias as well.

(See McNeil 1997 for application to estimating loss severity.)

According to the theory, a GPD will better fit the data that is further into the tail, and so a higher threshold may provide a better fit. But there is a tradeoff since selecting a higher threshold will make less data available to analyze, which will increase the prediction variance. Looking at graphs of fitted versus empirical data is the typical way to analyze this trade off and to select a threshold. Although other methods are available. (See Scarrott & MacDonald 2012 for an overview.)

Returning to account rating, discarding losses below a certain threshold has an intuitive appeal as well. It is often debated how relevant smaller losses are to the severity potential of high excess layers. Using this method can help provide a statistical framework for evaluating the most relevant data to use for predicting expected excess loss potential. As to whether this method can be used in practice, testing a bunch of individual accounts in different commercial lines of business, the GPD seemed to provide a good fit to the accounts' losses above a certain threshold, even where a GPD may not be the ideal loss distribution for the portfolio.

To fit the GPD, the likelihood formula shown above should not be used, as the likelihood is simply the probability density function. Setting the threshold parameter to the appropriate value already takes the left truncation of the data into account, that is, that no losses below the threshold are being included. The fitted distribution will be conditional on having a claim of at least the threshold. Multiplying the calculated severity at the policy layer obtained from the fitted GPD by the expected excess claim count at the threshold will yield the loss cost. Formulas 3.8-3.10 shown above for excess frequency can be used to credibility weight the excess claim count estimate as well.

This method can be implemented using any type of distribution (or distributions, as explained in section 4.6) for the portfolio. Recall that Bayes' formula is being used for credibility: f(Parameters | Data) = f(Data | Parameters) x f(Parameters). Credibility is performed by calculating the prior likelihood on the parameters. It is also possible to reparameterize the distribution and use other new parameters instead. In this case, the logarithm of the instantaneous hazard function (that is, $f(x) / s(x)$) will be used for the new parameters, the same number as the number of parameters in the distribution. These are used since they are approximately normally distributed, work well, and are also not dependent on the selected threshold since they are conditional values. Using these values, it is possible to solve for the original distribution parameters since there are the same number of unknowns as equations. Then the original distribution parameters can be used to calculate any value from the distribution, such as PDFs and CDFs. Since this is the case, that any distribution value can be calculated from these new parameters, they can be thought of as the new parameters of the distribution, and the prior likelihood can be calculated on these new parameters instead. As a trick, instead of actually solving for the original parameters, we can effectively "pretend" that they were

solved for by still using the original parameters as the input to the maximization routine but just calculating the prior likelihood on the new parameters, that is the instantaneous hazard values, since the results will be exactly the same. In practice, we suggest using the difference in the hazard functions for each addition parameter since it makes the parameters less correlated and seems to work better in simulation tests. In summary, the likelihood equation is as follows, assuming a two parameter distribution:

$$\begin{aligned} p1 &= log(\, f(t_1) \,/\, s(t_1)\,) \\ p2 &= log(\, f(t_1) \,/\, s(t_1) \;-\; f(t_2) \,/\, s(t_2)\,) \\ loglik &= \sum_i GPD(\, x_i, \alpha, \beta, threshold\,) \;+\; N(\, p1, h\,\alpha v1\,) \;+\; N(\, p2, h\,\alpha v2\,) \end{aligned} \tag{5.1}$$

Where *GPD* is the logarithm of the PDF of the GPD distribution, *N* is the logarithm of the normal PDF, t_1 and t_2 are the two points chosen to calculate the instantaneous hazards, x are the claim values, α and β are the fitted GPD parameters, *threshold* is the selected threshold, *h1* and *h2* are the credibility complements for the logarithm of the hazard functions from the portfolio, and *v1* and *v2* are the between variances for the complements.

The complement of each of these new parameters can be calculated from the portfolio distribution, even if it is not a GPD, and the between variances can be solved for in the same manner as described above in section 4.4. (This method is also helpful in allowing the threshold of the GPD to vary.)

Simulations were conducted using this method simulating from a lognormal distribution[9]. The results are shown below in Table 7.

Table 7: GPD performance for various layers

Layer (Limit xs Retention, In Millions)	Bias	Improvement in RMSE (From Using Portfolio Severity Estimate)
10 xs 10	+3.5%	34.9%
25 xs 25	+3.6%	31.3%
50 xs 50	+6.0%	28.3%

[9] Claims were simulated from a lognormal distribution with parameters 11 and 2.5, respectively. 1000 iterations were performed. The between standard deviation of the lognormal parameters was 1 and 0.5, respectively. This was used to calculate the between standard deviations of the transformed parameters, that is, the hazard functions. 50 claims were simulated, a threshold of 250 thousand was used, and there was an average of 14.5 claims above the threshold.

It is difficult to say how this method will perform relative to the more basic method explained above. This is an area for further research.

6. CONCLUSION

This paper discussed an alternative technique to considering all relevant information to produce the most accurate estimate of an account's loss expectation. In doing so, an actuary pricing an account or treaty can be confident that the most relevant and stable estimates are being used.

Appendix A

To derive the Buhlmann-Straub formulae for average capped severity, we first note that to calculate the variance of the average severity using the within variance formula shown above, the EPV needs to be divided by the claim count and then multiplied by the severity squared. (This can be seen from the fact that formula 3,11 multiplies by the claim counts as weights, but does not divide by them afterward.)

$$Var(Average\ Severity) = \frac{EPV \times \bar{S}^2}{c}$$

The formula for the variance of the average severity using the severity distribution is:

$$Var(Average\ Severity) = \frac{LEV2(cap) - LEV(cap)^2}{c}$$

Where LEV(x) is the limited expected value capped at x and LEV2 is the second moment of the limited expected value. Setting these two equations equal to each other and solving for the EPV produces the following:

$$EPV_{g,cap} = \frac{LEV2(cap) - LEV(cap)^2}{\bar{S}^2}$$

This EPV can be used in the VHM (between variance) formula instead of using the traditional Buhlmann-Straub EPV formula.

Appendix B

To derive the formula for the EPV for aggregate losses, we first need to derive the formula for the variance of the loss cost estimate. To calculate the variance of the loss cost, we first separate the frequency and severity components. The expected variance of the frequency is:

$$Var(f) = \frac{EPV_f}{e} \times \bar{F}$$

Where EPV_f is the within variance parameter for the frequency. We needed to divide by the exposures since, as explained by severity, the EPV formula multiplies by the exposures as weights but does not divide by them afterwards. We then multiplied by the expected frequency since the within variance formula we used was calculated as a percentage of the expected frequency. To get the expected variance of c, the claim count, this quantity needs to be multiplied by the square of exposures (since the claim count equals the frequency multiplied by the exposures):

$$Var(c) = \frac{EPV_f}{e} \times \bar{F} \times e^2 = EPV_f \times \bar{C}$$

Where $\bar{C}$ is the claim count expected using the exposure frequency.

The variance of the severity equals:

$$Var(s) = LEV2(cap) - LEV(cap)^2$$

Using the formula for aggregate variance, the variance of the aggregate losses can be calculated as:

$$Var(a) = [LEV2(cap) - LEV(cap)^2] \times \bar{C} + EPV_f \times \bar{C} \times \bar{S}^2$$

Where *a* are the aggregate losses. The variance of the loss cost per unit of exposure equals this divided by the exposures squared, which produces the following:

$$Var(l) = \frac{[LEV2(cap) - LEV(cap)^2] \times \bar{F} + EPV_f \times \bar{F} \times \bar{S}^2}{e}$$

This is the variance of the loss cost estimate. This will be used to back into the EPV parameter needed in the between variance formula. The variance of the loss cost (divided by the exposures) can also be written as the following, using the EPV parameter:

$$Var(L) = \frac{EPV_l}{e} \times L^p$$

Where EPV_t is the EPV of the loss cost. Setting these two equations equal to each other produces the following for the EPV:

$$EPV_{cap} = \frac{[LEV2(cap) - LEV(cap)^2] \times \bar{F} + EPV_f \times \bar{F} \times \bar{S}^2}{L^p}$$

Appendix C

To be able to convert the process variance from ground up to excess and vice versa, we first need to derive the formula for the excess variance-to-mean ratio relative to the ground up variance-to-mean ratio. To do this, the formula for aggregate variance can be used, where *G* is the aggregate cost, *N* is the claim count, and *X* is the severity: $V[G] = V[N]\ E[X]^2 + V[X]\ E[N]$.

In this case where the aggregate is the excess claim count, the severity is one if the claim pierces the threshold and zero otherwise. This severity follows a Bernoulli distribution which has a variance equal to p x (1 - p), where p is the probability of exceeding the threshold, which is the survival probability. Performing some algebra and then dividing by the excess frequency (E[N] x p) in order to derive the excess variance-to-mean ratio, produces the following[10]:

$$E[X] = p$$

$$V[X] = p\ (1 - p)$$

$$E[G] = Np$$

$$V[G] = V[N]\ p^2 + p\ (1 - p)\ N$$

$$XS\ VTM = V[G] \ / \ E[G] = V[N]\ p \ / \ N + 1 - p = GU\ VTM \times p + 1 - p$$

$$XS\ VTM = (GU\ VTM - 1) \times p + 1$$

Where *XS VTM* is the excess variance-to-mean ratio and *GU VTM* is the ground up variance-to-mean ratio. This is the formula that will be used for converting the process variance from ground up to excess.

To convert the between variance, the variance should be multiplied by the square of p, following the formula for the variance of the product of a constant (the expected value of p) and a random variable. To convert a between variance-to-mean ratio (as is used in the formulae of this paper), the numerator is multiplied by the square of p and the denominator is multiplied by p, and so the ratio needs to be multiplied by p.

These results can now be plugged into the Buhlmann-Straub formulae, which produces the results shown in the paper. The *EPV* formula works by calculating the excess variance-to-mean ratios, and then converting each to a ground up variance-to-mean ratio so that they are compatible before

[10] Thanks to Aaron Curry for helping me derive this formula

summing them together. The resulting *EPV* is then a ground up value. For the *VHM* formula, first the squared differences of the excess frequency are calculated. The *EPV*, which is the process variance, is converted to an excess EPV and then subtracted out from the total variance, which leaves over the squared differences related to the (excess) between variance. The results are then converted to ground up values by dividing by the survival probabilities so that they are compatible and are then combined as normal as in the original formula. The k credibility ratio can then be calculated for an excess threshold by converting the *EPV* and *VHM* values from ground up to excess and then using the original formula on these values[11].

[11] A simulation was also performed to double check the formulae and it was confirmed that the *EPV* and *VHM* values calculated were equivalent on average whether calculated from the ground up frequencies using the original formulae or from the excess frequencies using the revised formulae. It was also confirmed that the credibilities being calculated for the excess frequencies were optimal, and that either increasing or decreasing the credibility resulted in larger average errors from the known true values.

Appendix D

A related issue to loss rating credibility is selecting the most optimal basic limit to develop and multiply by an ILF. Choosing a lower basic limit helps the capped loss cost estimate be more stable, but involves the application of a more leveraged increased limit factor and also uses less of an account's individual experience (unless credibility weighting is being performed). The variance formulae and ideas discussed in this paper can be used to calculate the expected volatility of each capping point, and the capping point with the lowest volatility can then be chosen as the optimal.

The following formula can be used to calculate the variance of the loss cost estimate for the basic layer. (This formulae was discussed in Appendix B.)

$$Var(a) = [LEV2(cap) - LEV(cap)^2] \times \bar{C} + EPV_f \times \bar{C} \times \bar{S}^2$$

Note that the actual variance of an account's experience is not used, as this would be subject to a large amount of error due to volatility. Applying credibility as discussed will decrease this variance. Using the formula for a normal conjugate prior as an approximation, the inverse of the final variance equals the sum of the inverse of the within variance and the inverse of the between variance. (Bolstad 2007)

The variance of the ILF should include both the parameter estimation error as well as the variance of the differences between accounts to reflect the fact that a portfolio ILF may be less relevant for a specific account. The parameter uncertainty of the estimated parameters can be calculated by taking the matrix inverse of the Fisher information matrix, which contains the second derivatives of the likelihood. Most statistical packages have methods to calculate this automatically as well. These can be calculated numerically as well, as follows: the derivative of the likelihood can be calculated by taking the difference of the likelihood at a parameter value slightly above the MLE estimate and slightly below, and then dividing this by the difference of these two points chosen. Similarly, the second derivatives can be obtained by taking the derivative of these values.

Estimating the variance between accounts for the distribution parameters was discussed in section 4.4. The parameter variances for the within and between variances can be summed to derive the total parameter variance if credibility is not being performed. Otherwise, the total parameter variance can be calculated directly using the Fisher Information matrix resulting from the fitting the parameters using Bayesian credibility. Or it can be estimated by using the same variance formula for a conjugate normal that was mentioned. If the source of the ILFs are from ISO or a similar source, the parameter

error will be difficult to calculate, but can be assumed to be small relative to the between companies parameter variance, which will need to be judgmentally selected.

Once the parameter variance has been obtained, it can be used to calculate the variance of the ILF using simulation. (The delta method can be used as well, but will not be discussed here.)

Once we have the variance of the capped loss pick and the variance of the ILF, the variance of the loss cost estimate for the account layer can be calculated by using the variance of a product formula: $V[A \times B] = V[A]\,V[B] + V[A]\,E[B]^2 + V[B]\,E[A]^2$. Using all of this, the total variance of each potential loss pick can be calculated at each capping level, and the capping level with the lowest variance can be selected. This variance depends on the number of ground up claims and the retention and limit of the policy being priced (for each ILF table, that is). A table of the optimal capping levels can be built by these dimensions and then the appropriate capping level can be looked up for each account, or alternatively, this value can also be computed in real time when pricing an account.

To help illustrate this method, a simulation was performed with varying amounts of ground up claims on a fictional account with both a retention and limit of one million (without considering credibility). Figure 6 shows the estimated variance of the final estimated loss cost (divided by the average variance, so that all of the variances can appear on the same graph) at various capping levels for different number of ground up claims. Note how the curves decrease initially and then start to increase. At lower capping levels, the total variance is higher due using less information about the account's actual losses as well as the increased uncertainty in the ILF. At higher capping levels, the variance starts increasing again due to more volatility in the capped loss pick. The point in between with the lowest variance is the optimal capping point. Note how the variance changes at a slower rate and is more stable with more ground up claims. Figure 7 summarizes the results and shows the optimal capping points for each amount of ground up claims. As expected, a higher capping level should be chosen for larger accounts with more ground up claims.

Figure 6: Relative variance by capping level

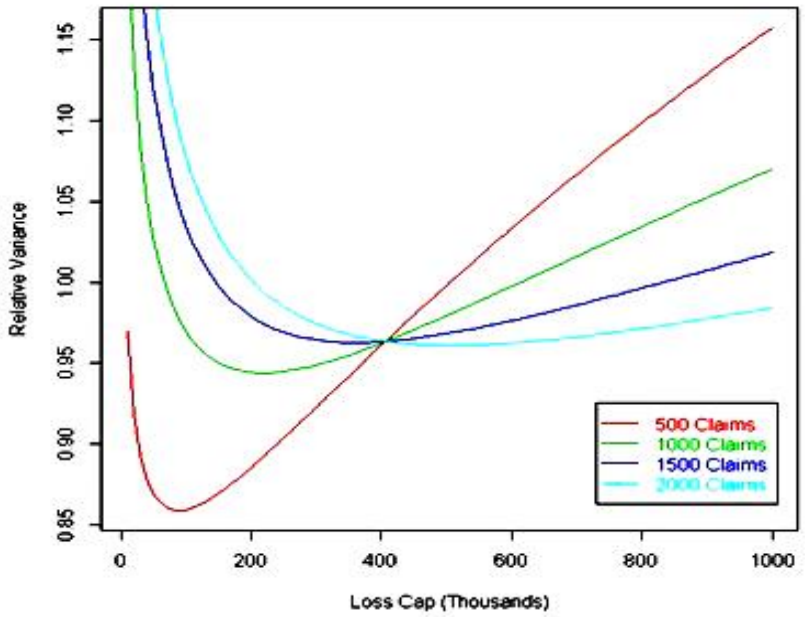
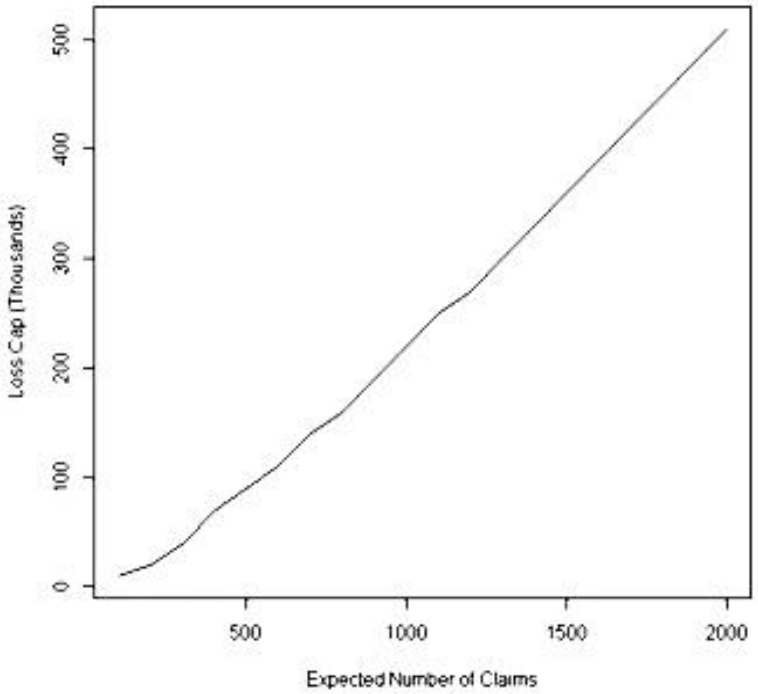


Figure 7: Optimal capping point



7. REFERENCES

[1] Bolstad, W., "Introduction to Bayesian Statistics (Second Edition)," 2007, p.106-108, 207-209

[2] Clark, D. R., "Credibility for a Tower of Excess Layers," Variance Casualty Actuarial Society, 2011, Volume 05, Issue 01, p.32-44. http://www.variancejournal.org/issues/05-01/32.pdf

[3] Dean, C., "Topics in Credibility Theory," Education and Examination Committee of the Society of Actuaries, 2005. https://www.soa.org/files/pdf/c-24-05.pdf

[4] Klinker, Fred., "GLM Invariants." Casualty Actuarial Society E-Forum, Summer 2011. 2011. https://www.casact.org/pubs/forum/11wforumpt2/Klinker.pdf

[5] Korn, U. "Credibility for Pricing Loss Ratios and Loss Costs." CAS E-Forum, Fall 2015. http://www.casact.org/pubs/forum/15fforum/Korn.pdf

[6] Korn, U. "Credibility and Other Modeling Considerations for Loss Development Factors." CAS E-Forum, Summer 2015. http://www.casact.org/pubs/forum/15sumforum/Korn.pdf

[7] Marcus, L. F., "Credibility for Experience Rating, A Minimum Variance Approach," CAS E-Forum, Summer 2010. https://www.casact.org/pubs/forum/10suforum/Marcus.pdf

[8] McNeil, Alexander J. "Estimating the tails of loss severity distributions using extreme value theory." ASTIN bulletin 27.01 (1997): 117-137. http://dx.doi.org/10.2143/AST.27.1.563210

[9] Scarrott, Carl, and Anna MacDonald. "A review of extreme value threshold estimation and uncertainty quantification." REVSTAT–Statistical Journal10.1 (2012): 33-60. https://www.ine.pt/revstat/pdf/rs120102.pdf

[10] Siewert, J. J., "A Model for Reserving Workers Compensation High Deductibles," CAS Forum, Summer 1996, pp. 217-244. https://www.casact.org/pubs/forum/96sforum/96sf217.pdf

[11] Venter, G., "Credibility Theory for Dummies," Casualty Actuarial Society Forum, Winter 2003, pp. 621–627. https://www.casact.org/pubs/forum/03wforum/03wf621.pdf

Biography of the Author

Uri Korn is the Industry Analytics Leader for AIG, Client Risk Solutions. Prior to that, he was an AVP & Actuary at Axis Insurance serving as the Research and Development support for all commercial lines of insurance. His work and research experience includes practical applications of credibility, trend estimation, increased limit factors, non-aggregated loss development methods, and Bayesian models. Uri Korn is a Fellow of the Casualty Actuarial Society and a member of the American Academy of Actuaries.

Removing Bias — The SIMEX Procedure

Thomas Struppeck, FCAS, MAAA

Abstract: SIMEX (Simulation-Extrapolation) is a very general technique that helps to correct for bias in estimates caused by errors in measurements of predictors. The method is well established in statistical practice, but seems to not be as widely known in actuarial circles. Using ordinary least squares regression as an example, the method is illustrated using some simple R code. The reader is encouraged to run the code (which is highlighted in grey) in R themselves to get hands-on experience using the method.

Keywords: SIMEX, MCSIMEX, Naïve Estimate, Unbiasedness, Bias

INTRODUCTION

Linear regression and its generalizations are commonly used tools when analyzing data. One variable, the response, is explained in terms of the other variables (the explanatory variables or regressors). The coefficient on the regressor is called its slope, and the goal of regression is to obtain a good estimate of it. Classically, regressors are assumed to have been measured with no error. When this assumption is violated, slope estimates can be biased. This has been known for some time and various methods have been developed to estimate and remove this bias. One such method, SIMEX, was introduced by Cook and Stefanski in their 1994 paper [CS]. One disadvantage of SIMEX is that an *a priori* estimate of the variance of the measurement errors is needed, and this is not always available.

This paper first presents an example of how the bias in ordinary least squares (OLS) regression arises in practice. Then this example is used to introduce the SIMEX procedure in a simple case. The example illustrates how observed noisy data can be used to estimate the regression line that accurate data would give if it could be observed or captured or obtained.

Historically, actuaries chose exposure and classification variables that were "…objective and relatively easy and inexpensive to obtain and verify" [WM]. The variables that required more judgment were used for underwriting or tiering. With the advent of more sophisticated rating methodologies, this line has come blurred. Now a great many variables may be used in a pricing model — and it is possible that there are missing values. If these values are imputed or modeled externally, the imputed value can be thought of as an observation with measurement error. As an example, "rural miles driven" might be the actual variable of interest, but only "total miles driven" is available. If a study indicates that for a given territory, half of all miles driven are rural, we might use half of "total miles driven" as a proxy for "rural miles driven". This proxy behaves like a measurement with error, and regressions

performed on the proxy will have slope parameters biased towards zero when compared to the true slope parameters for the underlying variable.

Methods like SIMEX can help remove bias from parameter estimates when using such data. For insurance data, the analyst has available the data that was recorded, not the actual data. The data that is recorded is called the observed data. The difference between the actual data and the observed data is called the noise. As we will see, the presence of noise may or may not introduce bias in the slope estimates depending on whether the noise is independent of the actual data or not. If we have an estimate of the variance of the noise, we can correct for it in our slope estimates.

Insurance data has often been rounded (say to the nearest thousand) or binned (say, 0-5, 6-10, 11-15) and while rounding or binning does introduce noise, we will see that it does not introduce bias. Other examples of insurance data such as miles driven (auto) or replacement cost for a building (homeowners) could be estimates (or proxies) and using these estimates (instead of the true values which are generally unavailable) can introduce bias. Estimates abound in reserving, consider for example case reserves based on legal opinions of potential jury awards; the difference between the actual award and the estimate could be viewed as noise.

ORDINARY LEAST SQUARES REGRESSION, THE GAUSS-MARKOV THEOREM, AND ATTENUATION

Ordinary Least Squares regression (OLS) is a commonly used tool that is very well behaved and very well understood, at least if the assumptions are all met. For instance, the Gauss-Markov Theorem says that in a regression if the errors have mean zero, have constant variance, and are uncorrelated, then the Best Linear Unbiased Estimator (more commonly, BLUE) of the slope coefficients is given by the OLS estimate. This says that (among other things) the slope estimates are unbiased, but in practice the slope estimates often are biased — how can this be?

The answer can be seen by looking closely at the assumptions:

1) The errors have mean zero.
2) The errors have constant variance.
3) The various errors are uncorrelated.

It is common for all three of these to be satisfied. The problem is with the word "errors".

The model being fit is: $Y_i = \beta X_i + \alpha + \varepsilon_i$

The "errors" are the ε_i, which are assumed to have mean 0 (or else we could change the intercept, α), to have the same variance for all *i*, and to be independent (and hence, uncorrelated.) Notice that the error is entirely associated with the dependent variable, Y_i. What is meant by that is that the independent variable, X_i, is assumed to be measured perfectly accurately. In practice this may or may

not be the case. Some regressors, such as policy limit or deductible may be known with certainty, but others, such as replacement cost, may have significant measurement error. If the error is independent of the actual value, then there will be attenuation (bias towards zero). Since the bias is towards zero, slope estimates will have their significance underestimated. That can cause otherwise significant regressors to appear insignificant. Surprisingly, if the error and true value are correlated, there may be no bias.

Let's look at an example:

```
set.seed (1001) # initialize the random number generator
x_actual <- runif(1000,10,20) # generate 1000 uniform (10,20) random variates
y_actual <- 3*x_actual + 2 # create the 1000 corresponding values of y
plot(x_actual,y_actual) # should all lie exactly on a line
abline(2,3,col="red")
(mod_a_a <- lm(y_actual~x_actual)) # coefficient estimates are exactly the correct (true) values
```

The reader is encouraged to run the sample blocks of code into R to follow along. Each block that involves generating random numbers will start with a call to set.seed[1].

We have generated 1000 uniform variates, X_actual, in the interval from (10,20). We then compute Y_actual = 3X_actual + 2. The points (X_actual,Y_actual) lie exactly along a line with slope 3. Indeed the regression shows the slope to be exactly 3. We will now add some noise to the Y_actual variables.

```
set.seed(1002)
# add some random noise to y_actual ... this is the y value that we observe
y_observed1 <- y_actual+rnorm(1000,0,5)
(mod_a_o1 <- lm(y_observed1~x_actual)) # the "extra" parentheses make R print the results
```

The slope coefficient on X_actual is estimated to be 2.911

```
summary(mod_a_o1)
```

Using the summary command, we can see the standard error of this estimate is 0.05237. We know that the true value of the slope is 3.000, which lies in the 95%-confidence interval.

Next, using the same Y-actual, we generate a different Y_observed (different noise) and estimate

[1] That function, set.seed, re-initializes the random number generator in R. This is done so that the random numbers that reader is using will match the ones used to make the examples. Different seeds are used each time so that no unintended correlations are introduced.

the new slope coefficient.

```
set.seed(1003) # cannot use the previous value or we will duplicate the old result!
# add some random noise to y_actual ... this is the y value that we observe
y_observed2 <- y_actual+rnorm(1000,0,5)
mod_a_o2 <- lm(y_observed2~x_actual)
summary(mod_a_o2)
```

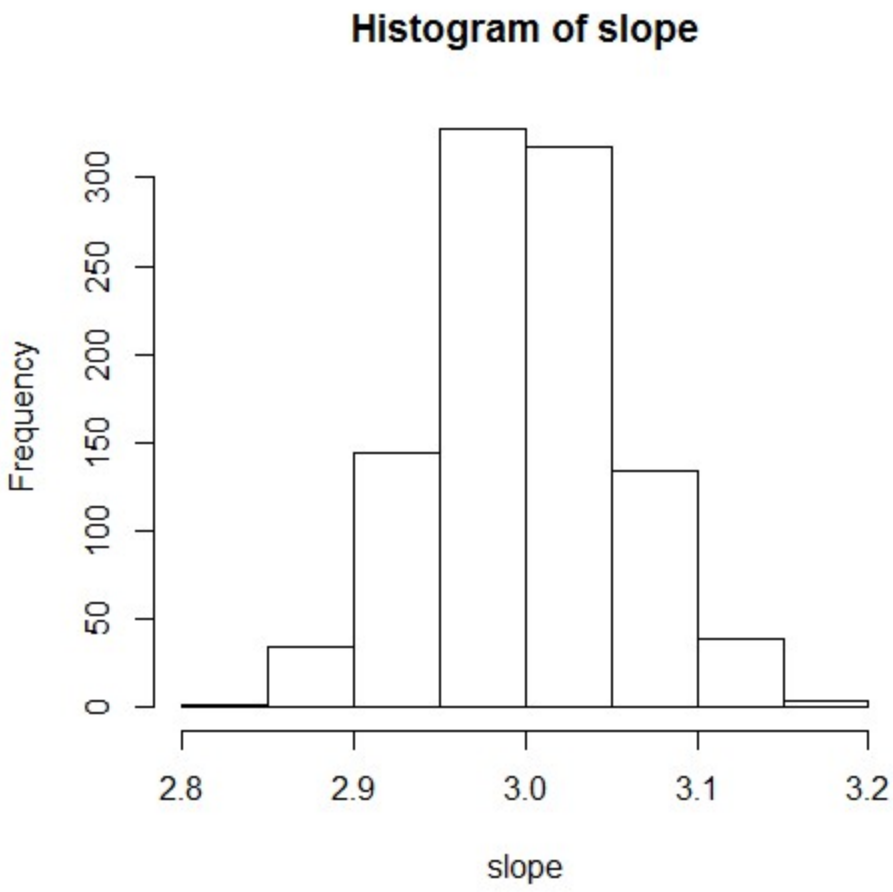
This time our slope estimate for X_actual was 2.985 which is closer to 3.000. To get a feeling for the sampling distribution of the slope coefficient, we will run 1000 regressions and record the slope estimates in a vector.

```
set.seed(1004)
slope <- rep(0,1000) # create a zero-filled vector of length 1000
for (i in 1:1000) {
    y_observed <- y_actual + rnorm(1000,0,5)
    mod_loop <- lm(y_observed~x_actual)
    slope[i]<- mod_loop$coeff[2]
}
summary(slope)
```

The average slope estimate is 3.000. That was luck, but most other seeds will also yield average estimates near the true value of 3.000. That is because the Gauss-Markov theorem tells us that ordinary least squares regression gives unbiased estimates.

Here is the histogram of the sampling distribution of the slope estimate:

```
hist(slope)
```



Now we will leave the noise added to Y fixed, and add some noise to X. By fixing the noise added to Y, we will be able to see how adding noise to X changes the result. (The variance of the noise added to X is twenty-five times smaller than that of the noise we added to Y, i.e. the standard deviation is one fifth as much.)

```
set.seed (1005)
y_observed3 <- y_actual+rnorm(1000,0,5) # We will leave this fixed from now on
x_observed1 <- x_actual+rnorm(1000,0,1)
(mod_o_o1 <- lm(y_observed3~x_observed1))
```

The slope estimate this time is way off, 2.681.

In this example, we have added normally distributed noise with mean 0 and standard deviation 1. The variance of the noise is what matters, not the exact distribution, so for example uniformly distributed noise on $[-\sqrt{3}, \sqrt{3}]$ (which also has standard deviation equal to one) should give similar results.

The estimate that we obtained is way outside the histogram from above. Perhaps we were unlucky? We will empirically estimate distribution of the slopes as we did before:

```
set.seed(1006)
for (i in 1:1000) {
    x_observed <- x_actual+rnorm(1000,0,1)
    mod_loop <- lm(y_observed3~x_observed)
    slope[i] <- mod_loop$coeff[2]
}
summary(slope)
```

The average estimate is 2.715, considerably smaller than the true value, 3.000. This bias toward zero is called attenuation bias. Our sample sizes (1,000) have been fairly large. Even larger sample sizes do not help in this situation. In other words, the slope estimates from OLS can be inconsistent[2] when there are errors in the regressors. Also, since the bias is towards zero, p-values and significance of regressors will be understated which, in a multivariable setting, could lead to a suboptimal choice of regressors.

The more noise that there is in the regressors, the worse that attenuation bias gets. What we would like to do is remove the noise from the data, but of course, that is impossible. We can however add more noise! Doing so increases the bias even more. We relate the amount of increase in the bias to the amount of additional (simulated) noise that we added. Once that relationship is established we use it to extrapolate back from the original noisy data to the ideal situation where there is no noise. That process, SIMulation followed by EXtrapolation, is the heart of the SIMEX algorithm. After a brief section on the theory, we will look at a hypothetical insurance data set and use SIMEX on it.

[2] An estimator is called inconsistent if it does not converge to the true value even as sample sizes go to infinity.

Observed values and Actual values

If we only have the noisy values, the variable X_actual can be thought of as a latent variable. The noisy values are the observed variable, so let's call these noisy values $X_{observed}$.

We can think of our model this way:

$$Y_{observed} = \beta X_{actual} + \alpha + \varepsilon_i$$

$$X_{observed} = X_{actual} + \delta_i$$

Conditional on the actual X value, the observed Y values have mean $\beta X_{actual} + \alpha$. There is noise ε_i added to Y, and we will assume that this noise has mean 0 and variance σ_Y^2.

We would like to estimate β and σ_Y^2, but we only observe the $Y_{observed}$ and the $X_{observed}$. While we do not actually directly observe the δ_i, often we will know something about the process that generates the δ_i. For example, if we are only know the year of an event which is equally likely to have occurred anytime during the year, we would probably use the mid-point of the year as the observed date; this is intended to make the error, as we have defined it, have mean 0. Or we may know from previous experience how far off, on average, the observed value is from the true value, *i.e.* the variance.

We will assume that δ_i has mean 0 and variance σ_δ^2. If we know its variance, and we know how it covaries with the actual X value, then we can, in the case of ordinary least-squares regression, actually compute the degree of attenuation.

Knowing how the errors covary with the measurements is not as unlikely as it might seem. The most common cases are perfect negative correlation and zero correlation. If the actual values are rounded or binned, the error and the actual value will be perfectly negatively correlated (for example, 4.2 gets always gets rounded to 4.0 and always produces an error of -0.2.) In other cases, assuming correlation zero is reasonable. As an example, consider the estimated replacement cost of a building. If we had ten appraisers evaluate the building, we would presumably get ten different estimates and these estimates would be centered around the actual value. We could estimate the variance of these estimates and use that as the variance of the error. The correlation between the estimate and the actual value would in this case be zero, assuming that a random appraiser from those ten did the measurement.

So, if repeated observations of a particular actual value always give the same observed value (*e.g.*

rounding) then the correlation will be negative 1. If on the other hand, repeated observations of a particular actual value give a range of observed values around the actual value (*e.g.* appraised values), then an assumption of correlation zero (uncorrelated noise) may be reasonable.

If we write $\sigma_{X\delta}$ for the covariance of X and δ, we have:

$$E(\hat{\beta}) = \beta \frac{\sigma_X^2 + \sigma_{X\delta}}{\sigma_X^2 + 2\sigma_{X\delta} + \sigma_\delta^2} \qquad \text{Equation 1 (General Case)}$$

Equation 1 can be thought of as a credibility weighted average of the true value of the slope parameter, β, and zero.

$$E(\hat{\beta}) = \beta\left(\frac{\sigma_X^2 + \sigma_{X\delta}}{\sigma_X^2 + 2\sigma_{X\delta} + \sigma_\delta^2}\right) + 0\left(\frac{\sigma_{X\delta} + \sigma_\delta^2}{\sigma_X^2 + 2\sigma_{X\delta} + \sigma_\delta^2}\right) \qquad \text{Equation 1A (Credibility Version)}$$

The weight on beta is the sum of the variance[3] of X and its covariance with δ and the weight on zero is the sum of the variance of δ and its covariance with X. Note that when the correlation is negative 1 (such as when the actual values has been rounded) Equation 1 becomes:

$$E(\hat{\beta}) = \beta \qquad \text{Equation 2 (Special case: rounded data.)}$$

So, in this case the estimator is unbiased. This equation is in fact the definition of unbiasedness.

When the covariance between the actual value of X and the error in measurement, $\sigma_{X\delta}$, is zero, we can see that the weights are just the respective variances. In this case, we have attenuation:

$$E(\hat{\beta}) = \beta \frac{\sigma_X^2}{\sigma_X^2 + \sigma_\delta^2} \qquad \text{Equation 3 (Special case: uncorrelated noise)}$$

When the covariance is equal to negative one times the variance of δ, a special case that we return to briefly at the end, all of the credibility is assigned to the true value β, and there is no attenuation. That these two cases need to be distinguished was first observed in 1950 by Berkson [B].

[3] To be exact, this is n/(n-1) times the sample variance.

The literature on measurement error refers to the biased estimator in Equation 1 as the naïve estimate. In the case of ordinary least squares regression, since we know the exact extent of the bias, we can simply correct for it. (See, for example, [CRSC] Chapter 2. or [Fu])

An alternative method which is easily implemented and much more generally applicable is SIMEX, which we now illustrate with an insurance-based example using hypothetical data.

Example 2:

Our data set will consist of blood levels of a toxin and exposure information obtained from exposure badges worn by 500 workers covered by a workers' compensation policy, which covers occupational disease claims. It is believed that the exposure is linearly related to the blood level in each worker. The data is from an older model of badge which was known to be inaccurate. These are being replaced by newer badges which are much more accurate. For this example, we have assumed that measurement error on the older model badges was known to be one unit. We wish to estimate the slope coefficient.

Going forward we will have be using the more accurate exposure information from the new badges and wish to infer expected blood level. We want the best possible estimate for beta.

The 500 workers' actual exposures and observed exposures are generated in the code below:

```
set.seed(1007)
actual_vec<-runif(500,5,10)
obs_vec<-actual_vec+rnorm(500,0,1) # add individual variation to get the badge reading for each worker
```

Now we will generate the (conditional) mean blood level for each worker. We selected an intercept of 10, but to make the exercise more exciting we will generate a random value for the true slope parameter (and reveal it at the end to see how we did).

```
set.seed(1008)
```

```
beta<-runif(1,2,4) # pick a uniformly distributed slope named beta in [2,4]
actual_bl<-beta*actual_vec + 10 # compute the actual mean for each individual
obs_bl<- actual_bl+rnorm(500,0,2) # add some random variation to the blood_level

naive_model<-lm(obs_bl~obs_vec)
(naive_slope<-naive_model$coeff[2])
```

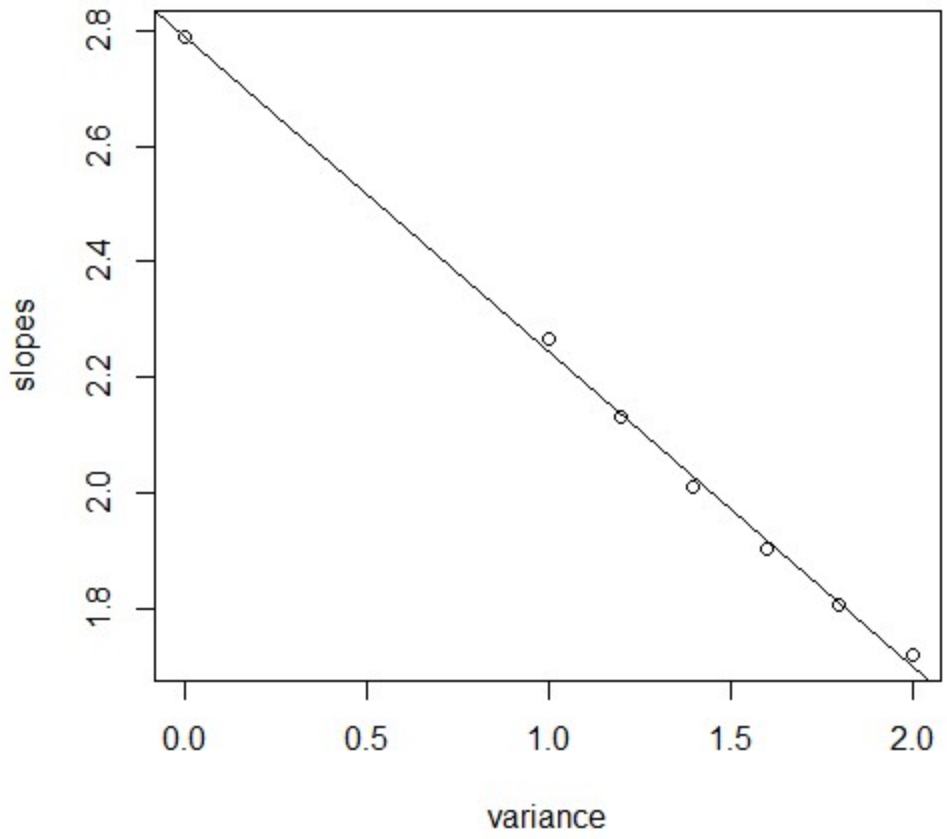
The slope estimate is 2.266. This is the naïve estimate. For this example we are given that the readings on the old badges vary from the mean with a standard deviation of 1 unit. We will add more noise to our observed data (increasing the variance to 1.2), compute the estimated slope and repeat many times (1000). This sample taken from the sampling distribution of $\hat{\beta}$ will by the law of large numbers have a mean very close to the expected value of $\hat{\beta}$. We then repeat that process for variances of 1.4, 1.6, 1.8, and 2.0.

```
set.seed(1009)
# SIMEX
slope<-numeric(1000) # allocate space for storing slopes
average_slope<-numeric(5) # space for averages
for (i in 1:5) { # 1:5 is 1,2,3,4,5
    extra_variance <- i/5 # generates 1/5, 2/5, 3/5, etc.
    for (j in 1:1000) {
            obs_plus_extra_noise <- obs_vec + rnorm(500,0,sqrt(extra_variance))
            slope[j] <- lm(obs_bl~obs_plus_extra_noise)$coef[2]
            }
    average_slope[i] <- mean(slope)
    }
variances<-1+0:5/5
simulated_slopes<-c(naive_slope,average_slope) #add naïve estimate to vector of slopes
```

```
(simex_mod<-lm(simulated_slopes~variances)) #find the slope and intercept of the best fit line
```

The naïve estimate was 2.266. We fit a linear model predicting the slope from the number of units of noise. The original data had one unit of noise … noiseless data would have zero units of noise; zero units of noise corresponds to the y-axis in the last regression above. In other words, the SIMEX estimate (of the slope) is the intercept from the linear model stored in sim_mod above. That is 2.791, significantly larger than the naïve estimate.

```
variance<-c(0,variances) #add ideal variance (0) to list of variances for plotting
slopes<-c(simex_mod$coeff[1],simulated_slopes) #add intercept from last regression for plotting
plot(variance,slopes)
abline(simex_mod$coeff[1],simex_mod$coeff[2]) #plot the regression line
```



The plot appears to show some concavity upwards. It can be shown that this is indeed the case and so SIMEX with linear extrapolation gives conservative estimates. Additionally, more sophisticated extrapolation techniques can obtain sharper and still conservative results. (See for example [CRSC], or the original paper by Cook and Stefanski [CS].) These more elaborate extrapolation techniques are implemented in the R package described briefly below in the section titled "The *simex* package in R". There a version of the above plot with a quadratic extrapolating function is shown.

One of the very nice by-products of SIMEX is this plot. The plot clearly illustrates what is happening, even more so if quadratic or other non-linear extrapolation method is used. This is often valuable in helping to explain how the method works and showing users or regulators that the method produces conservative estimates.

Time to reveal the true value of beta:

```
beta
```

So, the true value was 3.335.

Explaining the results to others.

Senior management or regulators, when presented reviewing a SIMEX analysis, may object to the addition of random noise into the data. It is important to quell such concerns. Here is a suggested approach:

The SIMEX procedure first computes the regular OLS estimate, the naïve estimate. This is done using the observed data with no additional noise.

We know that the naïve estimate will be biased if the errors are uncorrelated with the actual values. We want to correct for that bias and in order to estimate how much bias there is, we introduce additional noise. We then adjust the naïve estimate by the estimated bias adjustment.

It may be helpful to show the results of running the procedure a few times using different seeds in order to show that the choice of noise does not materially change the final estimates.

The *simex* package in R.

While working through the algorithm is a good way to learn and understand the method, there is no need to reinvent the wheel. A well-documented and versatile implementation of the algorithm is easily available in R. [R]

The *simex* package, written by Wolfgang Lederer and Helmut Küchenhoff [LK1], implements some more sophisticated extrapolation techniques, which improve the bias removal significantly. It also

includes a variation called *mcsimex* that can handle discrete data with misclassification[4]. The package has the nice feature that the user can simply pass it an object of type *lm* or *glm* and it will perform the procedure on the corresponding model. More details on this package can be found in [LK2].

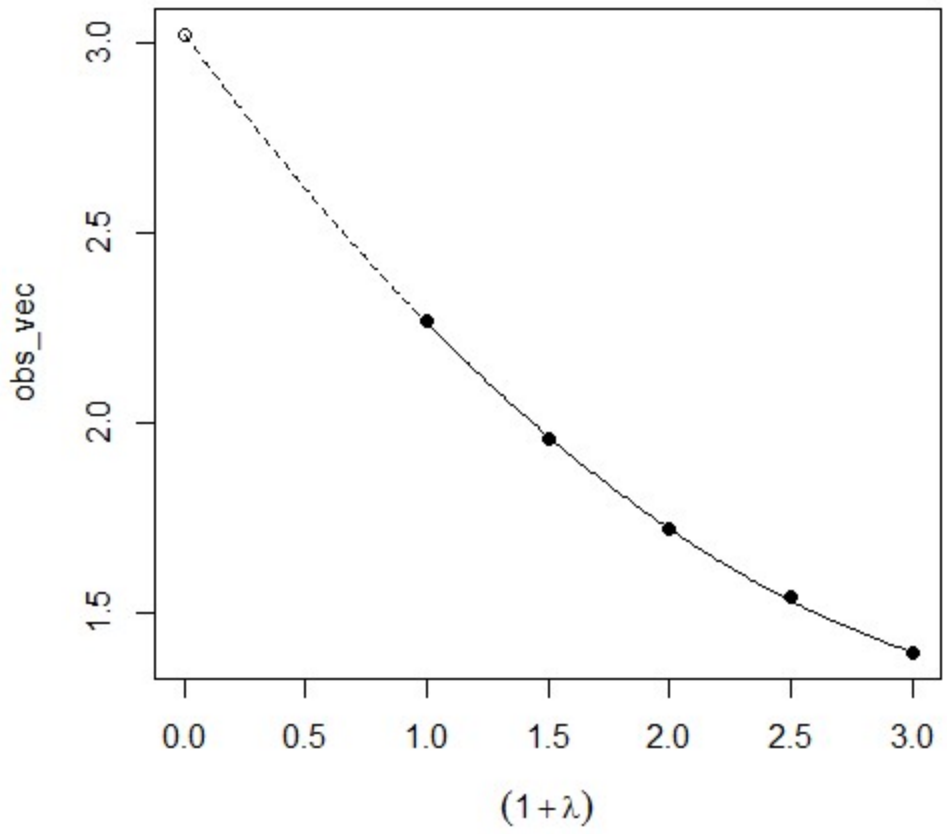
```
set.seed(1010)
library(simex) # this assumes that the simex package has already been loaded, if not, do so
# the call to simex below uses the default extrapolation method (quadratic), default number of
# iterations for each value of lambda (the amount of noise), with values of 0.5, 1.0, 1.5, and 2.0

# the given parameters are: the naïve model, the name of the variable with measurement error, the amount of error, finally the switch asymptotic=FALSE suppresses asymptotic variance estimation.
simex_model<-simex(naive_model,"obs_vec",1,asymptotic=FALSE)
simex_model
```

The model output shows the estimated coefficient on obs_vec of 3.022 which is a much better estimate of the true value (3.335) than either SIMEX with linear extrapolation (2.791) or the naïve model (2.266) produced.

To obtain the corresponding plot, we need only pass the simex model to the generic plot function, again using defaults for most parameters.

```
plot(simex_model)
```

[4] The MCSIMEX methodology was used in an NIH-funded study of predictors of periodontal disease [SB], which potentially could be of interest to dental insurers.



What about rounded data?

If regressors have been rounded, the resulting estimates from OLS regression are not biased! This may seem surprising since rounded data appears to look much like data with noise uniformly distributed on the interval (-1/2,1/2) added. The difference is subtle. This is the case where the noise and the actual value are correlated. An example may make this clearer. If the actual data was 9.4, we would always round that down to the observed rounded value of 9.0. In other words, knowledge of the true value, tells you what the observed value will be — they are correlated. Contrast this with adding random noise where the observed value could be either bigger or smaller for a fixed actual value. For a further discussion of this point see for example Faraway [F] or the original 1950 paper by Berkson [B].

CONCLUSION

SIMEX is a general method that allows one to estimate the attenuation bias present in a model. This paper illustrates the technique in the setting of ordinary least squares regression. The requirements to use SIMEX are an understanding of the size of the measurement error in the regressors and an understanding of how these are correlated with the actual values. Unless the errors are from rounding or binning, usually it will be reasonable to assume that the correlation is zero and that attenuation bias is occurring.

By adding even more noise to the regressors and observing the additional attenuation, it is possible to estimate the amount of attenuation that is already present in the data.

One important by-product of the analysis is a graph that is easy to explain and understand. That graph allows the method to be quickly explained in a non-technical way.

An implementation of SIMEX (and a related method, MCSIMEX) is available in the *simex* package in R. This publicly available implementation has a more sophisticated extrapolation procedure which removes even more of the attenuation from the slope estimates.

REFERENCES

[1.] [B] Berkson, J. (1950). Are there two regressions? Journal of the American Statistical Association 45, 165-180.
[2.] [CRSC] Carroll, R.J, Ruppert, D., Stefanski, L.A., & Crainiceanu, C.M. (2006). Measurement Error in Nonlinear Models. New York: Chapman & Hall/CRC.
[3.] [CS] Cook, J. and Stefanski, L. (1994). Simulation-extrapolation estimation in parametric measurement error models. Journal of the American Statistical Association 89, 1314-1328.
[4.] [F] Faraway, J.F. (2005). Linear Models with R. New York: Chapman & Hall/CRC.
[5.] [Fu] Fuller, W. (1987). Measurement Error Models. New York: Wiley.
[6.] [LK1] Wolfgang Lederer and Helmut Küchenhoff (2013). simex: SIMEX- and MCSIMEX-Algorithm for measurement error models. R package version 1.5. https://CRAN.R-project.org/package=simex.
[7.] [LK2] Wolfgang Lederer and Helmut Küchenhoff (2006) R News, Vol. 6/4, October 2006.
[8.] [R] R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
[9.] [SB] Slate, E.H. and Bandyopadhyay, D. (2009) Stat Med 28(28), 3523-38.
[10.] [WM] Werner, G. and Modlin, C. (2016). Basic Ratemaking. CAS.

Residual Loss Development and the UPR

Richard L. Vaughan, FCAS, FSA, MAAA

Abstract. Traditional reserve estimators such as chain-ladder and Bornhuetter-Ferguson model unpaid losses as a function of accident period versus lag to payment or reporting. The result of primary interest is expected future losses; these are derived from intermediate results such as lag factors and loss ratios.

With certain adjustments, traditional estimators may also be used for the statutory Unearned Premium Reserve, or UPR, for long-duration contracts. This reserve is governed by SSAP 65, the most important requirement of which is "Test 2", that earnings be recognized in proportion to the expected emergence of losses and expenses. Here the estimators model unincurred losses by issue month versus lag to incurral. The result of greatest interest is now the set of lag, or earnings, factors.

Adjustments are necessary to accommodate the decline in exposure due to cancellations, the deficiency of recent diagonals in the issue-to-incurral lag triangle due to unreported losses, and the unusual shapes and weight of tails for immature business. In particular, it is convenient to develop not losses *per se* but partial loss ratios to premium remaining in force, thus "factoring out" the effect of cancellations and leaving the results correct on a no-cancellation basis.

In this paper we suggest taking this adjustment one step further, and developing loss ratios to the product of premium in force and a set of positive *a-priori* earnings factors, the final earnings factors being the product of the *a-priori* factors and these "residual" earnings factors. In the case of automobile extended service contracts, there exists an excellent model for such *a-priori* factors, published by Kerper and Bowron in 2007 [1], but the technique is not dependent on any particular underlying model.

We demonstrate that this procedure (a) improves the robustness of the estimators to lack of perfect homogeneity in the data, (b) greatly simplifies the specification and calculation of tail factors, and (c) facilitates the use of reference factors to improve the estimates at lags where the experience data is sparse.

1. INTRODUCTION

1.1 Statutory Requirements for the Unearned Premium Reserve

Statement of Statutory Accounting Principles 65 (SSAP 65) defines three tests for the adequacy of the Unearned Premium Reserve (UPR) for contracts longer than 12 months in duration: Test 1, provision for refunds, Test 2, earnings to be proportional to emergence of losses and expenses, and Test 3, provision for unincurred losses and expenses. Of these, Test 2, which requires that the UPR be no less than the gross premium multiplied by the ratio of expected future to total losses, including expenses, is normally dominant. More importantly, a UPR created to satisfy, and just satisfy, Test 2 will release earnings in such a way as to make immature inception-to-date loss ratios useful predictors of ultimate loss ratios, and thus give management an accurate measure of the performance of the business in question.

1.2 Satisfaction in Aggregate

SSAP 65 need only be satisfied in aggregate, for all of a company's long-duration contracts together. But it is the common practice of writers of long-duration contracts to calculate the UPR at the contract level, using either predefined "strings" (vectors of UPR factors by lag) or formulas. Only then do they accumulate this UPR over accounting segments of interest, and finally sum it

across the entire company. The company's implied goal is that satisfaction of statutory requirements by most individual contracts will ensure satisfaction for most accounts and that satisfaction for most accounts will ensure satisfaction in aggregate.

The actuary testing the carried UPR typically starts by comparing average carried UPR factors with those indicated by experience, over reasonably homogeneous subdivisions of accounting segments, such as term groups. A close fit suggests, though it does not prove, that the carried UPR factors are satisfactory contract by contract, and will produce loss ratios that are reasonable predictors of ultimate experience. At the least it is, provisionally, a negative result in a management-by-exception sense, allowing attention to be focused elsewhere. The actuary also sums the indicated UPR's across each accounting segment, and finally across the entire company, to test the aggregate carried UPR for satisfaction of SSAP 65.

While the author believes this is a correct procedure it should be pointed out that it is really a matter of how the requirement that Test 2 be satisfied *in aggregate* is interpreted. Does this mean that the UPR must be no less than the aggregate inforce premium multiplied by aggregate future losses divided by aggregate ultimate losses (a "ratio of aggregates" definition), or does it mean that the UPR must be no less than the aggregate, over some exhaustive and mutually exclusive set of subdivisions, of premium multiplied by future losses divided by ultimate losses (an "aggregate of UPR's" definition)? We believe that the only sensible interpretation of the requirement is the latter. This is explained in Appendix A.

1.3 Estimation in Detail

Strings or formulas can give a UPR for each contract or for any collection of contracts, however small. But strings or formulas must be validated; this requires subdivisions that have enough experience to be credible, as well as being reasonably homogeneous.

Subdividing for homogeneity considers factors such as account, contract type (mechanical repair, GAP, etc.), insured product (automobile, boat, appliance, etc.), term in months, term in miles, and manufacturer's warranty. But using too many factors simultaneously may produce subdivisions with too few contracts for reliable estimates of UPR factors. So the actuary often must settle for reasonable, but not perfect, homogeneity.

Within a subdivision, we may use conventional methods to estimate UPR factors, loss ratios, and unincurred losses, provided certain adjustments are made to accommodate cancellations, unreported losses in policy month versus accident lag triangles, and the possibility of tail factors. For complete details of these adjustments see [2]. Here we propose a further adjustment to improve accuracy with imperfectly homogeneous data, to simplify greatly the projection of tails, and to weight the results against a simple set of reference factors to remove noise from the development at the later lags. The adjustment is similar to that used in the All-Terms Factors model, described in [2], to obtain residual

ATF's, and we call it "residual loss development". It requires a matrix of *a-priori* expected earnings factors, by issue month versus lag; for automobile service contracts; this may be derived contract by contract from exposure models such as that of Kerper and Bowron (KB) [1],[2].

2. BACKGROUND AND METHODS

2.1 Estimators of Earnings Factors

We consider estimators to be applied to experience data within a subdivision and to return earnings factors representing averages across the contracts in the subdivision. If these earnings factors resemble similar averages using the company's carried strings or formulas, we shall regard the strings or formulas as confirmed at that level of aggregation; otherwise, we shall take the results as evidence of a need for change.

Our suggested technique of residual loss development may be applied in conjunction with many underlying development methods. For concreteness in the discussion we assume that the estimators of earnings factors are in the chain-ladder family, and the estimators of future losses are in the Bornhuetter-Ferguson family, both of which have proven reliable in the context of extended service contracts. These families encompass variations in depth and weighting of average development factors, graduation of development or lag factors, and choice of expected loss ratios.

These estimators do require adjustment to cope with cancellations and with unreported losses in the issue-period-versus-incurral-lag triangle. For the purpose of this paper the most significant such adjustment is to develop, not losses *per se*, but partial loss ratios; this quite neatly produces earnings factors on a no-cancellation basis as required for SSAP 65 Test 2.

Losses include loss adjustment expenses where available; in practice the DCC component of such expenses may be negligible and the A&O component may be assumed to earn in parallel with the losses themselves. Issue periods and lags may be of any length but are normally (and will be assumed to be) of length one month.

The lag factors are normalized to total 1.00 and are then called earnings factors, since they represent the fraction of gross premium to be earned at each lag according to the principle of SSAP 65 Test 2. Their reverse cumulative sums represent SSAP 65 Test 2 UPR factors at the moment of issue and the beginning of each subsequent month; the first such factor will be 1.00. It is usually assumed that contracts are written uniformly throughout each month, so that there are potentially T+1 months with nonzero earnings factors (including the initial 1.00) if the term of the contracts is T. The experience may include some "goodwill" claims incurred even later; these may be handled by allowing the tail to extend beyond the term, or by treating all goodwill claims as if paid in the last month of the term and accepting the usually small misstatement of earnings in that month. Administrative systems usually require strings or formulas with exactly T factors after the initial 1.00.

2.2 Imperfect Homogeneity

It is not common for every subdivision of an account to be entirely homogeneous as regards factors that affect the earnings pattern. If all contracts in a subdivision have identical term in months, they still may differ in factors such as term in miles, manufacturers' basic warranties, manufacturers' power-train warranties, and odometer readings at issue. They may also differ in expected loss ratios, i.e. premium adequacy; while the loss ratio does not affect the earnings pattern of an individual contract, heterogeneity in loss ratios may affect the average earnings pattern of a subdivision.

Heterogeneity that is random over time, but stationary, presents little problem. Estimated earnings factors will be correct for the static mix, and, if it continues, will be correct on average for new contracts. Heterogeneity that changes over time is more serious. If contracts written in the early issue months of our lag triangle differ from those written later, then the resulting estimated earnings pattern may not be correct for either group of contracts, or even for a consistent mixture of the two, because each lag includes a different proportion of earlier and later contracts.

2.3 Residual Loss Development

By *residual loss development* we mean dividing the known cells of a loss triangle by an *a-priori* earnings pattern, developing the quotients, and multiplying the resulting earnings factors by the a-priori earnings pattern for both known and projected cells. The *a-priori* factors explain part of the observed pattern; we develop the residuals; the final estimate is the combination of a-priori and residual patterns.

Let the matrix $\boldsymbol{L} = [L_{ij}]$ represent losses for issue month i incurred at lag j, and reported through month $i+n$ (and therefore a "triangle" populated only for $i+j<=n+1$). Similarly, let $\boldsymbol{E} = [E_{ij}]$ represent exposures, such as premiums or numbers of contracts, for issue month i still in force, i.e. not cancelled, as of lag j. It is important to note that expirations, whether by months or miles, have no effect on the status of being "still in force"; the analysis is much simpler if the effect of expirations is allowed to emerge via the earnings factors rather than the exposure.

We are interested in completing $\boldsymbol{L}$ to a fully-populated "rectangle" $\boldsymbol{L^c}$ by estimating values for the future portion where $i+j>n+1$, as well as for the unreported portions of cells in the original $\boldsymbol{L}$. For SSAP 65 Test 2, we may be particularly interested in the earnings factors, or expected values of a row of $\boldsymbol{L^c}$, normalized to total 1. In completing $\boldsymbol{L}$ to $\boldsymbol{L^c}$ we may need future exposures; we assume that we have already completed $\boldsymbol{E}$ to a rectangle $\boldsymbol{E^c}$, for example by applying chain-ladder to $\boldsymbol{E}$ to estimate contract persistency rates, usually a straightforward procedure.

Finally we assume we have already estimated a vector $\boldsymbol{b} = (b_j)$ of cumulative *report* lag factors, for example by applying chain-ladder to a triangle of losses by incurral month versus report lag, again

usually a straightforward procedure. In the absence of reliable report-date data we may let $\boldsymbol{b}$ represent cumulative incurral-to-payment lag factors, in which case $\boldsymbol{L}$ should contain only losses paid through the valuation date. This amounts to defining report date as payment date. In the absence of report or payment dates, it is possible to estimate $\boldsymbol{b}$ from a single issue-to-incurral triangle, simultaneously with the earnings factors, though we do not discuss this further here.

There are then several possible procedures for completing $\boldsymbol{L}$ to $\boldsymbol{L^c}$, and estimating earnings factors:

- *Conventional loss development* applied directly to $\boldsymbol{L}$ (and possibly involving $\boldsymbol{E}$) yields first a vector of lag factors $\boldsymbol{f}=(f_j)$, then a set of ultimate losses by month i , and finally the future values of L_{ij} for each cell i,j. The factors f_i will not usually be satisfactory for SSAP 65 Test 2 analysis because they confound the effects of cancellations and losses. Moreover, the method fails for the issue-versus-incurral lag triangle $\boldsymbol{L}$ unless all losses are reported in the month of incurral, as the recent diagonals of $\boldsymbol{L}$ will be deficient because of unreported losses. A crude but common workaround is to suppress the last few diagonals, which sacrifices potentially useful recent information. A better approach is to adjust the latest diagonals for the expected unreported fractions, setting $L^a_{ij} = L_{ij} \,/\, b_{(n+2-i-j)}$, but the resulting lag factors f^a_j still confound cancellations with losses.
- *Exposure-adjusted loss development* completes the triangle $\boldsymbol{R}$ of loss ratios to inforce exposure, adjusted for the expected fraction reported to date, to the rectangle $\boldsymbol{R^c}$, where $R_{ij} = L_{ij}/E_{ij}*b_{(n+2-i-j)}$. This yields a vector of lag factors $\boldsymbol{g}=(g_j)$, which now reflect the emergence pattern of losses on a no-cancellation basis. If $\boldsymbol{R^c}$ is eventually converted to $\boldsymbol{L^c}$, the cancellations are accounted for by the decline in the exposures $\boldsymbol{E}$. This step, which is not always necessary if the lag factors are the main result of interest, must include the previously unreported fraction of "known" cells as well as the entirety of "future" cells.
- *Residual loss development* completes the triangle $\boldsymbol{R^*}$ of loss ratios to inforce exposure, adjusted for the expected fraction reported to date *and for the expected fraction earned in each period j*, to the rectangle $\boldsymbol{R^{*c}}$, where $R^*_{ij} = L_{ij}/E_{ij} \,*b_{(n+2-i-j)}*A_{ij}$, where $\boldsymbol{A}=[A_{ij}]$ gives the fraction of losses incurred in month i expected, *a-priori*, to emerge at lag j. This yields a vector of residual lag factors $\boldsymbol{h}=(h_j)$, which is multiplied by A_{ij} to produce a *matrix* of final lag factors $\boldsymbol{G}=[g_{ij}]$, on a no-cancellation basis.

The *a-priori* earnings matrix $\boldsymbol{A}$ may be derived by applying a formula to each contract and taking a weighted sum over the contracts remaining in force in each cell. Normalization over each row may be deferred until inside the loss development calculations.

For automobile service contracts, the Kerper-Bowron (KB) exposure model [1],[2] is an excellent basis for deriving $\boldsymbol{A}$. For contracts of other types (usually simpler), any reasonable earnings pattern, such as pro-rata after manufacturer's warranty, may be used; normally subject to a minimum positive value so that some earnings are possible at any lag.

Note that the technique described here is independent of the method used to estimate the earnings factors $\boldsymbol{f}$ from the original $\boldsymbol{L}$ and $\boldsymbol{E}$, or $\boldsymbol{g}$ from $\boldsymbol{R}$, or $\boldsymbol{h}$ from $\boldsymbol{R^*}$. Typically this will be some member of the chain-ladder family, with suitably-chosen depth of averaging, weighting, tail factors, and so forth, but it could be, e.g., a regression model. Similarly, the technique is independent of the

method used to *complete* the triangles. Typically this will be Bornhuetter-Ferguson, with Cape Cod expected loss ratios, or with Gluck decay factors giving a spectrum of possibilities from pure chain-ladder through pure Cape Cod. These methods are usually very satisfactory for Warranty business, with its characteristic high frequency and narrow size-of-loss distribution.

The Test 2 UPR factors, to be applied to gross inforce premium, are the reverse cumulative sums of the vector **g** or of the rows of the matrix **G**.

If a collection of contracts is perfectly homogeneous in earnings pattern, and losses emerge "noiselessly", then residual loss development will produce the same completed loss rectangle $\boldsymbol{L^c}$ as exposure-adjusted loss development, independent of $\boldsymbol{A}$, provided the rows of $\boldsymbol{A}$ are all proportional to each other. Each row of $\boldsymbol{G}$ from residual loss development will then be identical to the vector $\boldsymbol{g}$ from exposure-adjusted loss development. In the normal situation with random fluctuations in emerging losses this will no longer be true, in general, for the chain-ladder estimator. For any $\boldsymbol{L}$, a constant $\boldsymbol{A}$ will produce results from developing $\boldsymbol{R^*}$ identical to those from developing $\boldsymbol{R}$, so simple exposure-adjusted loss development is a special case of residual loss development, and the two methods may be handled by common program code.

The matrix $\boldsymbol{A}$ should normally be of full length to cover the terms of all contracts in the data, even if the data itself is immature, for reasons explained in Section 2.4.2 below.

If residual loss development is to have its expected beneficial effects, then the *a-priori* earnings factors in $\boldsymbol{A}$ should in fact reflect knowledge of the contracts included in each row of $\boldsymbol{L}$. Usually they will be averages over those contracts of formula earnings factors, or of earnings factors corresponding to UPR "strings" assigned to the contracts on an administrative system. Since they represent the expected emergence pattern of *losses* for that issue month, these averages ideally should be weighted not by contracts (which would ignore differences in expected loss costs) nor by actual premiums (which would ignore differences in expected loss ratios) but instead by theoretical premiums proportional to expected loss cost. Expected loss costs differ for two main reasons: the products insured (for example autos by manufacturer, make, and model) and the contract provisions (term, manufacturer's warranty, etc.). When the mix of products is not homogeneous over time, it may be necessary to derive an expected loss cost for each contract, using, for example, hierarchical credibility over the relevant characteristics. When just the contracts differ over time, it may be sufficient to use contract relativities estimated simultaneously with the KB or other *a-priori* earnings factors; this is part of the All-Terms Factors (ATF) model described in [2].

2.4 Advantages

Residual loss development has several useful features that improve or facilitate the estimation of earnings and UPR factors for long-duration contracts.

2.4.1 Handling of Heterogeneous Data.

Residual loss development circumvents the problem of imperfectly homogeneous data, with mix changing over time, precisely because $\boldsymbol{A}$ is a matrix, with a different *a-priori* emergence pattern for each issue month. If $\boldsymbol{A}$ correctly captures the changes in expected earnings patterns by issue month, these will be correctly propagated in the final development factors g_{ij} and the projected future losses L_{ij}. Each row of $\boldsymbol{G}$ will be a better representation of the earnings pattern of its contracts than the average vector of factors $\boldsymbol{g}$ that would have been produced by development of loss ratios to inforce premium alone. The *residual* factors h_j are still averages across issue months; the loss development process by itself cannot be expected to detect or compensate for heterogeneity. It is the matrix $\boldsymbol{A}$ that converts $\boldsymbol{h}$ into the matrix $\boldsymbol{G}$. Any positive matrix $\boldsymbol{A}$ will mechanically produce results, but only a reasonable $\boldsymbol{A}$ will produce reasonable results. Fortunately it is usually easy to construct a reasonable $\boldsymbol{A}$.

In this discussion we confine our attention to the estimation of earnings factors, not the completion of the loss triangle proper. But it goes without saying that residual loss development can improve the cell-by-cell accuracy of the completed loss triangle $\boldsymbol{L^c}$, since $\boldsymbol{G}$ contains a separate set of factors for each issue month, rather than a single average set.

2.4.2 Simplification of Tail Projections

Tails are necessary in projecting earnings factors for long-duration contracts whenever the available data is shorter than the longest term of the contracts in question. Projecting a tail usually requires considering its starting lag, its length, its shape, and its weight relative to the preceding known lags (or the number of such "lookback" lags to average to estimate the weight) [2]. Specifying length and shape requires information beyond that contained in the loss triangle, and this information must be passed as parameters to a program executing the loss development calculations. This complicates the logic of that program and of any programs calling it. Residual loss development, on the other hand, already captures all the information necessary to specify the shape of a tail in the *a-priori* earnings factor matrix $\boldsymbol{A}$, provided $\boldsymbol{A}$ includes factors for both known and projected lags. The tail of the *residual* earnings factors may simply be taken as constant (which gives pro-rata UPR factors). The result will be a separate tail for each row in the final matrix of earnings factors $\boldsymbol{G}$, and each such tail will have the shape specified by the corresponding row of $\boldsymbol{A}$.

2.4.3 Development of Sparse Triangles

When the data in a loss triangle is sparse, earnings factors estimated by loss development may be very irregular. A common solution to this problem is to take a weighted average of the raw earnings factors and a set of reference factors, usually derived by formula or from a larger volume of related experience. As with tail shapes, having to pass reference factors as separate parameters may complicate the logic of the loss development program and of all programs calling it. Residual loss

development circumvents this problem by including all necessary reference factor information in the matrix $\boldsymbol{A}$. The specified "credibility" weight is applied to the measured residual earnings factors, and its complement to a set of constant factors. The result will be a separate set of smoothed factors in each row of the final matrix of earnings factors $\boldsymbol{G}$.

2.5 Examples

2.5.1 Construction of illustrative triangles

To illustrate the above concepts we start with the following 12 x 12 matrix $\boldsymbol{E^c}$ of inforce exposures, both known and (already) projected; $\boldsymbol{E^c}$ incorporates growth, shows the effect of a moderate amount of cancellations, and, for verisimilitude, includes some randomness in written premiums and cancellations. A simplifying assumption here is that each month's writings take place at the beginning of the month, so that the first column represents written exposure and the remaining eleven columns represent exposure still in force after cancellations to date:

	0	1	2	3	4	5	6	7	8	9	10	11
1	216054	194448	183646	177164	172843	170683	168522	166361	164201	162040	162040	162040
2	179896	161907	152912	147515	143917	142118	140319	138520	136721	134922	134922	134922
3	206126	185513	175207	169023	164901	162839	160778	158717	156656	154594	154594	154594
4	254353	228918	216200	208570	203483	200939	198396	195852	193309	190765	190765	190765
5	208020	187218	176817	170576	166416	164336	162255	160175	158095	156015	156015	156015
6	195239	175715	165954	160096	156192	154239	152287	150334	148382	146430	146430	146430
7	236256	212631	200818	193730	189005	186642	184280	181917	179555	177192	177192	177192
8	211880	190692	180098	173742	169504	167385	165267	163148	161029	158910	158910	158910
9	242908	218617	206472	199184	194326	191897	189468	187039	184610	182181	182181	182181
10	218130	196317	185411	178867	174504	172323	170142	167960	165779	163598	163598	163598
11	270039	243035	229534	221432	216032	213331	210631	207930	205230	202530	202530	202530
12	263241	236917	223755	215858	210593	207961	205328	202696	200063	197431	197431	197431

In the present context we have the luxury of specifying the exact parameters generating the loss triangle $\boldsymbol{L}$. In addition to the exposure (the known part $\boldsymbol{E}$ of $\boldsymbol{E^c}$) these include a loss ratio, a loss emergence pattern for each issue month, and the cumulative fractions of losses reported as of each lag from incurral. To illustrate the effect of heterogeneity in a particularly simple way, we initially take all of these remaining parameters as fixed or "noiseless": a loss ratio of 70% for all issue years, cumulative fractions reported of **(0.3 0.7 0.9 1 1 1 …)**, and two separate loss emergence patterns, one for the first six months and one for the last six months, respectively as follows:

1	2	3	4	5	6	7	8	9	10	11	12
0.0152	0.0152	0.0152	0.0303	0.0758	0.1212	0.1515	0.1515	0.1515	0.1364	0.0758	0.0606
0.0169	0.0169	0.0169	0.0169	0.0169	0.0339	0.0847	0.1356	0.1695	0.1695	0.1695	0.1525

The discontinuity in earnings patterns might exist because the manufacturer has lengthened its factory warranty. The resulting $\boldsymbol{L}$ becomes:

	1	2	3	4	5	6	7	8	9	10	11	12
1	2291	2062	1948	3758	9166	14482	17874	17644	17415	13921	6015	2062
2	1908	1717	1622	3129	7632	12058	14882	14692	13051	9015	2146	0
3	2186	1968	1858	3585	8745	13817	17052	15150	11630	4427	0	0
4	2698	2428	2293	4424	10791	17049	18938	14541	6151	0	0	0
5	2206	1986	1875	3618	8825	12549	12046	5096	0	0	0	0

6	**2071**	**1864**	**1760**	**3396**	**7455**	**9161**	**4845**	**0**	**0**	**0**	**0**	**0**
7	**2803**	**2523**	**2383**	**2069**	**1570**	**1329**	**0**	**0**	**0**	**0**	**0**	**0**
8	**2514**	**2262**	**1923**	**1443**	**603**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
9	**2882**	**2334**	**1715**	**709**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
10	**2329**	**1630**	**660**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
11	**2243**	**865**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
12	**937**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**

This $\boldsymbol{L}$ is of course much smaller than the typical triangle analyzed for Warranty business, which might be of size 120 x 120 with contracts of term 72 or 84 months, but this should fit more comfortably on the reader's screen. We assume here that all contracts contributing to $\boldsymbol{L}$ have term 12 months, though some may expire earlier because of "miling out". Because of our assumption that contracts are written at the beginnings of months, just 12 columns are needed here.

Our known discontinuity in the earning pattern of the contracts being written might be expected to have been accompanied by a discontinuity in expected loss costs and therefore in either rates or loss ratios or both. For simplicity in this illustration we assume that the rates were adjusted to keep the loss ratios constant and that the effect is therefore buried in the growth of written exposures and is of no consequence to the model.

Here we identify report date with payment date and assume that our triangle $\boldsymbol{L}$ contains paid losses only, and that we have a separate triangle $\boldsymbol{P}$ of paid losses by incurral month versus payment lag, from which we have already estimated payment lag factors of (**0.3 0.4 0.2 0.1 0 0** ...) and the corresponding cumulative reported fractions **(0.3 0.7 0.9 1 1 1 ...)**.

2.5.2 Conventional loss development

From $\boldsymbol{L}$ alone we may apply conventional loss development (chain-ladder, loss weighted, no judgment adjustments) to obtain the following vector $\boldsymbol{f}$ of conventional lag factors:

0.0364 0.0301 0.0269 0.0420 0.0969 0.1524 0.1731 0.1525 0.1296 0.0985 0.0425 0.0190

The reader may confirm this calculation (and those to follow) using his or her preferred loss reserving software, or may consult Appendix B for the algorithms expressed as J language code.

These factors confound the effects of cancellations with the effects of the underlying loss emergence pattern and of the lag in reporting losses; they will not do to derive UPR factors. It is possible to correct for the effect of unreported losses at this stage while leaving the effects of cancellations and loss emergence pattern commingled; this yields an improved vector $\boldsymbol{f'}$:

0.0218 0.0197 0.0186 0.0308 0.0739 0.1219 0.1520 0.1501 0.1481 0.1316 0.0731 0.0585

This may be useful for a quick projection of future incurred losses, but is still not suitable for deriving SSAP 65 Test 2 UPR factors.

2.5.3 Exposure-adjusted loss development

To adjust for declining exposure as well as for unreported losses, we convert $\boldsymbol{L}$ into the triangle $\boldsymbol{R}$ of loss ratios to expected reported fractions of inforce exposure:

	1	2	3	4	5	6	7	8	9	10	11	12
1	0.0106	0.0106	0.0106	0.0212	0.0530	0.0848	0.1061	0.1061	0.1061	0.0955	0.0530	0.0424
2	0.0106	0.0106	0.0106	0.0212	0.0530	0.0848	0.1061	0.1061	0.1061	0.0955	0.0530	0
3	0.0106	0.0106	0.0106	0.0212	0.0530	0.0848	0.1061	0.1061	0.1061	0.0955	0	0
4	0.0106	0.0106	0.0106	0.0212	0.0530	0.0848	0.1061	0.1061	0.1061	0	0	0
5	0.0106	0.0106	0.0106	0.0212	0.0530	0.0848	0.1061	0.1061	0	0	0	0
6	0.0106	0.0106	0.0106	0.0212	0.0530	0.0848	0.1061	0	0	0	0	0
7	0.0119	0.0119	0.0119	0.0119	0.0119	0.0237	0	0	0	0	0	0
8	0.0119	0.0119	0.0119	0.0119	0.0119	0	0	0	0	0	0	0
9	0.0119	0.0119	0.0119	0.0119	0	0	0	0	0	0	0	0
10	0.0119	0.0119	0.0119	0	0	0	0	0	0	0	0	0
11	0.0119	0.0119	0	0	0	0	0	0	0	0	0	0
12	0.0119	0	0	0	0	0	0	0	0	0	0	0

The "noiselessness" of $\boldsymbol{L}$ and the differences between its first and last six months are obvious here. Developing $\boldsymbol{R}$ yields a vector $\boldsymbol{g}$ of lag factors adjusted for both declining exposure and unreported losses; these may be taken as a reasonable candidate for average earnings factors:

0.0171	0.0171	0.0171	0.0294	0.0720	0.1201	0.1515	0.1515	0.1515	0.1364	0.0758	0.0606

They differ from the previous vector $\boldsymbol{f}$ mainly in the early lags, where cancellations are concentrated.

Here, in applying chain-ladder, in averaging each column of development factors, the weights are the denominators (which are themselves loss ratios) multiplied by the numerator inforce exposures (see [2]). This properly recognizes the volume of experience contributed by each issue month, which otherwise would be flattened out when taking loss ratios.

2.5.4 Residual loss development

For residual loss development, we imagine that the actuary knows that the earnings pattern changes after six months – perhaps by applying a formula to the individual contracts – but that this knowledge is imperfect, so that the actuary's *a-priori* loss emergence patterns are not quite the same as the patterns underlying the actual data; we take his or her matrix $\boldsymbol{A}$ to be the following:

	1	2	3	4	5	6	7	8	9	10	11	12
1	0.0204	0.0204	0.0204	0.0408	0.0816	0.1224	0.1361	0.1361	0.1361	0.1224	0.0952	0.0680
2	0.0204	0.0204	0.0204	0.0408	0.0816	0.1224	0.1361	0.1361	0.1361	0.1224	0.0952	0.0680
3	0.0204	0.0204	0.0204	0.0408	0.0816	0.1224	0.1361	0.1361	0.1361	0.1224	0.0952	0.0680
4	0.0204	0.0204	0.0204	0.0408	0.0816	0.1224	0.1361	0.1361	0.1361	0.1224	0.0952	0.0680
5	0.0204	0.0204	0.0204	0.0408	0.0816	0.1224	0.1361	0.1361	0.1361	0.1224	0.0952	0.0680
6	0.0204	0.0204	0.0204	0.0408	0.0816	0.1224	0.1361	0.1361	0.1361	0.1224	0.0952	0.0680
7	0.0233	0.0233	0.0233	0.0233	0.0233	0.0465	0.0930	0.1395	0.1550	0.1550	0.1550	0.1395
8	0.0233	0.0233	0.0233	0.0233	0.0233	0.0465	0.0930	0.1395	0.1550	0.1550	0.1550	0.1395
9	0.0233	0.0233	0.0233	0.0233	0.0233	0.0465	0.0930	0.1395	0.1550	0.1550	0.1550	0.1395
10	0.0233	0.0233	0.0233	0.0233	0.0233	0.0465	0.0930	0.1395	0.1550	0.1550	0.1550	0.1395
11	0.0233	0.0233	0.0233	0.0233	0.0233	0.0465	0.0930	0.1395	0.1550	0.1550	0.1550	0.1395
12	0.0233	0.0233	0.0233	0.0233	0.0233	0.0465	0.0930	0.1395	0.1550	0.1550	0.1550	0.1395

For residual loss development we divide $\boldsymbol{R}$ by **A** to obtain the following triangle $\boldsymbol{R^*}$:

	1	2	3	4	5	6	7	8	9	10	11	12
1	0.5197	0.5197	0.5197	0.5197	0.6496	0.6929	0.7795	0.7795	0.7795	0.7795	0.5568	0.6236
2	0.5197	0.5197	0.5197	0.5197	0.6496	0.6929	0.7795	0.7795	0.7795	0.7795	0.5568	0
3	0.5197	0.5197	0.5197	0.5197	0.6496	0.6929	0.7795	0.7795	0.7795	0.7795	0	0
4	0.5197	0.5197	0.5197	0.5197	0.6496	0.6929	0.7795	0.7795	0.7795	0	0	0
5	0.5197	0.5197	0.5197	0.5197	0.6496	0.6929	0.7795	0.7795	0	0	0	0
6	0.5197	0.5197	0.5197	0.5197	0.6496	0.6929	0.7795	0	0	0	0	0
7	0.5102	0.5102	0.5102	0.5102	0.5102	0.5102	0	0	0	0	0	0
8	0.5102	0.5102	0.5102	0.5102	0.5102	0	0	0	0	0	0	0
9	0.5102	0.5102	0.5102	0.5102	0	0	0	0	0	0	0	0
10	0.5102	0.5102	0.5102	0	0	0	0	0	0	0	0	0
11	0.5102	0.5102	0	0	0	0	0	0	0	0	0	0
12	0.5102	0	0	0	0	0	0	0	0	0	0	0

Developing $\boldsymbol{R^*}$ (with LDFs weighted as described above) yields the residual lag factors $\boldsymbol{h}$:

0.0680	0.0680	0.0680	0.0680	0.0824	0.0889	0.1010	0.1010	0.1010	0.1010	0.0721	0.0808

Multiplying $\boldsymbol{h}$ by $\boldsymbol{A}$ yields a matrix of lag factors $\boldsymbol{G}$, adjusted for declining exposure, unreported losses, and expected emergence patterns:

	1	2	3	4	5	6	7	8	9	10	11	12
1	0.0153	0.0153	0.0153	0.0307	0.0743	0.1204	0.1518	0.1518	0.1518	0.1366	0.0759	0.0607
2	0.0153	0.0153	0.0153	0.0307	0.0743	0.1204	0.1518	0.1518	0.1518	0.1366	0.0759	0.0607
3	0.0153	0.0153	0.0153	0.0307	0.0743	0.1204	0.1518	0.1518	0.1518	0.1366	0.0759	0.0607
4	0.0153	0.0153	0.0153	0.0307	0.0743	0.1204	0.1518	0.1518	0.1518	0.1366	0.0759	0.0607
5	0.0153	0.0153	0.0153	0.0307	0.0743	0.1204	0.1518	0.1518	0.1518	0.1366	0.0759	0.0607
6	0.0153	0.0153	0.0153	0.0307	0.0743	0.1204	0.1518	0.1518	0.1518	0.1366	0.0759	0.0607
7	0.0176	0.0176	0.0176	0.0176	0.0214	0.0462	0.1048	0.1572	0.1747	0.1747	0.1248	0.1258
8	0.0176	0.0176	0.0176	0.0176	0.0214	0.0462	0.1048	0.1572	0.1747	0.1747	0.1248	0.1258
9	0.0176	0.0176	0.0176	0.0176	0.0214	0.0462	0.1048	0.1572	0.1747	0.1747	0.1248	0.1258
10	0.0176	0.0176	0.0176	0.0176	0.0214	0.0462	0.1048	0.1572	0.1747	0.1747	0.1248	0.1258
11	0.0176	0.0176	0.0176	0.0176	0.0214	0.0462	0.1048	0.1572	0.1747	0.1747	0.1248	0.1258
12	0.0176	0.0176	0.0176	0.0176	0.0214	0.0462	0.1048	0.1572	0.1747	0.1747	0.1248	0.1258

The average of these earnings factors, weighted by exposure and normalized to total 1, is:

0.0166	0.0166	0.0166	0.0237	0.0460	0.0807	0.1267	0.1547	0.16400	0.1569	0.102	0.0954

In this artificial example we know the earnings factors underlying the construction of $\boldsymbol{L}$:

	1	2	3	4	5	6	7	8	9	10	11	12
1	0.0152	0.0152	0.0152	0.0303	0.0758	0.1212	0.1515	0.1515	0.1515	0.1364	0.0758	0.0606
2	0.0152	0.0152	0.0152	0.0303	0.0758	0.1212	0.1515	0.1515	0.1515	0.1364	0.0758	0.0606
3	0.0152	0.0152	0.0152	0.0303	0.0758	0.1212	0.1515	0.1515	0.1515	0.1364	0.0758	0.0606
4	0.0152	0.0152	0.0152	0.0303	0.0758	0.1212	0.1515	0.1515	0.1515	0.1364	0.0758	0.0606
5	0.0152	0.0152	0.0152	0.0303	0.0758	0.1212	0.1515	0.1515	0.1515	0.1364	0.0758	0.0606
6	0.0152	0.0152	0.0152	0.0303	0.0758	0.1212	0.1515	0.1515	0.1515	0.1364	0.0758	0.0606
7	0.0169	0.0169	0.0169	0.0169	0.0169	0.0339	0.0847	0.1356	0.1695	0.1695	0.1695	0.1525
8	0.0169	0.0169	0.0169	0.0169	0.0169	0.0339	0.0847	0.1356	0.1695	0.1695	0.1695	0.1525

9	0.0169	0.0169	0.0169	0.0169	0.0169	0.0339	0.0847	0.1356	0.1695	0.1695	0.1695	0.1525
10	0.0169	0.0169	0.0169	0.0169	0.0169	0.0339	0.0847	0.1356	0.1695	0.1695	0.1695	0.1525
11	0.0169	0.0169	0.0169	0.0169	0.0169	0.0339	0.0847	0.1356	0.1695	0.1695	0.1695	0.1525
12	0.0169	0.0169	0.0169	0.0169	0.0169	0.0339	0.0847	0.1356	0.1695	0.1695	0.1695	0.1525

with normalized average

0.0161	0.0161	0.0161	0.0232	0.0444	0.0746	0.1159	0.1430	0.1611	0.1540	0.1258	0.1097

from which it becomes apparent that the fact that the actuary chose an *a-priori* earnings factor matrix slightly different from the actual earnings patterns underlying $\boldsymbol{L}$ had some effect, but small, on the final estimated earnings pattern. If the actuary had in fact used the actual earnings pattern as the *a-priori* pattern, the final estimated earnings pattern would replicate it exactly.

2.5.5 Tails

Now suppose we have only the last eight rows and the first eight columns of $\boldsymbol{L}$, but have all twelve columns of the last eight rows of $\boldsymbol{E^c}$ and of $\boldsymbol{A}$. Residual loss development will automatically append a tail to each row of the final projected factors, bringing that matrix to shape 8 x 12:

	1	2	3	4	5	6	7	8	9	10	11	12
5	0.0164	0.0164	0.0164	0.0329	0.0761	0.1245	0.1603	0.1603	0.1280	0.1152	0.0896	0.0640
6	0.0164	0.0164	0.0164	0.0329	0.0761	0.1245	0.1603	0.1603	0.1280	0.1152	0.0896	0.0640
7	0.0190	0.0190	0.0190	0.0190	0.0220	0.0479	0.1111	0.1666	0.1478	0.1478	0.1478	0.1330
8	0.0190	0.0190	0.0190	0.0190	0.0220	0.0479	0.1111	0.1666	0.1478	0.1478	0.1478	0.1330
9	0.0190	0.0190	0.0190	0.0190	0.0220	0.0479	0.1111	0.1666	0.1478	0.1478	0.1478	0.1330
10	0.0190	0.0190	0.0190	0.0190	0.0220	0.0479	0.1111	0.1666	0.1478	0.1478	0.1478	0.1330
11	0.0190	0.0190	0.0190	0.0190	0.0220	0.0479	0.1111	0.1666	0.1478	0.1478	0.1478	0.1330
12	0.0190	0.0190	0.0190	0.0190	0.0220	0.0479	0.1111	0.1666	0.1478	0.1478	0.1478	0.1330

2.5.6 Reference factors

Finally suppose that we replace our original $\boldsymbol{L}$ with a triangle of the same overall design but with random "noise" in each cell, so that the estimated earnings factors by residual loss development are somewhat erratic. By taking a weighted average of *residual* and *constant* earnings factors, we may obtain a weighted average of the *final* earnings factors from experience with the *a-priori* factors $\boldsymbol{A}$. It is convenient to define the weights given to experience as credibility-style factors $\boldsymbol{e}/(\boldsymbol{e}+k)$, where $\boldsymbol{e}$ is the total inforce exposure used in the estimation of each earnings factor and k is a constant selected, at the present time, by judgment. In the following example k = 500000. The "noisy" L is

	1	2	3	4	5	6	7	8	9	10	11	12
1	2853	1339	3536	475	7087	9807	3423	21225	9950	25801	3605	1722
2	2624	2877	2569	2911	9244	4584	21792	5848	17088	16688	1757	0
3	3899	3497	2663	7143	8355	26151	31550	24792	14522	1265	0	0
4	1585	3359	3849	6444	12068	16469	33179	12498	2681	0	0	0
5	2115	672	1858	2625	5589	20017	17969	610	0	0	0	0
6	3997	2477	298	1404	12537	10841	6107	0	0	0	0	0
7	5305	2551	3698	1272	824	1435	0	0	0	0	0	0
8	1341	3047	1572	245	1197	0	0	0	0	0	0	0
9	3674	2758	1665	1176	0	0	0	0	0	0	0	0
10	321	903	358	0	0	0	0	0	0	0	0	0
11	1146	807	0	0	0	0	0	0	0	0	0	0
12	596	0	0	0	0	0	0	0	0	0	0	0

The final earnings factors by pure residual loss development are:

	1	2	3	4	5	6	7	8	9	10	11	12
1	0.0173	0.0168	0.0175	0.0242	0.0676	0.1165	0.1746	0.1267	0.1215	0.2071	0.0524	0.0579
2	0.0173	0.0168	0.0175	0.0242	0.0676	0.1165	0.1746	0.1267	0.1215	0.2071	0.0524	0.0579
3	0.0173	0.0168	0.0175	0.0242	0.0676	0.1165	0.1746	0.1267	0.1215	0.2071	0.0524	0.0579
4	0.0173	0.0168	0.0175	0.0242	0.0676	0.1165	0.1746	0.1267	0.1215	0.2071	0.0524	0.0579
5	0.0173	0.0168	0.0175	0.0242	0.0676	0.1165	0.1746	0.1267	0.1215	0.2071	0.0524	0.0579
6	0.0173	0.0168	0.0175	0.0242	0.0676	0.1165	0.1746	0.1267	0.1215	0.2071	0.0524	0.0579
7	0.0199	0.0193	0.0201	0.0139	0.0195	0.0447	0.1206	0.1313	0.1398	0.2649	0.0862	0.1199
8	0.0199	0.0193	0.0201	0.0139	0.0195	0.0447	0.1206	0.1313	0.1398	0.2649	0.0862	0.1199
9	0.0199	0.0193	0.0201	0.0139	0.0195	0.0447	0.1206	0.1313	0.1398	0.2649	0.0862	0.1199
10	0.0199	0.0193	0.0201	0.0139	0.0195	0.0447	0.1206	0.1313	0.1398	0.2649	0.0862	0.1199
11	0.0199	0.0193	0.0201	0.0139	0.0195	0.0447	0.1206	0.1313	0.1398	0.2649	0.0862	0.1199
12	0.0199	0.0193	0.0201	0.0139	0.0195	0.0447	0.1206	0.1313	0.1398	0.2649	0.0862	0.1199

and the normalized average of these factors is:

0.0187	0.0181	0.0189	0.0187	0.0419	0.0781	0.1457	0.1291	0.1313	0.2380	0.0704	0.0910

The final earnings factors by weighted-average residual loss development are:

	1	2	3	4	5	6	7	8	9	10	11	12
1	0.0181	0.0177	0.0184	0.0283	0.0721	0.1194	0.1635	0.1307	0.1275	0.1631	0.0773	0.0638
2	0.0181	0.0177	0.0184	0.0283	0.0721	0.1194	0.1635	0.1307	0.1275	0.1631	0.0773	0.0638
3	0.0181	0.0177	0.0184	0.0283	0.0721	0.1194	0.1635	0.1307	0.1275	0.1631	0.0773	0.0638
4	0.0181	0.0177	0.0184	0.0283	0.0721	0.1194	0.1635	0.1307	0.1275	0.1631	0.0773	0.0638
5	0.0181	0.0177	0.0184	0.0283	0.0721	0.1194	0.1635	0.1307	0.1275	0.1631	0.0773	0.0638
6	0.0181	0.0177	0.0184	0.0283	0.0721	0.1194	0.1635	0.1307	0.1275	0.1631	0.0773	0.0638
7	0.0207	0.0202	0.0210	0.0162	0.0206	0.0454	0.1120	0.1342	0.1456	0.2069	0.1261	0.1310
8	0.0207	0.0202	0.0210	0.0162	0.0206	0.0454	0.1120	0.1342	0.1456	0.2069	0.1261	0.1310
9	0.0207	0.0202	0.0210	0.0162	0.0206	0.0454	0.1120	0.1342	0.1456	0.2069	0.1261	0.1310
10	0.0207	0.0202	0.0210	0.0162	0.0206	0.0454	0.1120	0.1342	0.1456	0.2069	0.1261	0.1310
11	0.0207	0.0202	0.0210	0.0162	0.0206	0.0454	0.1120	0.1342	0.1456	0.2069	0.1261	0.1310
12	0.0207	0.0202	0.0210	0.0162	0.0206	0.0454	0.1120	0.1342	0.1456	0.2069	0.1261	0.1310

with averages

0.0195	0.0191	0.0198	0.0218	0.0446	0.0799	0.1360	0.1326	0.1372	0.1865	0.1034	0.0997

2.4.3 UPR factors

The actuary will usually be interested in the incremental earnings factors $\boldsymbol{G}$ mainly for the fact that their reverse cumulative sums give factors to be applied to gross premium to obtain a UPR with desirable properties. In the context of UPR factors, I should mention the suggestion of John Sopkowicz that this technique might be even more usefully applied to claim counts versus inforce contract counts than to losses versus inforce premiums. This is an excellent point as it eliminates, as a source of noise, differences over time in rate adequacy, and we often do use analysis of claim count emergence patterns in related contexts, such as when using data with multiple terms to derive emergence patterns for each separate term in our All-Terms Factors model, or when analyzing frequency in ratemaking. The technique may indeed be applied whatever the definition of L and E.

Figure 1 compares the average UPR "curves" obtained from $\boldsymbol{L}$ by conventional loss development, conventional loss development adjusted only for unreported losses, exposure-adjusted

loss development, and residual loss development, with the actual average UPR curve implicit in the construction of $\boldsymbol{L}$.

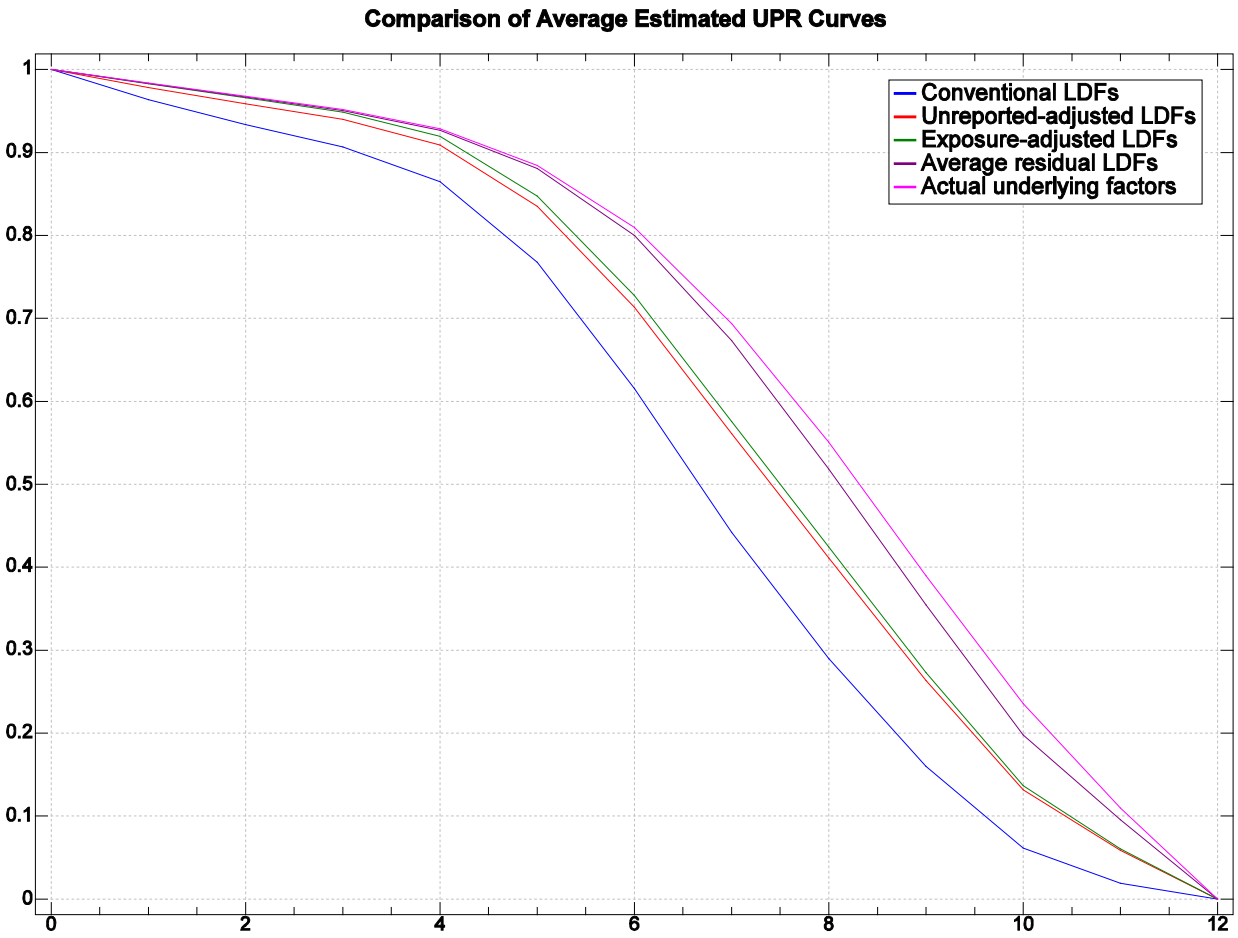


Fig. 1. UPR curves estimated from data in L and E

In this case residual loss development makes the UPR more conservative than simple exposure-adjusted loss development, reflecting the fact that the most recent six months' experience (which does not contribute to the estimated earnings factors by exposure-adjusted loss development at the later lags) is more conservative than the first six months' experience. The residual loss development shown here used the actuary's imperfect estimate of *a-priori* earnings factors; had it been perfect, the residual UPR curve would be identical to the actual curve.

Figure 2 compares the average UPR curves obtained from the last eight rows and first eight columns of $\boldsymbol{L}$, by conventional loss development with tails appended to each row to bring the projection to lag 12. Residual loss development facilitates such tails by providing the expected final shape in the a-priori earnings factor matrix, allowing the residual earnings factors to be extended with a constant tail.

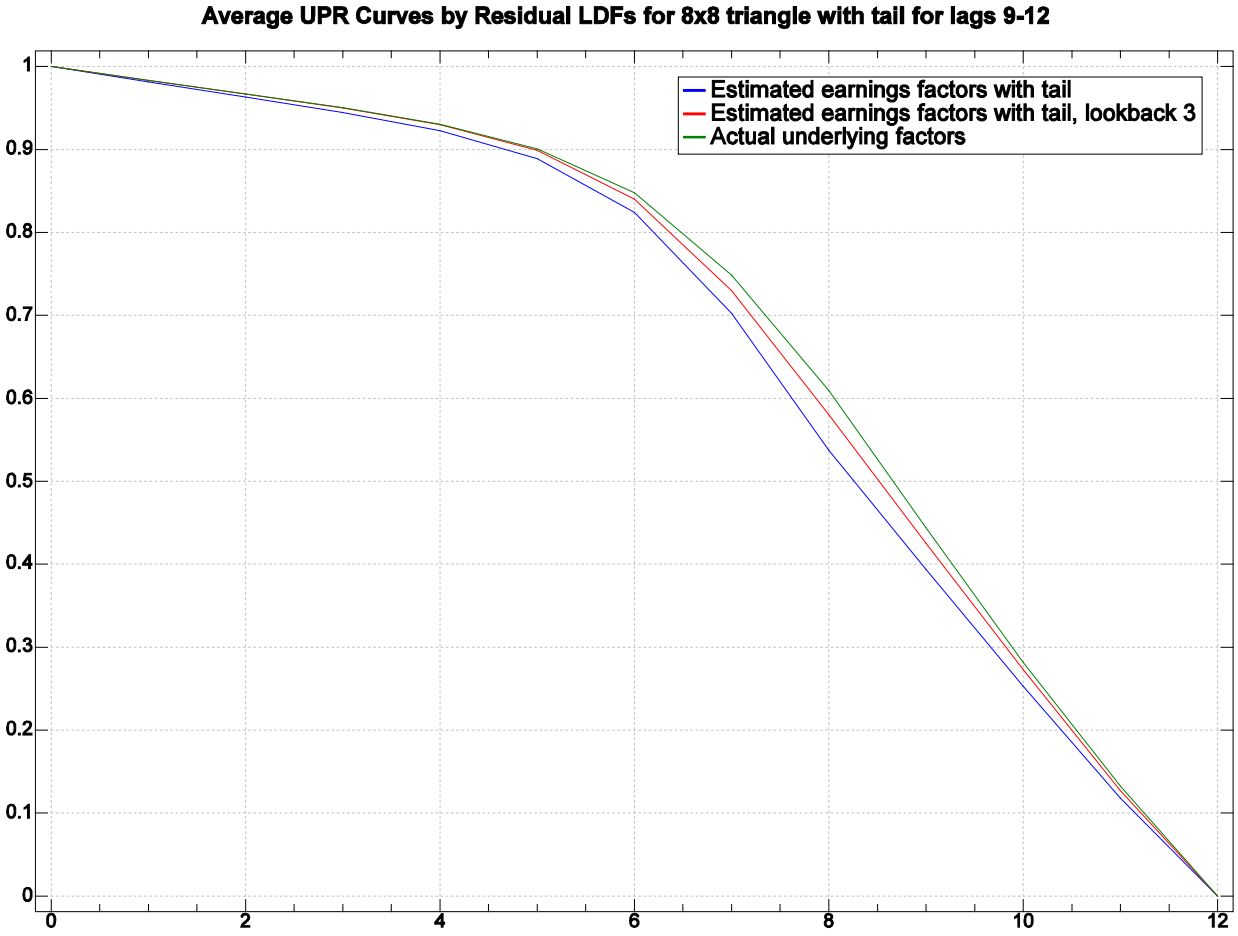


Fig. 2. UPR curve with tail after lag 8 months, compared with actual underlying UPR pattern

In the case at hand the tail is somewhat lighter than the actual last four lags (averaged across the last eight issue months), because the actuary's assumed *a-priori* earnings pattern for the last six months is lighter in the tail than the actual earnings pattern underlying $\boldsymbol{L}$. Because the earnings curve (the negative slope of the UPR curve) is normalized to total 1, the lighter tail causes earnings to increase at the earlier lags and makes the entire UPR curve less conservative. The differences between $\boldsymbol{A}$ and the actual earnings pattern of $\boldsymbol{L}$ were pronounced enough to show up in this illustration; in practice the actuary's formula-based $\boldsymbol{A}$ might be closer to the mark. It is also possible to adjust the weight in the tail to reflect the average residual earnings factors for a lookback period shorter than the entire known history of $\boldsymbol{L}$; we show this here with the "lookback 3" curve.

Figure 3 compares average UPR curves developed from a noisy triangle $\boldsymbol{L^r}$, together with the reference factors given by $\boldsymbol{A}$ and a normalized weighted average of the two. Residual loss development makes the use of reference factors particularly easy, since the weighting is done at the residual-factor stage and the complement of the weight is applied to constant factors.

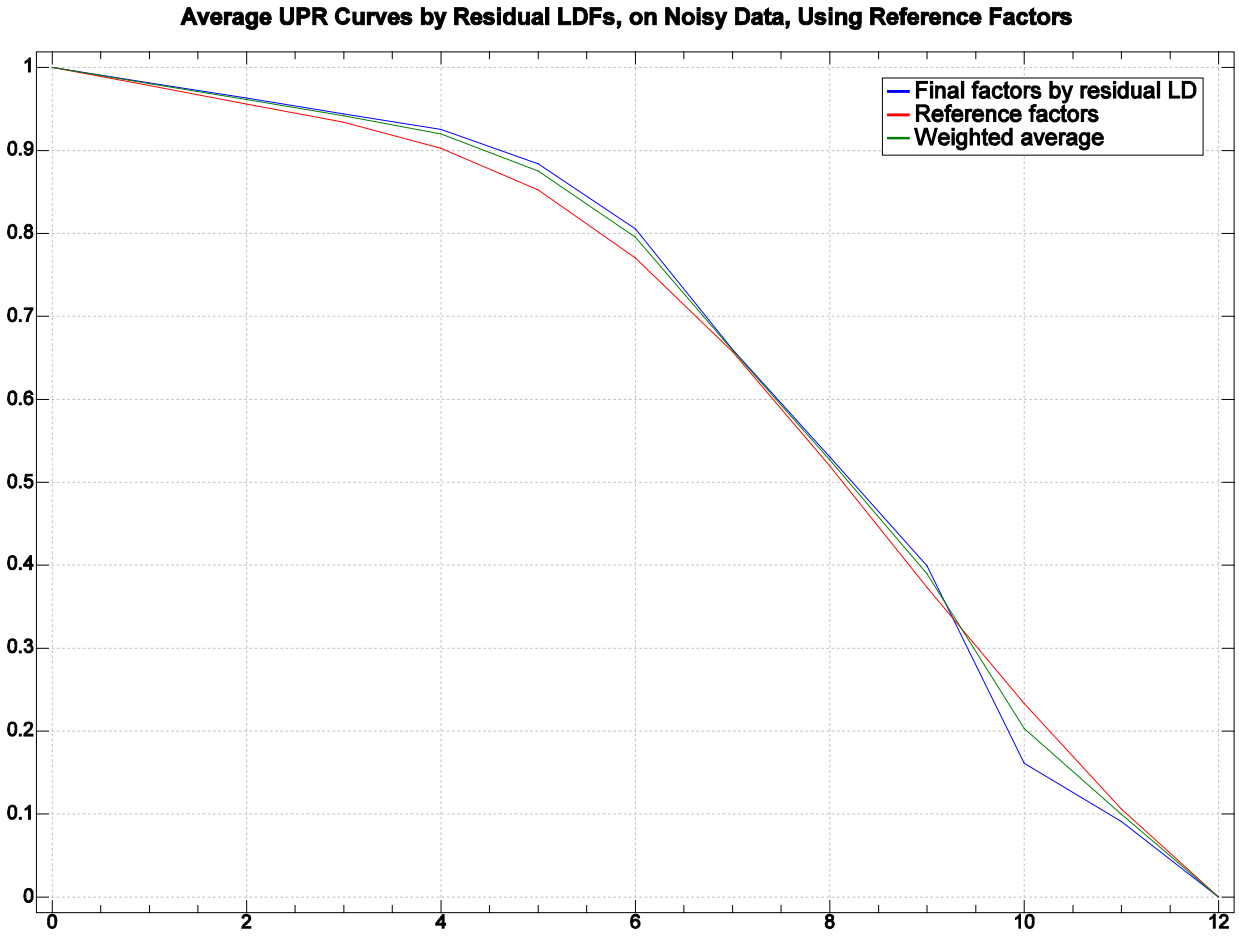


Fig. 3. UPR curve estimated from experience weighted against reference factors.

In this case the weighted average follows experience closely at the early lags, where several issue months contribute to the average development factors, and follows the reference factors at the later lags, where a smaller volume of exposure contributes to the averages.

3. CONCLUSIONS

Residual loss development is an enhancement to the adjustments required in the analysis of long-duration contracts for satisfaction of statutory UPR requirements and for providing accurate performance information to management, ownership, and regulators. It improves the analysis of imperfectly homogeneous segments, of immature segments requiring tail projections, and of segments with small volumes of experience. It may be possible to adapt this technique to policy or

accident year analysis of other lines of business, particularly those for which the lag to settlement has a practical maximum.

4. REFERENCES

[1] Kerper, John and Lee Bowron, An Exposure-Based Approach to Automobile Warranty Ratemaking and Reserving, CAS *Forum* 2007: 29-43

[2] Vaughan, Richard, The Unearned Premium Reserve for Warranty Insurance, CAS *E-Forum*, Fall 2014

APPENDIX A. Meaning of Satisfaction in Aggregate for SSAP 65 Test 2

In Section 1.2 we remarked that the requirement that SSAP 65 Test 2 be satisfied in aggregate for a company's long-duration contracts is subject to two interpretations, the *ratio of aggregates*, in which future and ultimate losses are considered only at the aggregate level, or the *aggregate of UPR's*, in which future and ultimate losses are considered separately for a collection of reasonably homogeneous subdivisions of the company's business, and the aggregate Test 2 criterion is built up from the separate criteria for the subdivisions, so that, in particular, satisfaction in aggregate may be guaranteed by satisfaction in detail.

To illustrate the difference, consider two company's books of business with in-force premiums P_0 and P_1, with future expected losses L_0 and L_1, and with expected ultimate losses U_0 and U_1, with all these quantities assumed to be greater than zero. Suppose each company's book of business satisfies SSAP 65 Test 2, with carried UPR equal to P_0L_0/U_0 for the first company and P_1L_1/U_1 for the second. This amounts to treating each company's business as internally homogeneous, or otherwise regarding its UPR as satisfying the ratio-of-aggregates definition. It is certainly reasonable to expect that a merger of the two companies would leave the combined carried UPR in compliance.

But the ratio of aggregates for the combined companies will lead to the same UPR as the aggregate of UPR's if and only if

$$P_0L_0/U_0 + P_1L_1/U_1 = (P_0+P_1)(L_0+L_1)/(U_0+U_1)$$

A bit of algebra shows that this is true if and only if

$$(U_1L_0 - U_0L_1)(P_0U_1 - P_1U_0) = 0$$

which holds if and only if the UPR factors are equal or the expected loss ratios are equal:

$$L_0/U_0 = L_1/U_1 \quad \text{or} \quad U_0/P_0 = U_1/P_1$$

Similarly the aggregate of the carried UPR's will be less than the ratio-of-aggregates Test 2 if and only if both the UPR factors and the loss ratios differ in the *same* direction, and the aggregate of the carried UPR's will be greater than the ratio-of-aggregates Test 2 if and only if both the UPR factors and the loss ratios differ in *opposite* directions.

So the two definitions do not produce the same result except when the expected loss ratios are equal (a reasonable assumption when combining contract-by-contract UPR's within a homogeneous subdivision, but often not valid across subdivisions) or when the UPR factors are identical (not common). But the difference will not often be material. An fairly extreme example might be a book of mature business with average UPR factor about 50% and loss ratio about 70%, to which is added an incipient book of business with UPR factor 1.00 and loss ratio 100%; if the inforce premium for the new book is one-fifth that of the mature business, then the ratio-of-aggregates UPR will be

about 4.8% greater than the aggregate of the separate UPR's. If the mature and incipient business were reversed, then the ratio-of-aggregates UPR would be about 3.8% less than the aggregate of the separate UPR's.

In principle the P_i's are known while the L_i's and U_i's must be estimated. Curiously, it may be harder to fix the value of P_i than to estimate L_i and U_i. More precisely, P_i depends on the definition of "in force"; which may attempt to exclude expired contracts or may exclude only cancelled contracts. Common estimators of L_i are indifferent to this choice, as long as each contract is considered in force at least as long as it is possible for a loss to emerge; they will simply return 0 for the estimated earnings factor at any later lag. If P_i is extended to P^*_i by inclusion of some expired contracts with 0 future losses, then $L^*_i = L_i$ while $U^*_i=(P^*_i/P_i)U_i$, so

$$(\text{Test 2})^*_i=P^*_iL^*_i/U^*_i = P_iL_i/U_i = (\text{Test 2})_i$$

But when we combine inhomogeneous blocks of business to get a ratio-of-aggregates UPR, the result will depend on the definition of inforce; for example

$$(\text{Test 2})^*_{agg} = (P^*_0 +P^*_1)(L^*_0 + L^*_1)/(U^*_0 + U^*_1) = (K_0P_0 + K_1P_1)(L_0+L_1)/(K_0U_0 + K_1U_1)$$

where $K_i = P^*_i/P_i$, and this will not, in general, equal

$$(P_0 + P_1)(L_0 + L_1)/(U_0 + U_1) = (\text{Test 2})_{agg}$$

For these reasons we conclude that the only reasonable interpretation of SSAP 65, where a company's business is subdivided into natural and reasonably homogeneous segments, is that the Test 2 criteria may be summed across such subdivisions to obtain the aggregate Test 2 criterion.

APPENDIX B. J Code

The illustrations in this paper were generated by a simple model of the procedure coded in the language J. To make the calculations replicable and to allow the reader to experiment with variations we include the code here.

A bit of a digression is in order here. About 60 years ago the late Kenneth Iverson, then teaching applied mathematics at Harvard, introduced the language APL in a small book entitled "A Programming Language"; this book later won Dr. Iverson the Turing Award. APL is basically a linearized mathematical notation convenient both for conveying algorithms in print and for parsing by a computer as a high level interpreted language. Dr. Iverson later joined IBM and in 1962 brought out the first APL interpreter, on an IBM mainframe. The language was extended by IBM and several other companies and was adopted by many users in the financial and actuarial communities. At one time it was not unusual for papers in the Proceedings of the Casualty Actuarial Society to include a page or two of APL code to illustrate their algorithms. This was especially convenient because of the extreme conciseness of APL.

APL takes arrays such as vectors and matrices as its fundamental objects and, partly for this reason, is admirably suited for many actuarial models such as life contingencies and P/C loss development. Many of APL's primitive functions are structural operations on arrays, not all of which are conveniently expressible in conventional mathematical notation. About 30 years after developing APL, Dr. Iverson, joined by Roger Hui, undertook to systematize the theory of operations on arrays and to create a new language, what APL would have been if he had it to do over again. The result is the language J. This language is a *tour de force*: elegant, concise, comprehensive, uncompromisingly systematic. The J interpreter and development environment are in the public domain, available free for all common computer platforms at the web site jsoftware.com, and are supported by Mr. Hui and by a large community of users.

The author recognizes that another language, R, has become a de facto standard for much work in CAS publications, and for good reason: R has an enormous library of contributed statistical packages, tested and validated, including several specifically actuarial packages. R also operates on arrays, and borrows some ideas from the original APL, but is not nearly as simple, consistent, or thorough in managing arrays as is J. For this reason actuaries looking for a language both to express their thoughts and to build libraries of models – or even just to prototype models eventually to be ported to other languages – would do well to consider J. A few days' experimentation, using the J interpreter interactively and writing small programs, will suffice to get started.

The following code is a simple model of residual loss development and at the same time of conventional loss development, loss development adjusted for unreported losses, and loss development adjusted for declining exposures, all of which may be treated as special cases. The

code is in the form of a script – a simple text file – which may be edited by any text editor and parsed by the J interpreter. Comments are preceded by the J word NB.

This is only a fragment of a complete library for triangle analysis, which would also include functions for managing contract and claims data, producing printed reports, etc., along with additional stochastic and deterministic estimators. In anticipation of these purposes this model defines a triangle as a structure containing not only the matrix proper but also additional information such as cumulative and lag status, cell sizes, and dates. The script starts out with a description of the triangle structure, sets the print precision, and loads a couple of addon packages that will be needed.

```
NB. This script contains operations on triangles of actuarial data.

NB. By triangle we mean a matrix of losses or similar quantities, together with
NB. structural and date information, represented as a vector of boxes containing:
NB. a.  Numeric matrix proper
NB. b.  1 if lagged, 0 otherwise [default: 1]
NB. c.  1 if cumulative, 0 otherwise [default: 1]
NB. d.  cell size on first axis [default: 1]
NB. e.  cell size on last axis (must be <=(d)) [default: (d)]
NB. f.  earliest month on first axis, as yyyymm
NB. g.  latest known month on last axis, as yyyymm
NB. h.  latest known or projected month on last axis, as yyyymm

NB. It is assumed that the initial cells along the first axis and the final known
NB. cells along the second axis are complete; if the cell size is greater than 1,
NB. the last cell on the first axis and the first cell on the second axis may be
NB. fragments.

NB. The triangle is so called because its known portion has three "corners"; when
NB. completed to include future periods it takes the shape of a rectangle.

(9!:11)10
require 'plot'
require 'stats/distribs'
```

Next we define some small functions useful in an actuarial context. Notice that the primitive objects of J itself are spelled with ASCII punctuation marks or with one (or occasionally more) letters or punctuation marks followed by '.' or ':'. This makes it impossible to overwrite them with user-defined objects, the names of which cannot include punctuation. These small functions are defined tacitly, that is, with no explicit reference to their left and right arguments x and y, though we refer to the arguments using x and y in the comments. J has extraordinary flexibility in the composition of functions, which facilitates functional programming of this type.

```
NB. Trimming and extending arrays
TrimB=: ]}.~[:+/[:*./\e.~     NB. Trim leading items found in x from vector y
TrimE=: TrimB&.|.             NB. Trim trailing items found in x from vector y
TrimV=: [ TrimE TrimB         NB. Trim items in x from both ends of vector y
TrimX=: (_1{.]),~]TrimE~_1{.] NB. Trim extra copies of last item from end of y
ExtE=: [{.],[#_1{.]           NB. Extend y to length x with copies of last item
ExtB=: ([:-[){.([#1{.]),]     NB. Extend y to length x with copies of first item
MinL=: ]{.~[>.[:#]            NB. Overtake y to give it a minimum length x
RowMat=:,.&.|:                NB. Make vect y into 1-row mat; leave mat unchanged
```

```
NB. Delimited character vectors and arrays of boxed character vectors
UnD=:[:<;._2 }.,({.~:{:)#{.   NB. CV y delimited by first char to vector of boxes
UnD1=:[:UnD,                  NB. CV y delimited by x to vector of boxes
UnD2=: [:>[ UnD1&.> LF UnD1 ] NB. CV y delimited by x and LF to matrix of boxes
UnCSV=: ','&UnD2              NB. CV y from .csv format to matrix of boxes
ToD1=:[:}:[:;[,&.>~[:":&.>]   NB. Vector y of boxes to CV delimited by x
ToD2=:[:;LF,&.>~([:<[:;ToD1)"1  NB. Matrix y of boxes to CV delim by x and LF
ToCSV=: ','&ToD2              NB. Matrix of boxes to .csv format

NB. Operations on months represented as numeric yyyymm:
ThisM=: 3 : '10000 100 #. 2{.6!:0 '''''
IncrM=: 4 : '10000 100#.0 1+10000 12#:(x-1)+10000 12#.10000 100#:y'
PrevM=: _1&IncrM
FollM=: 1&IncrM
LastM=:PrevM@:ThisM
NextM=:FollM@:ThisM
DiffM=: -&(10000 12&#.@:(10000 100&#:)@:<:)

NB. Other supporting functions
Fill=:([:#[){.],[}.~[:#[:,]   NB. Fill vector y with defaults from vector x
Round=: ([:<.0.5+]) : ([*[:<.0.5+%~)  NB. Round y to nearest multiple of x
```

Now we come to some functions that manipulate triangle of the specified structure. These functions are defined explicitly, as a series of lines which may reference the arguments x and y explicitly. The expression 3 : 0 specifies that a monadic function (i.e., with right argument only) or an ambivalent function (may have a right argument only, or both arguments) is defined by the following code up to the first line consisting of a single right parenthesis. The meaning of various control structures should be evident.

```
TriDflts=: 3 : 0  NB. Complete triangle structure y with default values
if. 0=L.y do. y=.,<y end.
't lg cm c0 c1 e k p'=.((0 0$0);1;1;1;1;0;(LastM '');0) Fill y
c1=.c1<.c0
if. e=0 do. e=.>.(1-c1*{:$t) IncrM k end.
if. p=0 do.
    if. lg do.
        if. 3=#$t do. xx=.({:$t)<:((c0%c1)*i.0{$t)+/((c0%c1)*i.1{$t)+/i.2{$t
        else. xx=.({:$t)<:((c0%c1)*i.#t)+/i.{:$t
        end.
        p=.((-.*./0=,t*xx)*c1*({:$t)-1) IncrM k
             NB. If cells below diag all 0, dflt p for lagged tri equals k
    else. p=.(c1*({:$t)->.(1+k DiffM e)%c1) IncrM k
             NB. Default p for non-lagged tri is determined by its length
    end.
end.
z=.t;lg;cm;c0;c1;e;k;<p
)


LagTri=: 3 : 0  NB. Convert y from date to lag triangle
z=.'t lg cm c0 c1 e k p'=.TriDflts y
if. lg=0 do.
    if. 3=#$t do. z=.((i.#t)|.!.0"0 2 ((c0%c1)*i.1{$t)(|.!.0"0 1)"2
t);1;cm;c0;c1;e;k;<p
    else. z=.(((c0%c1)*i.#t)|.!.0"0 1 t);1;cm;c0;c1;e;k;<p
    end.
end.
)
```

```
UnlagTri=: 3 : 0  NB. Convert y from lag to date triangle
z=.'t lg cm c0 c1 e k p'=.TriDflts y
tt=.((>.1+p DiffM e)%c1){."1 t
if. lg=1 do.
    if. 3=#$t do. z=.((-(c0%c1)*i.1{$t)(|.!.0"0 1)"2 (-i.#t)|.!.0"0 2
t);0;cm;c0;c1;e;k;<p
    else. z=.((-(c0%c1)*i.#t)|.!.0"0 1 tt);0;cm;c0;c1;e;k;<p
    end.
end.
)


CumTri=: 3 : 0  NB. Make triangle y cumulative
tri=.'t lg cm c0 c1 e k p'=.TriDflts y
if. cm do. z=.tri
elseif. lg do. z=.(<1) 2}(<(1 KnownPart tri)*+/\"1 t) 0}tri
elseif. 1 do. z=.(<1) 2}(<+/\"1 t) 0}tri
end.
)


DiffTri=: 3 : 0  NB. Make triangle y incremental
tri=.'t lg cm c0 c1 e k p'=.TriDflts y
if. -.cm do. z=.tri
elseif. lg do. z=.(<0) 2}(<(1 KnownPart tri)*(}.-}:)"1 (0,.t)) 0}tri
elseif. 1 do. z=.(<0) 2}(<(}.-}:)"1 (0,"1 t)) 0}tri
end.
)


KnownTri=: 3 : 0  NB. Known, or known or projected, part of triangle y
NB. x is 0 for known part, 1 for known or projected part [default: 0]
0 KnownTri y
:
tri=.'t lg cm c0 c1 e k p'=.TriDflts y
if. lg do. z=.($t)(]0}~[:<[{.[:>0{]) LagTri x KnownTri UnlagTri tri
else. z=.((<.(0>.1+(x{k,p) DiffM e)%c1){."1 t);lg;cm;c0;c1;e;k;<x{k,p
end.
)


KnownPart=: 3 : 0  NB. Flags known, or known and projected, part of triangle y
NB. x is 0 for known part, 1 for known and projected part [default: 0]
0 KnownPart y
:
tri=.'t lg cm c0 c1 e k p'=.TriDflts y
($t){.>0{x KnownTri (<([=])t) 0}tri
)


BandTri=: 4 : 0  NB. Band of x diagonals of triangle y, lagged
tri=.'t lg cm c0 c1 e k p'=.LagTri TriDflts y
z=.t*(KnownPart tri)*.-.KnownPart (<(-(x<.#t)*c0) IncrM k) 6}tri
)
```

Now the main function for the purposes of this paper, and the source of most of the illustrated factors::

```
LDFs=: 3 : 0  NB. LDF's, etc, by residual (or simpler) loss development
NB. y is (loss triangle in standard format);(inforce exposures);(incremental
```

```
NB.   vector of report lag factors, OR triangle of losses by incurral month
NB.   versus report lag) [default for last two items: 1;,1]
NB. x is (depth);(vector or matrix proportional to a-priori lag factors);
NB.   (cred constant)(maximum tail lookback);(Boolean vector with flags for
NB.   unreported adjustment, inforce exposure adjustment, and residual adjustment);
NB.   (maximum length of vector of report lag factors, if calculated)
NB.   [default: _;(($%){:$>0{>0{y);0;_;0 0 0;6]
NB. z is (matrix with rows ldfs, uldfs, cfs, residual lag factors);(matrix
NB.   of final lag factors by accident period versus lag);<(vector of incremental
NB.   report lag factors)

NB. The credibility constant K produces credibility factor E%E+K, where E is
NB. the sum of the denominator exposures at each lag, before a-priori adjustment.
NB. For indications from experience only, use a-priori lag factors of length no
NB. greater than the experience and set the credibility constant to zero.
NB. To append a tail to the indications of experience, use a-priori lag factors
NB. extending to the end of the tail and set the credibility constant to zero.

NB. For a weighted average of experience and a-priori factors, set the credibility
NB. constant >0.  If the a-priori pattern includes factors beyond the end of the
NB. experience data, these will automatically be given full weight, thus appending
NB. a tail, while the earlier factors will be a weighted average.

'' LDFs y
:
'tri exp rpt'=.('';'';,1) Fill y
'd ap K lkb flags supp'=.(_;'';0;_;'';6) Fill x
tri=.'t lg cm c0 c1 e k p'=.DiffTri LagTri KnownTri TriDflts tri
'rptadj expadj apadj'=.0 0 0 Fill flags
'm n'=.$t
if. 0 e.$ap do. ap=.($%){:$>t end.
ap=.(%+/)"1 n MinL"1 m ExtE RowMat ap
N=.{:$ap                      NB. Length including tail
if. 0 e.$exp do. exp=.m$1 end.NB. Default exposure is constant
exp=.N ExtE"1 ,.exp           NB. Matrix of exposures
if. 2=#$>{.rpt do. rpt=.(%+/)supp{.RLFs rpt;'' end.  NB. Report lag factors
rfs=.(i.m)|.!.0"0 1 (m,n)$+/\.(-n){.|.rpt  NB. Matrix of reported fractions
E=.((m,n){.exp^expadj)*(rfs^rptadj)*(m,n){.ap^apadj  NB. Adjusted exposures
lrs=.t%E                      NB. Loss ratios to adjusted exposures
lrstri=.CumTri (<lrs)0}tri
w=.}."1 (n{."1 exp^expadj)*rfs^rptadj  NB. Adj to weights for avg ldf's
nums=.w*}."1 d BandTri lrstri NB. Numerators of ldf's
dens=.w*}:"1 d BandTri (<(-c1) IncrM k) 6}lrstri  NB. Denominators of ldf's
ldfs=.((+/nums)%+/dens),1     NB. Weighted average ldf's
uldfs=.*/\.ldfs               NB. Ultimate ldf's
cfs=.%uldfs                   NB. Completion factors
lfs=.(}.-}:)0,cfs             NB. Lag factors
xx=.d BandTri n{."1 exp       NB. Inforce exposure used in obtaining lag factors
Z=.N{.(+/xx)([%+)K            NB. Credibility factors by lag
rfs=.(n$(+/%#)lfs),(N-n)$((+/%#)(-lkb<.n){.lfs)  NB. Reference lfs
   NB.  These are constant at %#lfs for first n factors, then equal the
   NB.  average lfs over the lookback period for any tail.
LFS=.(%+/)(Z*N{."1 lfs)+"1 (1-Z)*"1 rfs  NB. Cred-adj lag factors
flfs=.(%+/)"1 LFS*"1 ap^apadj NB. Final lag factors
z=.(ldfs,uldfs,cfs,:lfs);flfs;<rpt
)
```

This carries the model through the earnings factor stage which is the focus of this paper. Additional functions may be included to estimate loss ratios by issue month, earnings factors from a single issue month versus incurral lag triangle, and future losses cell by cell, to project persistency, cancellations, and refunds, and so forth, and to use these results for such purposes as financial projections, estimates of ultimate loss and refund ratios, and tests of UPR factors for satisfaction of SSAP 65,

beyond the scope of this paper. But the code shown above should serve to document the residual loss development algorithm for the actuary with some knowledge of J.