Pricing Multiple Property Cover Based on a Bivariate Lognormal Distribution

by Farrokh Guiahi

ABSTRACT

This paper applies a bivariate lognormal distribution to price a property policy with property damage and business interruption cover subject to an attachment point, separate deductibles, and a combined limit. Curve-fitting tasks for univariate probability distributions are compared with the tasks required for multivariate probability distributions. This is followed by a brief discussion of the data used, data-related issues, and adjustments. Selection of a parametric multivariate size of loss distribution, estimation of the parameters of the selected distribution, and goodness of fit are discussed in reference to the bivariate lognormal distribution. Finally, an algorithm is provided for estimating the average loss cost based on a bivariate lognormal distribution by taking into consideration the loss-sensitive features of the policy.

KEYWORDS

Loss-sensitive features, multivariate statistical analysis, bivariate lognormal

1. Introduction

The traditional approach to pricing a property cover—the approach used by underwriters—relies upon using tables of rates and scales. Scales are used to provide credit for the deductible amounts. Apart from the papers of Salzmann [8] and Ludwig [6], there is relatively little published material on the subject of property rating by CAS actuaries. The aforementioned papers discussed the construction of scales and use of the scales for the purposes of rating simple property covers. Casualty actuaries rely on the use of a frequency and severity approach as the basis for pricing insurance products, and are normally less engaged in pricing property covers. There is no paper in the *Proceedings of the Casualty Actuarial Society* (*PCAS*) that specifically guides actuaries in pricing multiple property covers with loss-sensitive features. Hence the reason for writing this paper is to address this need, using a frequency and severity approach.

For policies providing multiple cover, the loss may arise from distinct sources. For example, given a policy providing a property damage (PD) cover as well as a time element (TE) or business interruption cover, losses may arise from damage to a property location and/or loss of income due to business interruption. The claim payment by the insurer is tempered by the loss-sensitive features of the policy. The loss-sensitive provisions considered here are due to the interaction of separate deductibles, an attachment point, and a single limit upon insured losses.

The expected loss cost or pure premium is based on determining separately the frequency and the severity. These two elements are multiplied to determine the pure premium. By loading the pure premium for expenses and profit, the technical premium for a policy is determined. This paper concentrates mainly on the severity component of the pure premium—loss per claim—referred to as the average loss cost (ALC). ALC is calculated using a suitable size of loss distribution (SOLD) as well as the specification of a claim payment function (CPF). A CPF serves to constrain loss payments according to loss-sensitive features of the policy such as the deductible and the limit. Here the term ALC is used synonymously with the expected value of the CPF.

The following outline describes Sections 2 through 7 of this paper: In Section 2, curve fitting for the univariate and multivariate probability distributions are briefly compared. In Section 3, data and data related issues are discussed. Section 4 addresses the selection, estimation, and fit of a bivariate distribution to the sample losses. Section 5 is a discussion of the complexity of the CPF for a multiple cover policy. In Section 6, an algorithm is given for estimating the ALC based on a bivariate lognormal distribution. Finally, Section 7 provides some concluding statements.

2. Comparison of univariate and multivariate curve-fitting tasks

An important aspect of pricing an insurance cover is the consideration of a suitable SOLD. Tasks related to the determination of a SOLD are selection, estimation, fit, and implementation. In this section, curve-fitting tasks for univariate distributions are briefly compared to those for multivariate distributions.

Losses emanating from one source may be modeled by univariate probability distributions. In these cases, curve fitting is accomplished under a univariate setting. However, when there are multiple distinct sources of loss, as for instance in losses arising from a multiple cover policy, then the multivariate probability distributions are better suited to represent the SOLD. In these latter cases, curve fitting is done under a multivariate setting.

A multivariate approach is required in order to properly price policies with loss-sensitive features. As an example, consider the following:

Let Y_1, a random variable, denote a PD loss arising from a property cover during an exposure period, and let random variable Y_2 denote an individual TE loss from a property cover; then random variables or distributions of interest are:

(a) a univariate distribution for Y_1,
(b) a univariate distribution for Y_2,
(c) a univariate distribution for the sum $Y_1 + Y_2$, and
(d) a bivariate distribution for

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}.$$

Having information about the bivariate distribution Y, a 2×1 column vector—case (d) above—enables one to determine the marginal distributions, i.e., the cases (a) and (b) above, as well as the sum (transformation) case (c). Having information about the three univariate distributions—cases (a), (b), and (c)—may not suffice to price multiple covers with an attachment point, separate deductibles, and a combined limit. This situation is better understood when one examines the CPF as described in Section 5. Hence, bivariate distributions, case (d) above, play a vital role in pricing multiple covers.

Curve fitting methodology with regard to a SOLD normally involves the following tasks:

(a) selection of suitable parametric probability distributions to represent loss or losses,
(b) estimation of the parameters of the selected distributions based on historical loss experience,
(c) evaluation of the goodness of fit, and
(d) computation of the statistics related to the fitted distribution function such as the mean, the standard deviation, and the ALC.

A brief comparison of the above tasks for univariate probability distributions to those required for multivariate distributions is made below.

Regarding (a), there have been many univariate parametric distributions referenced in the actuarial publications as potential candidates for the SOLD. The list includes lognormal, Pareto, and Weibull. By comparison, the use of multivariate SOLD in the *PCAS* is less common. Multivariate normal is a popular distribution with statisticians. In finance, multivariate normal has been used to represent the joint distribution of stock returns as well as a model for pricing compound options; see Jarrow and Turnbull [3]. Multivariate normal can be applied in the insurance field after suitable transformation of loss components.

Regarding (b), the estimation of parameters, the maximum likelihood method may be utilized. However, in the case of multivariate distributions, the number of parameters to be estimated is much larger. For example, in the case of the lognormal family of distributions, the parameters needed are:

Distribution	*Number of Parameters to be Estimated*
Univariate Lognormal	2
Bivariate Lognormal	5
Multivariate Lognormal with 4 loss components	14

Apart from the need to estimate many more parameters, there is the issue of the required sample size. This is referred to as the "curse of dimensionality" problem. The sample size needed to fit a multivariate function grows exponentially with the number of variables, i.e., higher-dimension spaces are inherently sparse. Larose [5] provides an example: "The empirical rule tells us that in one dimension, about 68% of normally distributed variates lie between 1 and -1, whereas for a 10-dimensional multivariate normal distribution, only 0.02% of the data lie within the analogous hypersphere."

Regarding fit, (c), in univariate situations there are well-known procedures such as the Kolmogorov-Smirnov test for evaluating the fit. Many of these univariate fit procedures do not have corresponding multivariate counterparts. Hence, more innovative schemes are needed to assess the goodness of fit.

Finally, with regard to (d), the computation of ALC, in univariate cases, can be expressed in terms of standard available functions in many instances. For example, when the SOLD is a univariate lognormal, the computation of the ALC, a single limit integral, can be expressed in terms of an exponential function and a cumulative distribution function (cdf) of univariate normal (see Section 5). However, as will be shown in Section 6, when the SOLD is a bivariate lognormal distribution, then the computation of the corresponding ALC, a double integral, is a more difficult task requiring a tailor-made solution.

To summarize, curve fitting in a multivariate setting is more complex. There are more parameters to be estimated, and one needs larger sample sizes. Innovative procedures are needed for evaluating the fit and computing derived statistics based on multivariate probability distributions.

3. Data and data related issues

Any actuarial study should be based on well-defined objectives. Having established the objectives of the study, the next step usually involves an analysis of suitable data to shed light on the problem at hand. In order to estimate frequency and severity components of pure premium, claim and policy data are needed. Common concerns regarding the data are the volume of data, availability of relevant attributes, and the quality of data.

The analysis performed upon the data is dependent on the knowledge of the team involved in areas such as statistical modeling, actuarial science and insurance.

The data referenced in this paper was collected for a specific property project. For competitive reasons, the data used in that property project were altered in order to be used for illustrative purposes in this paper. Therefore, the estimated parameter values cannot be used as actual parameters or as benchmarks.

Before embarking on a curve fitting process, some preprocessing (cleaning) of the data is usually warranted. The tasks to be undertaken depend on the nature of the available data as well as the scope of the project. Thus, these preparatory tasks vary considerably depending upon the prevailing circumstances.

The operations performed on the data used in this paper were as follows: (a) exclusion of certain claims due to either incompleteness of the data, or being out of the scope of the project, (b) adjustment of the data, and (c) summarization of the data in order to gain an overview of the data.

The exclusion of certain claims was based on the following criteria:

(a) The incurred loss amount was negative.

(b) The claim belonged to a class that will not be underwritten in the future by the company.

(c) The claim arose from blanket written risks where a single limit may be applicable to buildings at different locations (in these instances, it is not possible to associate a PD loss amount with its corresponding building value amount, due to incompleteness of information gathered).

(d) The claim was due to a catastrophe cover, a CAT loss, CAT losses were modeled separately.

(e) The claim was a boiler machinery (BM) claim, which was also analyzed separately.

Regarding the adjustment of the data, the data were collected on a ground-up basis. Hence, losses were measured from the first dollar and grossed up in those cases where the company had less than a 100% share (not fully participating in the risk). For the purpose of curve fitting, the PD and TE losses were trended so they would be on a current level basis. Most losses considered were closed (paid), especially for the older accident years. The evaluation date was also subsequent to the latest accident year considered. Hence there were relatively few open claims left in the data

Table 1. Summary of total unadjusted ground up losses. Line of business: Property, non-CAT, non-BM losses

PD\TE	+	NA
+	$S(+,+)$ 1.827 B, 1.251 B	$S(+,\mathrm{NA})$ 1.774 B
NA	$S(\mathrm{NA},+)$ 0.065 B	$S(\mathrm{NA},\mathrm{NA})$

PD: Property Damage
TE: Time Element (Business Interruption)

set. The property claims studied tended to settle quickly (average time to close was about 13 months). The remaining open claims were not developed individually to an ultimate basis.

To gain an overview of the data used, the data was organized as tables. Tables 1 and 2 were constructed to provide some insight with regard to frequency and severity. The dollar amount of the losses pertains to loss figures prior to adjustment for trend.

A few remarks with regard to these two tables are in order. Table 1 is a two-way table used to summarize information with regard to severity. It consists of four cells referred to as $S(+,+)$, $S(+,\mathrm{NA})$, $S(\mathrm{NA},+)$, and $S(\mathrm{NA},\mathrm{NA})$.

The cell $S(+,+)$ arose from individual claims, where both the PD loss amount and the TE loss amount were **strictly** positive. The total loss figures were 1.827 billion dollars for the PD losses and 1.251 billion dollars for the TE losses over the period used in the study.

The losses contributing to the cell $S(+,\mathrm{NA})$ had a PD loss amount component that was **strictly** positive in each instance. However, the TE field accompanying the PD loss was populated by either a blank or a zero. For this cell, it was not possible to identify correctly the reasons for having a blank or zero value. Thus, it was not possible to identify correctly if the PD loss came from a PD and TE cover policy with no accompanying TE loss, or if the PD loss arose from a PD-only cover policy. Losses contributing to the cell $S(+,\mathrm{NA})$ were labeled as PD-only losses and the NA associated with the TE implies "not applicable." The total for these PD-only losses was 1.774 billion dollars.

The cell $S(\mathrm{NA},+)$ is analogous to cell $S(+,\mathrm{NA})$ with the role of PD and TE switched. A TE loss contributing to this cell was **strictly** positive with the accompanying PD field having either a blank or a zero value. The total value of these TE-only losses was 0.065 billion dollars.

The cell $S(\mathrm{NA},\mathrm{NA})$ is a void cell, indicating neither a PD nor a TE loss situation.

Table 1 is a helpful overview of the severity (ALC). One can estimate ALC from the sample data as the ratio of total losses divided by number of losses. In this case, Table 1 provides information with regard to the numerator of this ratio. It is understood that zero losses are excluded from both the numerator and the denominator of the ratio used to estimate the ALC. In this paper, ALC is computed based on the knowledge of a fitted probability distribution, a SOLD, derived from individual trended incurred loss amounts as explained below. The figures in Table 1 are meaningful only for comparing different loss categories and should not be viewed in any absolute sense.

For policies providing PD-only cover, the severity curve needed should be based on the individual PD losses appearing in the cells $S(+,+)$ and $S(+,\mathrm{NA})$. The PD losses contributing to these two cells were not necessarily identically distributed. A PD loss in the cell $S(+,+)$ tended, on average, to be larger than a PD loss from cell $S(+,\mathrm{NA})$. In order to price a PD-only cover, one should use **all** the available PD losses. In this case, a univariate probability distribution is needed. Similarly, for policies providing TE-only cover, one should use the individual TE losses contributing to the cells $S(+,+)$ and $S(\mathrm{NA},+)$. Once again, the curve fitting is done in a univariate setting.

To price a multiple cover policy—PD and TE—one needs to make use of all the loss data, i.e., the individual losses contributing to cells $S(+,+)$, $S(+,\mathrm{NA})$, and $S(\mathrm{NA},+)$. In this paper, the focus is on the application of the multivariate techniques to estimate a suitable bivariate dis-

Table 2. Summary of total unadjusted claim count. Line of business: Property, non-CAT, non-BM losses

PD\TE	+	NA
+	$F(+,+)$ 1,672	$F(+,\mathrm{NA})$ 12,257
NA	$F(\mathrm{NA},+)$ 165	$F(\mathrm{NA},\mathrm{NA})$

PD: Property Damage
TE: Time Element (Business Interruption)

tribution based on losses arising from the cell $S(+,+)$ **only**.

The claim count data were summarized by the Table 2. The four cells $F(+,+)$, $F(+,\mathrm{NA})$, $F(\mathrm{NA},+)$, and $F(\mathrm{NA},\mathrm{NA})$ of Table 2 are defined in a similar fashion as the four cells in Table 1. The cell $F(+,+)$ shows the number of claims for which both the PD loss and the accompanying TE loss were strictly positive. The total number of claims was 1,672 over the period under study. $F(+,\mathrm{NA})$ presents claim counts related to the PD-only losses, which were 12,257. The $F(\mathrm{NA},+)$, with a figure of 165, corresponds to the TE-only claim counts. Once again, the cell $F(\mathrm{NA},\mathrm{NA})$ represents a "not applicable" or a void cell.

Table 2 may be helpful to give an overview of the frequency. One can estimate the frequency from sample data as the ratio of Number of Claims divided by an appropriate exposure amount. In this instance, Table 2 provides information with regard to the numerator of this ratio. Alternatively, the frequency can be modeled based on the mean of a Poisson or negative binomial random variable divided by a suitable exposure amount. The estimation of the frequency component of the pure premium is not within the scope of this paper.

4. Multivariate methodology: selection, estimation, and fit

In this section, issues related to selecting, estimating the parameters, and fitting a bivariate probability distribution to the data are discussed.

An important point worth emphasizing at the outset is with regard to the perspective on statistical inference. In this paper the approach to inference is **exploratory** as opposed to being confirmatory. Hence, there is more emphasis on the use of graphical tools to provide informal support for the reasonableness of the stated position, rather than to provide a proof for the stated position. Other common uses of exploratory tools are to highlight anomalies or outliers in the data. The exploratory tools presented here should be of interest to practicing actuaries. For the sake of completeness, the more technical materials are presented in the appendices.

With regard to curve fitting, the distribution selected to model the losses is a choice that cannot be proven correct. The underlying distribution that produces losses cannot be known without doubt, and the best that can be done is to show that the selected distribution is reasonable.

Let

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix},$$

a 2×1 vector, denote a random vector that presents the loss data, where Y_1 denotes a PD loss, and Y_2 denotes a TE loss. The loss vector Y may also be written as a row vector by writing $Y' = (Y_1, Y_2)$ where the prime (shown as a superscript) denotes the operation of transposing a column to a row.

The Y realizations considered here were pairs of strictly positive losses related to the cell $S(+,+)$ in Table 1. The losses were adjusted for trend for the purpose of curve fitting. The focus of this paper is solely with regard to fitting a bivariate distribution to losses arising from cell $S(+,+)$. For cells $S(+,\mathrm{NA})$ and $S(\mathrm{NA},+)$, univariate probability distributions can be fitted to those losses. But the subject is not discussed further in this paper.

If no prior knowledge is available with regard to suitable distributions for the components of Y, i.e., Y_1 and Y_2, then it may be a good idea to start by trying to fit a few standard distributions to these components prior to selecting a distribu-

Figure 1. Plots to support the normal assumption for X_1

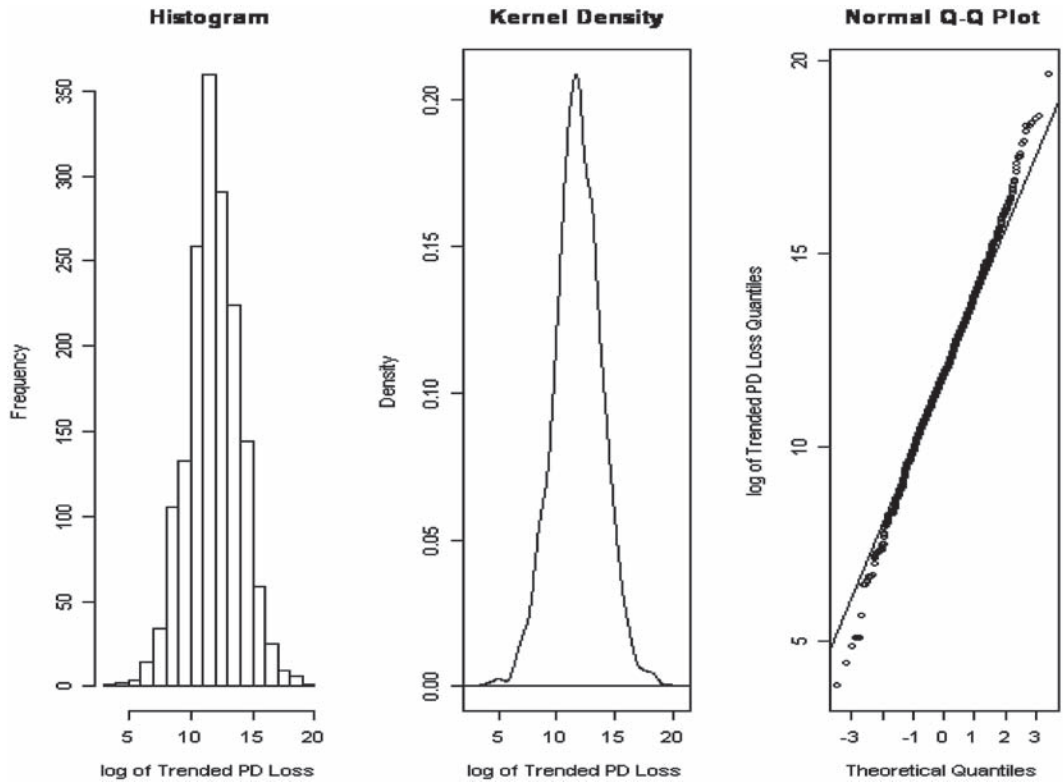


Table 3. A comparison of univariate fit statistics

Trended PD Loss			Trended TE Loss		
Distribution	KS	AD	Distribution	KS	AD
Lognormal	0.0242	1.25	Lognormal	0.0222	1.47
Gamma	0.2560	229.87	Gamma	0.2560	243.06
Extreme Value	0.3355	355.57	Weibull	0.3375	221.90
Pareto	0.3913	424.87	Extreme Value	0.3551	372.24

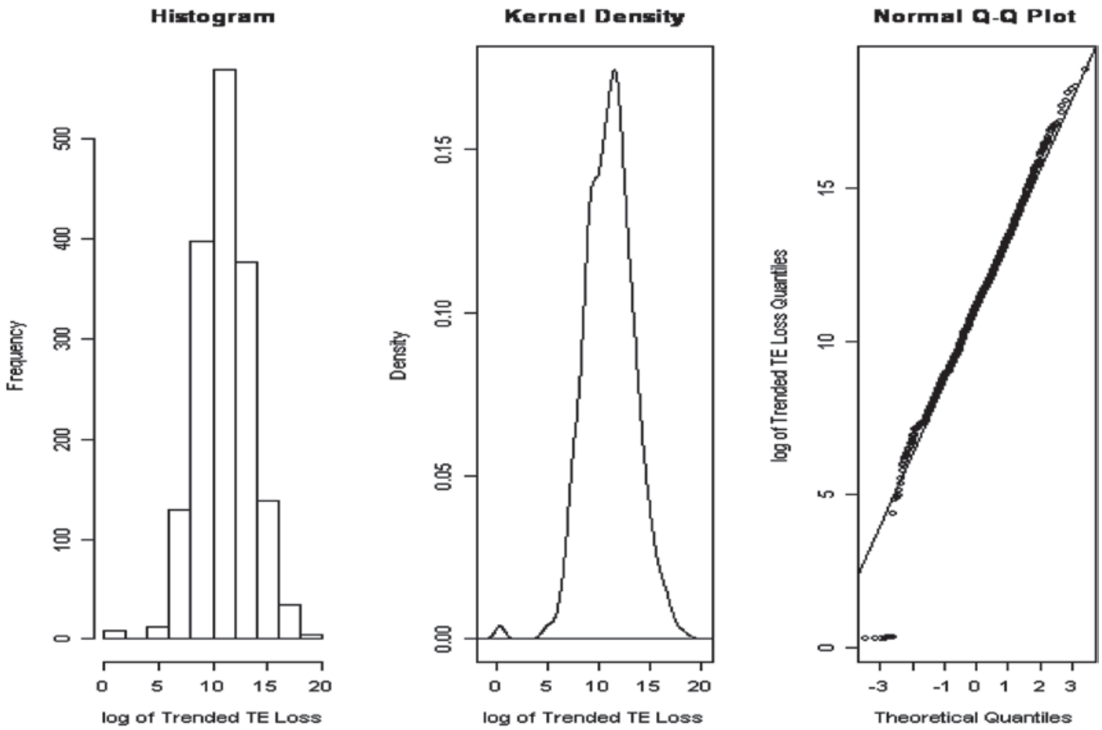
tion for Y. The software Crystal Ball, an add-on of Microsoft Excel, can be used to fit a number of standard univariate distributions to the data. Table 3 provides a summary of the goodness-of-fit statistics for the trended PD losses and the trended TE losses arising from claims contributing to cell $S(+,+)$. Among the four fitted distributions, the lognormal provided the best fit based on the Kolmogorov-Smirnov (KS) criteria. The Anderson-Darling (AD) statistics have also been provided as an alternative to KS criteria for assessing the goodness of fit. These statistics are further discussed in Appendix 1.

Table 3 suggests considering the lognormal distributions for Y_1 and Y_2. In order to check the validity of the lognormal distributions, it is more convenient to work with the transformed version of these variates. Transform Y_1 and Y_2 according to $X_1 = \log(Y_1)$ and $X_2 = \log(Y_2)$. If the data suggests that it is reasonable to assume that X_1 and X_2 are normally distributed, then, based on the above transformation, it is reasonable to state that Y_1 and Y_2 are lognormally distributed.

Clues to the shape of the underlying distribution for X_1 may be obtained by plotting a histogram and a kernel density of X_1. In addition, a QQ-Plot can be utilized to examine if the data supports informally whether X_1 is distributed according to a hypothesized probability distribution. Appendix 2 provides more information about these tools.

Figure 1 displays the histogram, kernel density, and QQ-Plot for X_1, based on the log of

Figure 2. Plots to support the normal assumption for X_2



the trended PD losses contributing to the cell $S(+,+)$. The left panel and the middle panel exhibit an approximately bell-shaped curve for the distribution of X_1. The kernel density provides a smoothed version of the histogram, as the shape of histograms tend to be affected by the specification of the number of bins and their spacing on the horizontal axis. The right panel of Figure 1, the QQ-Plot, compares the expected (according to normal) and actual quantiles. The majority of the points lie about a straight line. Hence, the plots in Figure 1 suggest that it is reasonable to assume that X_1 may be represented as normal.

Figure 2 is similar to Figure 1, but applies to X_2. It is based on the log of the trended TE losses contributing to the cell $S(+,+)$. Once again the plots in Figure 2 suggest that it is reasonable to assume that X_2 may be normally distributed.

Figures 1 and 2 support informally the reasonableness of the assumption that X_1 and X_2 are normally distributed; hence it is natural to consider a bivariate normal, X, to represent (X_1, X_2).

The bivariate probability distributions considered in this paper are:

1) bivariate lognormal distribution Y,
2) bivariate normal distribution X, and
3) standardized bivariate normal Z.

These three distributions are related to each other according to

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} \exp(X_1) \\ \exp(X_2) \end{pmatrix} = \begin{pmatrix} \exp(\mu_1 + \sigma_1 Z_1) \\ \exp(\mu_2 + \sigma_2 Z_2) \end{pmatrix}$$

where μ_is and σ_is are parameters of a bivariate normal X as defined below.

The distributions of Y and Z may be derived from X. Hence, the focus will be on the bivariate normal distribution X.

The bivariate normal distribution X has parameters (μ, Σ) where

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$$

is the mean vector, and

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

is the variance-covariance matrix. The interpretation of the five parameters of X—(μ_1, μ_2, σ_1, σ_2, and ρ)—in terms of the moments of the components of X are as follows:

$$\mu_1 = E(X_1), \qquad \mu_2 = E(X_2),$$

$$\sigma_1^2 = \mathrm{Var}(X_1), \qquad \sigma_2^2 = \mathrm{Var}(X_2), \quad \text{and}$$

$$\rho = \frac{\mathrm{Cov}(X_1,X_2)}{\sigma_1\sigma_2}$$

where ρ denotes the correlation between X_1 and X_2.

The bivariate lognormal Y is derived from X by exponentiation of the X_1 and X_2 components of X. It has the same parameters (μ,Σ) as X.

The standardized bivariate normal Z is derived from X by standardizing the components of X, i.e., by replacing X_is by $Z_i = (X_i - \mu_i)/\sigma_i$, $i = 1,2$. It has a single parameter ρ.

If the data supports the notion that it is reasonable to assume a bivariate normal distribution for X, then it follows that it is also reasonable to assume that Y has a bivariate lognormal distribution.

In order to examine if the data supports the assumption of the bivariate normality for X, it is necessary to compute certain statistics of interest from the data as described below. These statistics depend upon the estimate of the parameters (μ,Σ), i.e., upon the maximum likelihood estimates $(\hat{\mu},\hat{\Sigma})$. The maximum likelihood estimates of the five parameters $(\hat{\mu}_1,\hat{\mu}_2,\hat{\sigma}_1,\hat{\sigma}_2,\hat{\rho})$ of a bivariate normal X are

$$\hat{\mu} = \begin{pmatrix}\hat{\mu}_1\\ \hat{\mu}_2\end{pmatrix}, \qquad \hat{\mu}_1 = \bar{x}_1 = \frac{1}{n}\sum_{i=1}^{n} x_{1,i},$$

$$\hat{\mu}_2 = \bar{x}_2 = \frac{1}{n}\sum_{i=1}^{n} x_{2,i},$$

$$\hat{\Sigma} = \frac{(n-1)}{n}S, \qquad S = \begin{pmatrix} s_{11} & s_{12}\\ s_{21} & s_{22}\end{pmatrix}$$

$$s_{11} = \frac{1}{n-1}\sum_{i=1}^{n}(x_{1,i}-\bar{x}_1)^2,$$

$$s_{22} = \frac{1}{n-1}\sum_{i=1}^{n}(x_{2,i}-\bar{x}_2)^2,$$

$$s_{12} = s_{21} = \frac{1}{n-1}\sum_{i=1}^{n}(x_{1,i}-\bar{x}_1)(x_{2,i}-\bar{x}_2),$$

$$\hat{\sigma}_1 = \sqrt{\frac{(n-1)}{n}s_{11}}, \qquad \hat{\sigma}_2 = \sqrt{\frac{(n-1)}{n}s_{22}},$$

$$\hat{\rho} = \frac{s_{12}}{\sqrt{s_{11}s_{22}}},$$

where

$$x_{1,i} = \log(i\text{th trended PD loss}), \qquad 1 \le i \le n,$$

and

$$x_{2,i} = \log(i\text{th trended TE loss}), \qquad 1 \le i \le n.$$

The matrix S above is the unbiased estimator for the Σ, while $\hat{\Sigma}$ is the maximum likelihood estimator for Σ, which is a biased estimator.

Based on the trended losses contributing to the cell $S(+,+)$ only, the maximum likelihood estimate of the five parameters were

$$\hat{\mu}_1 = 11.830, \qquad \hat{\mu}_2 = 11.057,$$

$$\hat{\sigma}_1 = 2.086, \qquad \hat{\sigma}_2 = 2.399, \quad \text{and} \quad \hat{\rho} = 0.646.$$

The informal support of the bivariate normality of X lies in the use of two graphs: Figure 3 (Ellipse) and Figure 4 (QQ-Plot), see below. The theory underlying the constructions of these two figures rests upon a result, referred here to as Theorem 1 (see below), from the subject of multivariate statistical analysis (see Johnson and Wichern [4]).

THEOREM 1 *If the random vector X has a bivariate normal distribution with parameters (μ,Σ), then the random variable Q, a quadratic form,*

$$Q = (X_1-\mu_1, X_2-\mu_2)'\Sigma^{-1}\begin{pmatrix}X_1-\mu_1\\ X_2-\mu_2\end{pmatrix} \tag{1}$$

is distributed as a chi-square distribution with 2 degrees of freedom.

Based on Theorem 1, provided that X is assumed to be distributed as a bivariate normal, then the following probability statement can be made: The chance that

$$Q \le \chi_2^2(0.95) \tag{2}$$

Table 4. Ellipse coverage probabilities: probability of coverage

Expected	Actual
95%	94.1% (1,574 out of 1,672 pairs)
90%	89.7%
50%	56.2%

is 95%, where $\chi^2_2(0.95)$ denotes the 95th quantile of a chi-square distribution with two degrees of freedom.

In order to create Figure 3 based on inequality (2), two changes must be made. First, the X_1 and X_2 random variables are replaced by the Z_1 and Z_2 variables, the standardized version of the X_1 and X_2. Second, the parameters (μ,Σ) are replaced by their sample estimates $(\hat{\mu},\hat{\Sigma})$ in order to compute the values of Z_1 and Z_2 from the data. After making these two changes, then the inequality (2) becomes

$$\hat{Q} = \frac{1}{1-\hat{\rho}^2}[Z_1^2 - 2\hat{\rho}Z_1Z_2 + Z_2^2] \leq \chi^2_2(0.95). \tag{3}$$

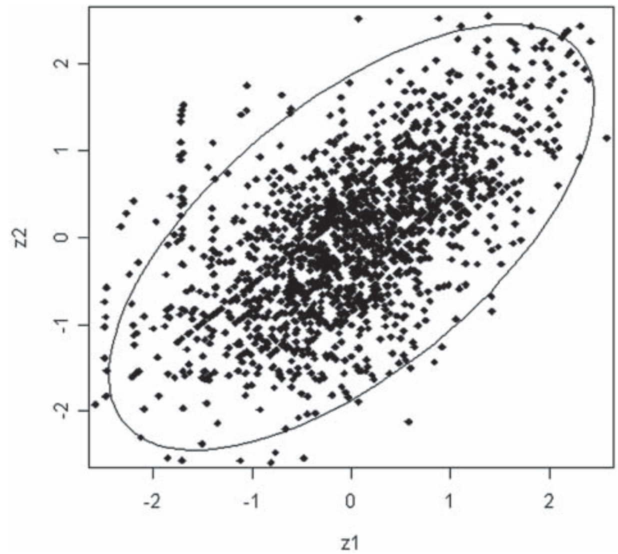
The ellipse in Figure 3 is based on plotting the observed values of (z_1, z_2) subject to the inequality (3).

Table 4 provides additional figures for the 50% and 90% expected and actual coverage probabilities based on the assumption of the bivariate normality of X. There is a close agreement between expected and actual coverage probabilities in Table 4, especially for the higher probability levels. Thus, Figure 3 and Table 4 support informally that it is reasonable to state that X has a bivariate normal distribution.

If the premise of bivariate normality is not tenable, then a search should be made for another bivariate distribution, or alternatively, use copulas in conjunction with marginal distributions of Y_1 and Y_2 as deemed appropriate; see Nelson [7] regarding copulas.

In this paper, a **single** bivariate lognormal distribution has been considered to represent the loss distribution. No distinction has been made

Figure 3. Ellipse with 95% probability coverage



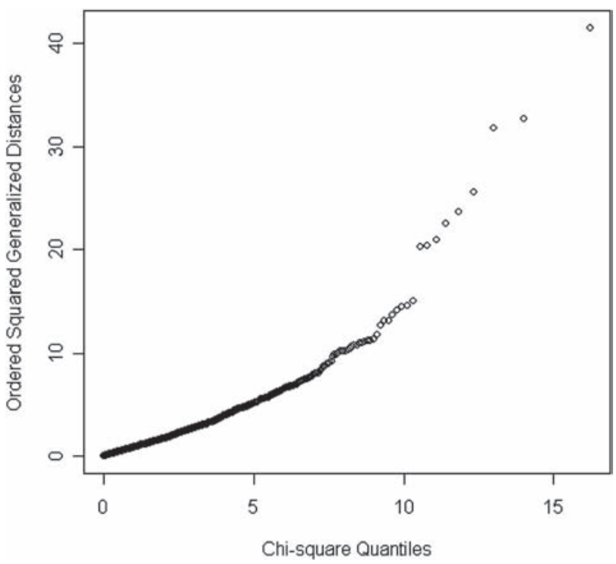
with respect to the risk attributes. Thus, the same distribution is implied for risks with differing construction types, protection, or size. The size may be represented by the location value and/or the business interruption limit. If it is desirable to have a family of bivariate lognormal distributions, differing by risk characteristics, then one way to accomplish this goal may be by incorporating these risk attributes into the parameters (μ,Σ). For example, to account for the effect of the size of the risk, one can replace the parameters (μ,Σ) by parameters (μ^*,Σ) according to

$$\mu^* = \begin{pmatrix}\mu_1^*\\ \mu_2^*\end{pmatrix} = \begin{pmatrix}\beta_{0,1} + \beta_{1,1}\log(\text{PD Value})\\ \beta_{0,2} + \beta_{1,2}\log(\text{TE Limit})\end{pmatrix}.$$

Here, the regression-like parameters $\beta_{0,1}$, $\beta_{1,1}$, $\beta_{0,2}$, and $\beta_{1,2}$ need to be estimated from the data. A nonlinear estimation procedure, similar to the GLM approach, may be used to estimate all the model parameters. However, this topic is not pursued further in this paper.

Another graphical procedure for illustrating informally the reasonableness of the assumption of the bivariate normality is the use of the QQ-Plot—Figure 4 below. In order to create this figure, it is necessary first to introduce the notion of

Figure 4. QQ plot for bivariate normality



"Squared Generalized Distance" statistics, d_j^2s, (see Johnson and Wichern [4]):

$$d_j^2 = (X_{1,j} - \hat{\mu}_1, X_{2,j} - \hat{\mu}_2)' S^{-1} \begin{pmatrix} X_{1,j} - \hat{\mu}_1 \\ X_{2,j} - \hat{\mu}_2 \end{pmatrix}, \quad j = 1,2,\ldots,n. \tag{4}$$

These statistics like the random variable Q, in Equation (1), are quadratic forms which can be computed from the data since the parameters (μ,Σ) of the bivariate normal have been replaced by their sample estimates $(\hat{\mu},S)$, see above. Assuming that X has a bivariate normal distribution, Theorem 1 implies that the d_j^2 statistics constitute a sample of size n that is approximately distributed as a chi-square with two degrees of freedom. The approximate nature of this result is due to the fact that the parameters (μ,Σ) have been replaced by their estimates. The QQ-Plot, Figure 4, was obtained by plotting the expected quantiles from a chi-square distribution with two degrees of freedom on the horizontal axis against the actual observed quantiles of d_j^2s as plotted on the vertical axis. (Refer to Appendix 3 for more details.)

Figure 4 shows that the majority of points lie close to a straight line, suggesting the reasonableness of the assumption of the bivariate normality.

5. Specification of claim payment function

The potential insured loss amounts Y_1 (PD loss) and Y_2 (TE loss) are affected by loss-sensitive provisions of a policy. In this paper, the loss-sensitive provisions are comprised of a PD deductible D_1, a TE deductible D_2, an attachment point A, and a combined limit of L. The interaction of losses, $Y' = (Y_1, Y_2)$, with loss-sensitive provisions (D_1, D_2, A, L) is defined through the CPF, denoted by $g(Y_1, Y_2)$, see below.

Before discussing the specification of a CPF in a multivariate setting, it may be instructive to review the form of the CPF in a univariate situation. For a single cover policy with a deductible D, and limit L, let the random variable X denote an insured loss. Then the CPF, designated by $h(X)$ is

$$\begin{aligned} h(X) &= \min\{\max(X - D, 0), L\} \\ &= \min(X, D+L) - \min(X, D). \end{aligned} \tag{5}$$

The expected value of the $h(X)$, $E[h(X)]$, denotes the severity component of pure premium—the ALC, and can be stated as

$$E[h(X)] = \mathrm{LEV}(D+L) - \mathrm{LEV}(D) \tag{6}$$

where $\mathrm{LEV}(c)$ is the limited expected value of the loss amount X subject to a cap c. It is computed as

$$\mathrm{LEV}(c) = \int_0^\infty \min(x,c)\, dF(x)$$

where F denotes the cdf of the insured loss X.

In the case of X being distributed as a lognormal with parameters (μ, σ^2), then the $\mathrm{LEV}(c)$ function is

$$\begin{aligned} \mathrm{LEV}(c) &= e^{\mu + 1/2\sigma^2}\Phi\left(\frac{\log(c) - \mu - \sigma^2}{\sigma}\right) \\ &\quad + c\left[1 - \Phi\left(\frac{\log(c) - \mu}{\sigma}\right)\right] \end{aligned} \tag{7}$$

where Φ is the cdf of a standard normal.

Thus, in the lognormal case, it is easy to compute the expected value of the CPF using (6) and (7). Simply use the exponential function and the cdf of the standard normal, Φ. However, the computation of the ALC for multiple cover policies is more complicated, as explained below.

For a multiple cover policy with a potential PD loss amount y_1 and a TE loss amount y_2, subject to loss-sensitive features, (D_1, D_2, A, L), the CPF as denoted by $g(y_1, y_2)$ cannot be expressed by a single formula and has to be defined piecemeal. What is needed is a breakup of the positive quadrant, $R_+^2 = [0,\infty) \times [0,\infty)$, into ten regions, as defined below. Then, the value of $g(y_1, y_2)$ can be determined depending upon the region where the point (y_1, y_2) will reside.

The ten required regions are defined as

$$R_1 = \{(y_1,y_2) : y_1 \le D_1,\ y_2 \le D_2\}$$

$$R_2 = \{(y_1,y_2) : y_1 \le D_1,\ D_2 < y_2 \le D_2 + A\}$$

$$R_3 = \{(y_1,y_2) : y_1 \le D_1,\ D_2 + A < y_2 \le D_2 + A + L\}$$

$$R_4 = \{(y_1,y_2) : y_1 \le D_1,\ y_2 > D_2 + A + L\}$$

$$R_5 = \{(y_1,y_2) : D_1 < y_1 \le D_1 + A,\ y_2 \le D_2\}$$

$$R_6 = \{(y_1,y_2) : D_1 + A < y_1 \le D_1 + A + L,\ y_2 \le D_2\}$$

$$R_7 = \{(y_1,y_2) : y_1 > D_1 + A + L,\ y_2 \le D_2\}$$

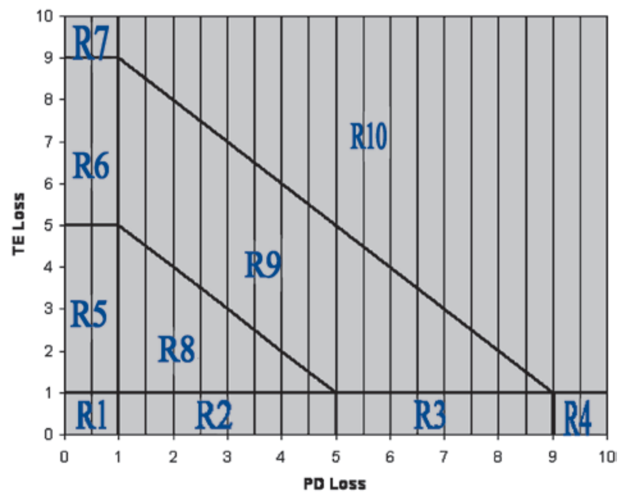
$$R_8 = \{(y_1,y_2) : y_1 > D_1,\ y_2 > D_2,\ y_1 + y_2 \le D_1 + D_2 + A\}$$

$$R_9 = \{(y_1,y_2) : y_1 > D_1,\ y_2 > D_2,\ D_1 + D_2 + A < y_1 + y_2 \le D_1 + D_2 + A + L\}$$

$$R_{10} = \{(y_1,y_2) : y_1 > D_1,\ y_2 > D_2,\ y_1 + y_2 > D_1 + D_2 + A + L\}.$$

The diagram (Figure 5) depicts these regions. These regions are used to integrate the function $g(y_1, y_2)$ over the positive quadrant, as explained in Appendix 4. For illustrative purposes, the values chosen for these loss-sensitive features are $D_1 = 1$, $D_2 = 1$, $A = 4$, and $L = 4$.

Figure 5. Regions of integration



The following will illustrate how $g(y_1, y_2)$ is computed according to the region. For region 1, $R_1 = \{(y_1,y_2) : y_1 \le D_1,\ y_2 \le D_2\}$, the value of $g(y_1, y_2)$ is zero, since both losses are below their respective deductibles. For region 2, $R_2 = \{(y_1,y_2) : y_1 \le D_1,\ D_2 < y_2 \le D_2 + A\}$, the value of $g(y_1, y_2)$ is also zero. In this case y_1, the PD loss, is below its deductible D_1. The TE loss, y_2, exceeds its deductible D_2, but due to the imposition of the attachment A, the constraint: $y_2 - D_2 \le A$, there is no loss payment to be made. For region 3, $R_3 = \{(y_1,y_2) : y_1 \le D_1,\ D_2 + A < y_2 \le D_2 + A + L\}$, the CPF is given by $g(y_1, y_2) = \min\{\max(y_2 - (D_2 + A), 0), L\}$. Similar expressions for $g(y_1, y_2)$ may be written in terms of y_1, y_2, D_1, D_2, A and L depending on the region applied. In a computing environment, expressions for the $g(y_1, y_2)$ may be written in terms of if-then-else based on the region.

6. Computation of average loss cost based on bivariate lognormal

In this section, an algorithm is outlined for computing the ALC, the severity component of the pure premium. The following formula, (8), may be used as the basis of calculating the pure premium. This formula takes into consider-

Table 5. Loss-sensitive provisions and related average loss cost

D_1	D_2	A	L	ALC
100,000	300,000	1,000,000	5,000,000	558,095
300,000	100,000	1,000,000	5,000,000	553,353
100,000	300,000	0	5,000,000	821,033
300,000	100,000	0	5,000,000	799,109
100,000	300,000	1,000,000	6,000,000	617,968
100,000	300,000	1,000,000	10,000,000	794,020
100,000	100,000	2,000,000	5,000,000	452,736
50,000	50,000	1,000,000	5,000,000	599,483
0	0	0	5,000,000	985,479
50,000	50,000	100,000	5,000,000	860,974
100,000	100,000	500,000	5,000,000	688,031
100,000	100,000	1,000,000	5,000,000	583,402
250,000	250,000	500,000	5,000,000	630,820
100,000	100,000	2,000,000	5,000,000	452,736
100,000	100,000	1,000,000	1,000,000	191,943

ation the way the data was organized in Section 3 by Table 1 and Table 2.

$$\text{Pure Premium} = \text{Freq}_{\text{PD}}\text{ALC}_{\text{PD}} + \text{Freq}_{\text{TE}}\text{ALC}_{\text{TE}} + \text{Freq}_{\text{PD\&TE}}\text{ALC}_{\text{PD\&TE}} \quad (8)$$

where Freq_{PD}, Freq_{TE}, and $\text{Freq}_{\text{PD\&TE}}$ are frequency values, and ALC_{PD}, ALC_{TE}, and $\text{ALC}_{\text{PD\&TE}}$ are ALC values.

The ALC_{PD} and ALC_{TE} can be computed using appropriate univariate SOLD based on trended loss data fitted to cells $S(+,\text{NA})$ and $S(\text{NA},+)$, respectively. In this paper, the interest lies only in explaining how to estimate $\text{ALC}_{\text{PD\&TE}}$ based on a bivariate lognormal distribution.

In Section 4, the data was fitted to losses arising from cell $S(+,+)$. A bivariate lognormal distribution was considered to represent the SOLD, and its five parameters were estimated from the data. In Section 5, the form of the CPF, the function $g(y_1,y_2)$, was described in terms of the interaction of loss data (y_1,y_2) with loss-sensitive features, (D_1,D_2,A,L). Finally, an expression for ALC can be provided by defining the ALC as a double integral

$$\text{ALC}_{\text{PD\&TE}}(D_1,D_2,A,L;\mu,\Sigma) = \iint g(y_1,y_2) f_{Y_1,Y_2}(y_1,y_2)\,dy_1\,dy_2 \quad (9)$$

where (D_1,D_2,A,L) is loss-sensitive provisions of the multiple cover policy,

(μ,Σ) are the parameters of the bivariate lognormal distribution **replaced** by their sample estimates,

$g(y_1,y_2)$ is the CPF as described in Section 5, and

$f_{Y_1,Y_2}(y_1,y_2)$ is the density function of a bivariate lognormal (see Section 4 and Appendix 3).

What is required is an estimate for the above double integral. Since the computation of (9) is rather technical, it has been supplied in Appendices 4 and 5.

Table 5 provides estimates of $\text{ALC}_{\text{PD\&TE}}(D_1,D_2,A,L;\mu,\Sigma)$ by considering differing loss-sensitive provisions. Such a table is useful in assessing the impact of (D_1,D_2,A,L) upon $\text{ALC}_{\text{PD\&TE}}(D_1,D_2,A,L;\mu,\Sigma)$.

It should be noted that the figures in Table 5 are based on the value of the estimated parameters, as well as selection of the value of N (N = 100). The variable N implicitly controls the error incurred in estimating ALC (see Appendix 4 for the definition and use of the N term). The computation time to estimate ALC in Table 4 was a matter of a few seconds for the case of $N = 100$. For the case of $N = 1000$ it was less than two minutes.

The algorithm outlined above estimates ALC, a double integral, using a numerical procedure that approximates the double integral (9) by a suitable sum as explained in Appendix 4. An alternative numerical procedure to estimate ALC can be based on simulation. The simulation approach requires generating random pairs (y_1, y_2) according to a bivariate lognormal whose parameters (μ, Σ) have been estimated from the data (see Section 4). Then, for each simulated pair (y_1, y_2), the value of CPF, $g(y_1, y_2)$, is calculated based upon the ten regions in which (y_1, y_2) resides (see Section 5). This procedure is repeated a number of times and the corresponding generated $g(y_1, y_2)$ values are averaged to provide an alternative estimate for the ALC. An important step in the simulation procedure is generating samples from a bivariate lognormal distribution. Appendix 6 provides an algorithm for doing this and discusses briefly some advantages as well as drawbacks of the simulation approach.

7. Conclusions

This paper discussed pricing a property policy with multiple cover for PD and TE subject to an attachment point, separate deductibles, and a combined limit relying on a bivariate lognormal distribution. Comparisons were made between univariate and multivariate curve fitting. A methodology needed to fit a bivariate lognormal to the data was developed. An algorithm was given for estimating the ALC, a double integral. Hopefully, this paper will encourage other actuaries to contribute further to the methodology needed for pricing multiple cover policies and the applications of multivariate statistical techniques to the actuarial field.

Acknowledgement

The author has benefited from discussions of this subject with Shane Vadbunker, and in particular would like to thank him for writing an Excel VBA program to estimate average loss cost based on the bivariate lognormal distribution.

References

[1] Bickel, P. J., and K. A. Doksum, *Mathematical Statistics*, Vol. I (2nd edition), Upper Saddle River, NJ: Prentice Hall, 2001.

[2] Drezner, Z., and G. O. Wesolowsky, "On the Computation of the Bivariate Normal Integral," *Journal of Statistical Computer Simulation* 35, 1990, pp. 101–107.

[3] Jarrow, R., and S. Turnbull, *Derivative Securities* (2nd edition), Cincinnati, OH: South-Western College Publishing, 2000.

[4] Johnson, R. A., and D. W. Wichern, *Applied Multivariate Statistical Analysis* (5th edition), Upper Saddle River, NJ: Prentice Hall, 2002.

[5] Larose, T. D., *Data Mining Methods and Models*, Hoboken, NJ: Wiley-Interscience, 2006.

[6] Ludwig, S. J., "An Exposure Rating Approach to Pricing Property Excess-of-Loss Reinsurance," *Proceedings of the Casualty Actuarial Society* 78, 1991, pp. 110–145.

[7] Nelson, R. B., *An Introduction to Copulas* (2nd edition), New York: Springer, 2006.

[8] Salzmann, R. E., "Rating by Layer of Insurance," *Proceedings of the Casualty Actuarial Society* 50, 1963, pp. 15–26.

[9] SAS Institute, Inc., SAS OnlineDoc 9.1.3 documentation, 2004.

[10] Silverman, B. W., *Density Estimation for Statistics and Data Analysis*, New York: Chapman and Hall, 1986.

[11] Venables, W. N. and B. D. Ripley, *Modern Applied Statistics with S.* (4th edition), New York: Springer, 2002.

Appendix 1. Kolmogorov-Smirnov and Anderson-Darling goodness-of-fit statistics

Kolmogorov-Smirnov (KS) or Anderson-Darling (AD) statistics are used to test if a given sample of observations conforms to a hypothesized univariate distribution of interest. In Section 4, these statistics were calculated for four standard distributions available in Crystal Ball in order to **rank** informally the fitted distributions. The smaller the value of the KS (or AD) statistic the better the fit is deemed to be. A description of how these statistics are calculated from the data is given below; see the references for further information.

Kolmogorov-Smirnov (KS): Let $x_1, x_2, \ldots, x_n$ denote a random sample from a population with cumulative distribution function F_0, the "hypothesized" cdf.

Let $x_{n1}, x_{n2}, \ldots, x_{nn}$ denote the corresponding order statistics, and let $\hat{F}_n$ denote the empirical distribution function corresponding to the above sample. In particular, $\hat{F}_n$ is defined by

$$\hat{F}_n(x) = \frac{1}{n}\sum_{1}^{n} I(x_i \leq x)$$

where

$$I(x_i \leq x) = \begin{pmatrix} 1, & \text{if} \quad x_i \leq x \\ 0, & \text{if} \quad x_i > x \end{pmatrix}.$$

The KS statistic D_n is defined as

$$\begin{aligned} D_n &= \sup |\hat{F}_n(x) - F_0(x)| \\ &= \max_{1 \leq i \leq n} \max \left\{ \frac{i}{n} - F_0(x_{ni}), F_0(x_{ni}) - \frac{(i-1)}{n} \right\}. \end{aligned}$$

Refer to Bickel and Doksum [1] for more details.

Anderson-Darling (AD): AD provides an alternative goodness-of-fit statistic to KS.

Let $u_{ni} = F_0(x_{ni})$, $1 \leq i \leq n$, then AD is defined as

$$\begin{aligned} \text{AD} &= \int_{-\infty}^{\infty} \frac{(\hat{F}_n(x) - F_0(x))^2}{F_0(x)(1 - F_0(x))} dF_0(x) \\ &= -n - \frac{1}{n}\sum_{1}^{n} [(2i-1)\log(u_{ni}) \\ &\qquad + (2n+1-2i)\log(1-u_{ni})]. \end{aligned}$$

For more details, refer to the SAS Institute [9] documentation on PROC UNIVARIATE.

Appendix 2. Histogram, Kernel density, and QQ-plot

Figures 1 and 2 display three graphs: histogram, kernel density, and QQ-Plot. These graphs are useful for evaluating informally the assumption of normality of the transformed losses.

Histogram and kernel density provide insight with regard to the shape of the underlying density. The QQ-Plot is used to check informally the validity of a specified assumed distribution.

The software R was used to plot these graphs. R is available free through the Internet under the General Public License. R a useful tool for doing statistical analysis with many nice graphic capabilities. It may be downloaded from the site www.r-project.org. The R functions used to plot these graphs were: hist(), density(), and qqnorm() respectively.

Histogram: Histograms provide a rough glimpse of the shape of a density function. The following description of the histogram is from Silverman [10].

Given a sample of observations $x_1, x_2, \ldots, x_n$, select an origin point x_0 and a bin width h. Define the bins to be the intervals $[x_0 + mh, x_0 + (m+1)h)$, for positive and negative integers m.

The histogram is then defined by

$$\hat{f}(x) = \frac{1}{nh}(\text{Number of } x_i \text{ in same bin as } x).$$

Kernel Density: Kernel density estimators, as compared to histograms, provide smoother estimates of the density function.

A kernel, $K(x)$, is any nonnegative function with the following property

$$\int_{-\infty}^{\infty} K(x)dx = 1.$$

Given a kernel $K(x)$ and a positive number h called a smoothing parameter, then the kernel estimator is defined as

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right).$$

See Silverman [10] for more details.

QQ-Plot: Normal Case: Let $x_1, x_2, \ldots, x_n$ be a sample of observations with corresponding order statistics denoted by $x_{n1}, x_{n2}, \ldots, x_{nn}$. To check informally the assumption of the normality of the data, a QQ-Plot is constructed. Plot the points

$$\left(\Phi^{-1}\left(\frac{i}{n+1}\right), x_{ni}\right), \qquad 1 \leq i \leq n,$$

where Φ^{-1} denotes the inverse of the cdf of a standard normal. The Excel function NORMINV may be used to compute the values of Φ^{-1}.

If the points in the QQ-Plot lie about a straight line, then there is informal support for the assumption of normality.

For Figure 1, the data used were $x_i = \log(i\text{th Trended PD loss}),\quad 1 \le i \le n,$ and for Figure 2, the data applied were $x_i = \log(i\text{th Trended TE loss}),\ 1 \le i \le n$. In both cases the transformed trended losses arose from those losses contributing to the cell $S(+,+)$ only.

Appendix 3. Bivariate distributions and related results

A good reference book on the subject of multivariate statistical analysis is the text by Johnson and Wichern [4]. The definitions and results stated here are based mainly on that book.

The bivariate lognormal distribution Y is derived from bivariate normal distribution X by exponentiation of the X_1 and X_2 components of X. That is

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} \exp(X_1) \\ \exp(X_2) \end{pmatrix}.$$

It has the same parameters (μ,Σ) as X.

The standardized bivariate normal distribution Z is related to the bivariate normal distribution X according to

$$Z_1 = \frac{(X_1-\mu_1)}{\sigma_1}, \qquad Z_2 = \frac{(X_2-\mu_2)}{\sigma_2}$$

where μ_1, μ_2, σ_1, and σ_2 are parameters of the bivariate normal X as defined in Section 4.

The density of a bivariate normal distribution X is

$$f_{X_1,X_2}(x_1,x_2) = \frac{1}{2\pi}\frac{1}{|\Sigma|^{1/2}}\exp\left(-\frac{1}{2}Q\right)$$

where $|\Sigma| = \sigma_1^2\sigma_2^2(1-\rho^2)$ is the determinant of the variance-covariance matrix Σ, and $Q = (x-\mu)'\Sigma^{-1}(x-\mu)$ is a quadratic form in x_1 and x_2.

The standardized bivariate normal Z has density

$$f_{Z_1,Z_2}(z_1,z_2) = \frac{1}{2\pi\sqrt{1-\rho^2}} \times \exp\left[-\frac{1}{2(1-\rho^2)}\left(z_1^2 - 2\rho z_1 z_2 + z_2^2\right)\right].$$

It should be noted that this density has only one parameter, namely ρ, and plays an important role in the computation of the ALC, as explained in Appendix 4.

The density of the bivariate lognormal distribution Y is

$$f_{Y_1,Y_2}(y_1,y_2) = \frac{1}{2\pi|\Sigma|^{1/2}}\frac{1}{y_1 y_2}\exp\left(-\frac{1}{2}Q\right),$$

where

$$Q = (\log(y_1)-\mu_1, \log(y_2)-\mu_2)'\Sigma^{-1}\begin{pmatrix}\log(y_1)-\mu_1 \\ \log(y_2)-\mu_2\end{pmatrix}$$

is a quadratic form, and μ and Σ are the parameters of the corresponding bivariate normal distribution.

Figure 3 (Ellipse) and Figure 4 (QQ-Plot) were based on Theorem 1. A more complete version of that theorem is stated here.

THEOREM 1 *Let X be distributed as a bivariate normal with parameters (μ,Σ) where $|\Sigma| > 0$. Then*

(i)

$$Q = (X_1-\mu_1, X_2-\mu_2)'\Sigma^{-1}\begin{pmatrix}X_1-\mu_1 \\ X_2-\mu_2\end{pmatrix}$$

is distributed as the chi-square distribution with 2 degrees of freedom, denoted by χ_2^2,

(ii) The bivariate normal distribution assigns probability $1-\alpha$ to the interior of the ellipse

$$\left\{x = \begin{pmatrix}x_1 \\ x_2\end{pmatrix} : (x_1-\mu_1, x_2-\mu_2)'\Sigma^{-1}\begin{pmatrix}x_1-\mu_1 \\ x_2-\mu_2\end{pmatrix} \le \chi_2^2(1-\alpha)\right\},$$

where $\chi_2^2(1-\alpha)$ denotes the $(1-\alpha)$th quantile of χ_2^2.

The quadratic form, Q of Theorem 1 (i), may also be re-stated as

$$Q = \frac{1}{1-\rho^2}[Z_1^2 - 2\rho Z_1 Z_2 + Z_2^2].$$

This alternative form of Q was used to plot the ellipse given in Figure 3 upon replacing the parameters μ and Σ by their respective sample estimates $\hat{\mu}$ and $\hat{\Sigma}$.

Finally, regarding Figure 4, the x values of the points used to plot the QQ-Plot were the quantiles of a chi-square with two degrees of freedom evaluated at $(i-0.5)/n$, $1 \le i \le n$. The corresponding y values were based on ordered statistics corresponding to the squared generalized distance statistics, d_j^2s.

Appendix 4. An algorithm for computing average loss cost based on a bivariate lognormal distribution

The formula (9) above is reproduced here as (A4.1)

$$\mathrm{ALC}_{\mathrm{PD\&TE}}(D_1, D_2, A, L; \mu, \Sigma) = \iint g(y_1, y_2) f_{Y_1,Y_2}(y_1, y_2)\, dy_1\, dy_2 \qquad \text{(A4.1)}$$

where (D_1, D_2, A, L) are the loss-sensitive features of the policy,

(μ, Σ) are the parameters of the bivariate lognormal distribution, estimated from the data using $(\hat{\mu}, \hat{\Sigma})$,

$g(y_1, y_2)$ is the CPF, see Section 5, and

$f_{Y_1,Y_2}(y_1, y_2)$ is the bivariate density of lognormal (see Appendix 3).

In order to estimate the double integral in (A5.1), it would be helpful to break up this task into a number of steps as follows:

(a) Transform the variables (y_1, y_2) to (z_1, z_2),

(b) Divide the plane into N^2 rectangles (grids) as defined below,

(c) Compute the value of the CPF at the center of each rectangle,

(d) Compute the probability that a pair of losses—a random vector—will fall in the rectangle, and finally,

(e) Multiply the value of the CPF from step (c) by the probability from step (d) and sum these terms for all N^2 rectangles to obtain an estimate for ALC.

It may be helpful to provide more details with regard to the above steps.

Step (a). Replace the bivariate normal distribution parameters by their maximum likelihood estimates (see Section 4). Then use the following relationship

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} \exp(\hat{\mu}_1 + \hat{\sigma}_1 z_1) \\ \exp(\hat{\mu}_2 + \hat{\sigma}_2 z_2) \end{pmatrix}.$$

Step (b). For a standard normal, the univariate case, the chance that it takes a value outside the interval $[-3, 3]$ is small, being 0.0026998. By being conservative, it can be stated that **practically** no values of Z_is, $i = 1, 2$ would lie outside the interval $[-10, 10]$.

In order to evaluate the double integral given by (A4.1) in (z_1, z_2), the plane R^2 is partitioned into N^2 rectangles as described below. Suitable values for N would be 100 or 1000, giving rise to 10,000 or 1,000,000 rectangles.

The process of creating these N^2 rectangles begins by partitioning the horizontal and the vertical axes into intervals.

On the horizontal axis, partition the interval $[-10, 10]$ into subintervals using points $z_{1,i}$, where

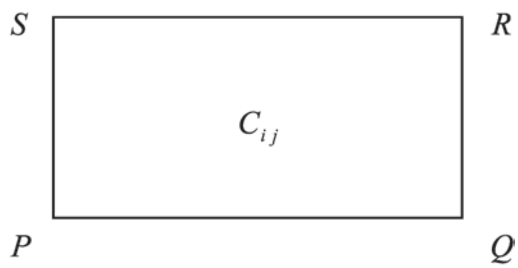
$$z_{1,i} = -10 + \frac{20}{N} i, \qquad 1 \le i \le N.$$

The same type of partitioning is done on the vertical axis by using points $z_{2,j}$, where

$$z_{2,j} = -10 + \frac{20}{N} j, \qquad 1 \le j \le N.$$

The rectangle I_{ij}, $1 \le i \le N$ and $1 \le j \le N$ (see Figure 6) has four corner points given by $P = (z_{1,i-1}, z_{2,j-1})$, $Q = (z_{1,i}, z_{2,j-1})$,

Figure 6. Rectangle I_{ij}



$R = (z_{1,i}, z_{2,j})$, and $S = (z_{1,i-1}, z_{2,j})$. The center of the rectangle is at the point $C_{ij} = (\bar{z}_{1,i}, \bar{z}_{2,j})$ where

$$\bar{z}_{1,i} = \frac{z_{1,i-1} + z_{1,i}}{2} \quad \text{and}$$
$$\bar{z}_{2,j} = \frac{z_{2,j-1} + z_{2,j}}{2}.$$

Step (c). For each rectangle I_{ij} defined above, the value of the CPF, $g(y_1, y_2)$, is evaluated at the point C_{ij}, the center of the rectangle, using the formula

$$g(\exp(\hat{\mu}_1 + \hat{\sigma}_1 \bar{z}_{1,i}), \exp(\hat{\mu}_2 + \hat{\sigma}_2 \bar{z}_{2,j})). \qquad \text{(A4.2)}$$

The value of $g(y_1, y_2)$, depends upon where the point C_{ij} resides, according to the ten regions specified in Section 5 above.

Step (d). It is necessary to compute the probability that the random vector Z, a standardized bivariate normal, would lie in the rectangle I_{ij}, i.e., $P(I_{ij})$. This probability can be determined in terms of the cdf of a standardized bivariate normal F as

$$F(z_1, z_2) = P(Z_1 \le z_1 \cap Z_2 \le z_2).$$

Thus,

$$P(I_{ij}) = F(z_{1,i}, z_{2,j}) - F(z_{1,i}, z_{2,j-1}) - F(z_{1,i-1}, z_{2,j}) + F(z_{1,i-1}, z_{2,j-1}). \qquad \text{(A4.3)}$$

Equation (A4.3) can be evaluated by estimating the cdf of a standardized bivariate normal evaluated at the four corner points P, Q, R, and S of the rectangle (see Figure 6 above). There is no standard function available in Excel for computing $F(z_1, z_2)$. Fortunately, there is an algorithm provided by Drezner and Wesolowsky [2] for estimating $F(z_1, z_2)$ that can be programmed relatively easily (see Appendix 5 for more details).

Step (e). Finally, the double integral in (A4.1) is replaced by the sum

$$\sum_{k=1}^{10} \sum_{I_{ij} \varepsilon R_k} \{g(\exp(\hat{\mu}_1 + \hat{\sigma}_1 \bar{z}_{1,i}), \exp(\hat{\mu}_2 + \hat{\sigma}_2 \bar{z}_{2,j}))\} P(I_{ij}). \qquad \text{(A4.4)}$$

A program written in Excel VBA was used to calculate the sum given by (A4.4).

Appendix 5. Drezner and Wesolowsky algorithm for computing the cumulative distribution function of a standardized bivariate normal distribution

Below is an outline of the Drezner and Wesolowsky algorithm [2] for the case $K = 5$, where K is the number of points used in the Gaussian quadrature. The interested reader should refer to the above paper in order to gain additional insights into their algorithm. The Excel VBA program referred to in this paper uses their algorithm as sketched below to calculate ALC.

Let $F_{Z_1,Z_2}(z_1, z_2; \rho)$ be the cdf of a standardized bivariate normal distribution with the parameter ρ. The corresponding density is

$$f_{Z_1,Z_2}(z_1, z_2; \rho) = \frac{1}{2\pi\sqrt{(1-\rho^2)}} \times \exp\left[-\frac{1}{2(1-\rho^2)}\left(z_1^2 - 2\rho z_1 z_2 + z_2^2\right)\right].$$

See Appendix 3.

Let $\Phi(z)$ denote the cdf of a standard univariate normal. In Excel it is the function NORMDIST.

The tail probability for a standardized bivariate normal is denoted by $L(z_1, z_2; \rho)$, where

$$L(z_1, z_2; \rho) = \int_{z_1}^{\infty} \int_{z_2}^{\infty} f_{Z_1,Z_2}(u, v)\, du\, dv.$$

Drezner and Wesolowsky state that

$$F_{Z_1,Z_2}(a,b;\rho) = L(a,b;\rho) + \Phi(a) + \Phi(b) - 1. \tag{A5.1}$$

Furthermore, Drezner and Wesolowsky approximate $L(a,b;\rho)$ according to

$$L(a,b;\rho) \approx 2\pi\rho \sum_{i=1}^{5} w_i f_{Z_1,Z_2}(a,b;\rho x_i) + \Phi(-a)\Phi(-b) \tag{A5.2}$$

where

x_i	w_i
0.04691008	0.018854042
0.23076534	0.038088059
0.5	0.045270394
0.76923466	0.038088059
0.95308992	0.018854042.

Since $f_{Z_1,Z_2}(z_1,z_2;\rho)$ can be easily programmed, and $\Phi(z)$ is a standard function (at least in Excel), then it is not hard to compute (A5.1) provided that $L(a,b;\rho)$ can be approximated based on formula (A5.2).

Appendix 6. Notes on simulation

Simulation can provide an alternative method for computing ALC. An advantage of the simulation approach is **flexibility**. One disadvantage of a simulation is that it may require longer computing times. Another disadvantage is that it may require storing the values of the intermediate simulated losses in order to compute the ALC.

It is easy to generate independent univariate normal variates. The more challenging requirement is to generate **correlated** normal variates. Fortunately, this can be accomplished on the basis of a known result from Linear Algebra called the Choleski decomposition (refer to Venables and Ripley [11]). Choleski decomposition states that a positive definite matrix can be written as a product of a lower triangular matrix and its transposition. In particular, one can write

$$\Sigma = LL'$$

where Σ is a variance-covariance of a bivariate normal distribution, and L is a lower triangular matrix as defined below.

The necessary steps to simulate a pair of (y_1, y_2) values according to a bivariate lognormal distribution with parameters (μ, Σ) are:

Step 1. Generate two independent standard (univariate) normal variates and label them as V_1 and V_2. The Excel functions RAND and NORMINV can help to accomplish this task.

Step 2. Let

$$V = \begin{pmatrix} V_1 \\ V_2 \end{pmatrix}$$

based on Step 1.

Define

$$X = \mu + LV,$$

where

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \qquad L = \begin{pmatrix} \sigma_1 & 0 \\ \sigma_2\rho & \sigma_2\sqrt{1-\rho^2} \end{pmatrix}.$$

Then, X is distributed as bivariate normal with parameters (μ, Σ). If the values of (μ, Σ) are not known, then their estimates should be substituted.

Step 3. Let

$$Y = \begin{pmatrix} \exp(X_1) \\ \exp(X_2) \end{pmatrix}$$

where

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

has been computed according to Step 2.

Step 4. Repeat Steps 1, 2, and 3, n times as necessary.