Empirical Investigation of Insurance Claim Dependencies

Emiliano A. Valdez, Ph.D., F.S.A., F.I.A.A.
School of Actuarial Studies
Faculty of Commerce and Economics
University of New South Wales
Sydney, AUSTRALIA
e-mail: e.valdez@unsw.edu.au

20 January 2006

Abstract

There has been a great deal of recent interest in the modeling of insurance risks to incorporate the presence of dependencies, and some early work in the literature has demonstrated that for a typical dependent insurance portfolio, ignoring dependencies can have a direct impact on the tail or extremes of the overall portfolio loss distribution. The tail of the loss distribution is a typical concern to the actuary. To date, in spite of the growing number of papers in the literature on dependence, there is no known published work that provides for an empirical investigation that validates the presence of dependencies in an insurance portfolio. In this paper, we use mixture models, which are commonly applied in the modeling of dependent credit risks, to facilitate in the understanding of claim dependencies. The empirical data used in this paper comes from a portfolio of automobile insurance provided by the General Insurance Association which is an organization of general insurance companies in Singapore. Because of the volume of data from our source, we reasonably selected a portfolio of policies drawn from a randomly selected insurance provider. To measure up the presence of claim dependencies, the most reasonable statistic to use is the relative risk, a measure that is widely popular in medical statistics and is used to gauge how the presence of a particular insurance risk induces another insurance risk to go on claim. Our calibration results indicate some presence of dependencies; relative risk is in the neighborhood of 14%. To accommodate for the presence of covariates, we also introduce mixture models with covariates as explained in this article. Not surprisingly, because the premium is the actuary's best guess of the degree of riskiness of an insurance risk, it provides for the single most important factor that influences the presence of claim dependencies.

Acknowledgements: This research has been funded and supported by the Actuarial Education Research Fund (AERF) and the Casualty Actuarial Society (CAS), and the author thanks both organization for providing the funding. The author also extends appreciation for the research assistance provided by Mr. Mitchell Wills of the University of New South Wales. Submitted to: *Journal of Casualty Actuarial Science.*

1 Introduction and motivation

The aim of this work is to provide for an empirical investigation of the presence of "dependencies" in individual claims for a portfolio of insurance contracts. For apparent reasons of simplicity and tractability, it is typical to assume that the claims are independent in traditional risk theory. On the other hand, it is well recognized that the assumption of independence may be unrealistic as claims may exhibit some form of "dependencies". Individual risks may be exposed to the same physical or economic environment that may be driving resulting claims to be in some sense, dependent. A classic example of this is in home insurance where coverage is provided to a group of residences with contiguous houses in a particular location or region. These homes may share a common exposure to disasters (windstorm, hurricane, fire, earthquake, etc.) that may lead to possible catastrophic financial damages. As yet another classic example is in group insurance coverage where an organization provides for health and life insurance to its employees who may be working in a manufacturing plant. A severe mechanical breakdown or explosion can have damaging repercussions on the life and health of the employees.

Our work focuses on portfolios of automobile insurance contracts and the possible presence of claim dependencies here may be qualitatively justified in several respects. First, for most car accidents, it is the case that more than one party may be involved. In a collision for example, at least two car drivers would be involved, and additionally, there may be other passengers of the vehicles in the accident. Second, poor weather conditions sometimes contribute to poor driving and traffic conditions possibly causing multiple number of accidents. Thirdly, it is possible that in some neighborhood, some roads and traffic lights are poorly engineered, and until these poor road conditions are corrected, there can be multiple occurrences of car accidents in the neighborhood. Lastly, it is possible for an insured vehicle to have multiple accidents leading to multiple claims within the policy period.

Consider a portfolio of n insurance risks and for each individual risk, denote the Bernoulli indicator for claims with I_1, I_2,..., I_n. Denote the probability of a claim by $q_k = P\left(I_k = 1\right)$ for $k = 1, 2, ..., n$ so that $1 - q_k = P\left(I_k = 0\right)$ is its complement, the probability of no claim.

Clearly, the total number of claims in the portfolio is $S = \sum_{k=1}^{n} I_k$. The indicator for claims has traditionally been referred to in the actuarial literature as the claim frequency. In contrast, the claim severity refers to the amount or cost when an accident occurs. In this,article we shall be concerned with dependencies on the claim frequency rather than on the claim severity. As pointed out by Denuit, Lefevre and Utev. (2002), there are many situations whereby the dependence affects the occurrences of claims rather than the claim severities. Our investigation then is concerned with the modeling of the joint multivariate distribution of the claim occurrence vector $\mathbf{I} = (I_1, ..., I_n)'$.

The dimension n for insurance portfolios is not uncommon to be large and in drawing inference from data where the number of policy exposure is generally large, the issue of tractability remains a concern. Our dataset alone provides for a total number of exposures of 210,446 in a period of nine calendar years. We chose to use dependence through mixing for tractable reasons and additionally, for its intuitive appeal. The construction is based on the fundamental assumption that, conditional on an unobserved random vector $\mathbf{Z}$, the random variables $I_k |\mathbf{Z}$ for all $k = 1, ..., n$ are independent. These unobservables help to explain the forces that may be driving the dependencies. For these types of models, they have been referred to in the literature on credit risks as mixture models. See, for example, McNeil, Frey and Embrechts (2005). Mixture models have been used to understand the presence of dependencies among the risks in credit portfolios. The applicability in the insurance setting is not surprising because there are similarities between the occurrence of claims and the occurrence of credit default.

We need measures of dependence to further help us understand the presence of claim dependencies. The correlation measure usually helps explain the presence of linear relationships on the data. This measure, for instance, can be computed pair-wise and if we assume exchangeability, this same correlation is applicable for any pairs of claims. All throughout the analysis presented here, exchangeability has been assumed for our data. As alternative measure, we also use the relative risk measure to quantify how the claim that arises from a single policy can adversely affect the probability that another policy will also claim. Relative risk measures have been widely used in medical statistics to help understand how different factors affect the occurrence of diseases. In probabilistic terms, the relative risk measure is defined as

$$\delta = \frac{P\left(I_k = 1 \,|I_{k^*} = 1\right)}{P\left(I_k = 0 \,|I_{k^*} = 1\right)}$$

for any pairs $k \neq k^*$. Our data reveals observed correlations of about 4% and relative risk measures of about 14%. We also found that when premiums are additionally considered as covariates, the degrees to which these dependencies increase with the increase in premiums. The premium provides the single most important covariate that influences the presence of claim dependencies. This is not at all surprising considering the fact that premium is usually condered the actuary's best estimate of the overall degree of riskiness. Furthermore, the weak dependencies are not at all surprising either; this is because generally there is a large volume of policy exposures for which we are capturing overall dependence.

The results of our work may have far-reaching implications to the practicing actuary who may be concerned about the financial consequences of assuming independent claims when in fact there may be forces driving claim dependencies. The actuary can have better information to make more informed decisions about pricing, assessment of solvency, claims prediction, to name just a few. This paper suggests innovative procedures for examining the presence of such dependencies.

The rest of the paper has been organized as follows. First, in Section 2, we explore some of the different models of claim dependence that have appeared in the literature. As already pointed out earlier, our primary interest is modeling the dependence on the claim frequency and so we focus on the multivariate structure of the binary outcome of the claim occurrences. Many of these models may not be practicable to implement; some leads to over-parameterization and therefore non-identifiability of the model parameters, while others are not tractable enough to handle the dimensionality issue earlier described. Next in Section 3, we discuss the construction of the claim dependence model through mixing, or mixture models. Section 4 discusses our data, the procedures of model estimation, and of course, the results of these estimations. Finally, in Section 5, we give brief statements as final remarks.

2 Models of claim dependence

While it is true that there has been a growing number of papers in the literature addressing the issue of dependencies on claims, surprisingly no paper has provided empirical evidence to examine and test the validity of the independence assumption. The notion of claim dependencies has increasingly become an important part of the modeling process. WANG (1998) suggests a set of statistical tools for modeling dependencies of risks in an insurance portfolio. ALBRECHER AND KANTOR (2002) and VALDEZ AND MO (2002) both examine the impact of claim dependencies on the probability of ruin under the copula framework. HEILMANN (1986) and HURLIMANN (1993) investigated the effect of dependencies of risks on stop-loss premiums. Several generalizations and alternative models of dependence have since followed including, DHAENE AND GOOVAERTS (1997) and MULLER (1997), addressing their impact on stop-loss premiums. Other models have included the works of COSSETTE, GAILLARDETZ, MARCEAU AND RIOUX (2002) and GENEST, MARCEAU AND MESFIOUI (2003) where claim dependence have been addressed in the framework of individual risk models. Furthermore, using the notion of a stochastic order, the recent papers by PURCARU AND DENUIT (2002, 2003) provide excellent discussion of dependencies in claim frequency for credibility models. More recently, FREES AND WANG (2005) considered a generalized linear model framework for modeling marginal claims distributions and allowed for dependence using t-copulas.

Now, consider again a portfolio of n insurance policies during some well-defined fixed reference period. Denote by $\mathbf{I} = (I_1, I_2, ..., I_n)'$ the random vector of claim indicators, that is, for each policy k, $k = 1, ..., n$, comes with a random variable giving an indication for claims:

$$I_k = \begin{cases} 0, & \text{if no claim occurs} \\ 1, & \text{if a claim occurs} \end{cases}.$$

Its joint probability function is therefore

$$p(\mathbf{I}) = P(I_1 = i_1, ..., I_n = i_n) \text{ for } i_k \in \{0, 1\}, \ k = 1, ..., n. \tag{1}$$

Its mean vector will be $\mathbf{q} = (q_1, q_2, ..., q_n)'$ where each $q_k = P(I_k = 1)$ gives the probability of a claim for policy k and clearly the marginal claim probabilities.

In the following examples, we consider various models of claim dependence. Some more other complex models of claim dependence have appeared in the literature but these are extremely difficult for data calibration.

Example 2.1: *The multivariate Bernoulli*

One can always view the random vector $\mathbf{I}$ with joint probability function $p(\mathbf{I})$ in (1) and with marginal claim probabilities $\mathbf{q}$ as having the most general possible case of a multivariate Bernoulli. Indeed, it is a special case of a multinomial distribution. In this case, one would have to estimate the probabilities for all possible combinations of the indicator random variable and there are a total of $2^n - 1$ parameters needed. Clearly, in this case the random variable I_k has a Bernoulli(q_k) where q_k can be computed from the joint probabilities using

$$q_k = \sum_{i_1=0}^{1} \cdots \sum_{i_{k-1}=0}^{1} \sum_{i_{k+1}=0}^{1} \cdots \sum_{i_n=0}^{1} p(\mathbf{I})$$

with $i_k = 0$ or 1 for each $\mathbf{I}$ in the sums. Because of the non-independence of the Bernoulli random variables, the sum is no longer a Binomial but is rather a correlated binomial random variable. Because of the large number of parameters to estimate, particularly as n gets large, this is not often useful for many practical applications.■

Example 2.2: *The Frechet classes*

Consider the Frechet space of all indicator random vectors $\mathbf{I}$ with given marginal claim probabilities $\mathbf{q}$. The extremal joint distribution functions within the Frechet space consist of the Frechet upper and lower bounds. It is well-known that for any random vector $\mathbf{X} = (X_1, X_2, ..., X_n)'$, its joint distribution function is bounded by

$$\max\left(0, F_{X_1}(x_1) + \cdots + F_{X_n}(x_n) - n + 1\right) \le F_{\mathbf{X}}(\mathbf{x}) \le \min\left(F_{X_1}(x_1), \cdots, F_{X_n}(x_n)\right).$$

In the context of the indicator random vector, the upper bound corresponds to the distribution of a so-called comonotonic random vector defined by

$$\mathbf{I}^{c\backslash} = (I_1^c, I_2^c, ..., I_n^c)' = \left(I\left(U \ge 1 - q_1\right), ..., I\left(U \ge 1 - q_n\right)\right)'$$

where U is a Uniform on $[0,1]$ random variable and $I(\cdot)$ is an indicator function which gives a value of 1 if the statement is true and 0 otherwise. The lower bound, on the other hand, provided that $\sum_{k=1}^{n} q_k \leq 1$, which often is true in practice as pointed out by Denuit, Lefevre and Utev (2002), corresponds to the case of a counter-comonotonic random vector whose elements are mutually exclusive, that is,

$$P(I_k = 1) = q_k \text{ and } P(I_k = 1, I_l = 1) = 0 \text{ for all } k \neq l \in \{1, 2, ..., n\}.$$

These Frechet distribution bounds are used typically for describing perfect positive associations (in the comonotonic case) and perfect negative associations (in the counter-comonotonic case). Random vectors of the Frechet classes have been discussed and considered in the actuarial literature by Dhaene and Goovaerts (1997), Muller (1997), Hu and Wu (1999), and Dhaene, et al. (2000a, 2000b).■

Example 2.3: *A common global shock model*

Marceau, Cossette, Gaillardetz and Rioux (1999) considered a construction of the individual risk model by introducing a common global shock that can induce claims among the individual policies and hence dependence on the claims. In the construction, each Bernoulli random variable I_k is decomposed as

$$I_k = \min(J_k + J_0, 1) \text{ for each } k \in \{1, 2, ..., n\}$$

where $J_0, J_1, ..., J_n$ are independent Bernoulli random variables with $P(J_k = 1) = r_k$ for $k \in \{0, 1, 2, ..., n\}$. Clearly, the individual policies share the common global shock variable J_0 and this induces the association among the policies. In particular, I_k is also clearly a Bernoulli random variable with parameter

$$P(I_k = 1) = r_0 + (1 - r_0) r_k.$$

However simple this model construction may be, we find that without additional assumptions, this model leads to non-identifiable parameters.■

Example 2.4: *A common local shock model*

Extending the construction of Marceau, et al. (1999) introduced in the previous example, one can also construct common shocks only to certain subgroups of the insured groups. For example, suppose we could sub-divide the n individual risks into m different sub-classes with each sub-class having n_j individual risks, for $j \in \{1, 2, ..., m\}$. Here then we have $n = \sum_{j=1}^{m} n_j$. A local common shock J_{kj} is then introduced for each member in a subclass $j \in \{1, 2, ..., m\}$ as follows:

$$I_{kj} = \min\left(J_{kj} + J_k, 1\right) \text{ for each } j \in \{1, 2, ..., m\}, k \in \{1, ..., n_j\}$$

where both J_{kj} and J_k are independent Bernoulli random variables with $P\left(J_k = 1\right) = r_k$ and $P\left(J_{kj} = 1\right) = r_{kj}$. The random variables I_{kj} remain Bernoulli random variables but are no longer independent. A variation to such construction is to simultaneously introduce the common global shock and in which case we have the further decomposition as

$$I_{kj} = \min\left(J_{kj} + J_k + J_0, 1\right) \text{ for each } j \in \{1, 2, ..., m\}, k \in \{1, ..., n_j\}.$$

Such construction can have an intuitive appeal because of its interpretation. In effect, we are saying that an individual claim arises because of a shock either introduced locally from the policy belonging to a homogeneous group or introduced globally from the common environment shared by all the policies, or it could be caused by some characteristic that is just peculiar to the individual policy. This makes sense for instance in a housing insurance where a claim can arise because a fire has been caught in the neighborhood or because a typhoon or storm or hurricane or some other natural disaster swept the entire region.■

Example 2.5: *Archimedean copula models*

This construction has been motivated by GENEST ET AL. (2000) where the dependence in the individual risks has been introduced in the random vector **I** by specifying the joint distribution using the copula representation $P\left(I_1 \leq i_1, ..., I_n \leq i_n\right) = C_\varphi\left(F_1\left(i_1\right), ..., F_n\left(i_n\right)\right)$ where $F_1, ..., F_n$ denotes the marginal distribution functions of $I_1, ..., I_n$, respectively, and where C_φ is an Archimedean copula which has the form

$$C_\varphi\left(u_1, ..., u_n\right) = \varphi^{-1}\left(\varphi\left(u_1\right) + \cdots + \varphi\left(u_n\right)\right) \tag{2}$$

where the generator φ is a mapping $\varphi : [0, 1] \to [0, \infty)$ that is continuous, strictly decreasing and satisfying $\varphi(0) = \infty$ and $\varphi(1) = 0$. Its inverse is denoted by φ^{-1} and its inverse must satisfy the property of complete monotonicity defined by $(-1)^m d^m\varphi^{-1}(s)/ds^m \geq 0$ (i.e. alternating signs of derivatives) for all integers m. This complete monotonocity property ensures that (2) defines a proper multivariate distribution for any portfolio size n. See Nelsen (1999) for many examples of generators inducing Archimedean copulas. One of the major criticisms of Archimedean copulas is being unable to specify a model rich in dependence parameters. Many Archimedean copulas have single parameter that is supposed to capture all the different forms of dependencies possible. Nevertheless, they are more easily tractable than many other copula forms. In particular, some find them useful in situations where the portfolios are "relatively small and homogeneous." See Frey and McNeil (2003).■

Example 2.6: *Latent variable models and copulas*

Latent variable models have more frequently appeared in the finance literature for modeling dependent credit default but they can also be helpful in constructing models for claim dependence as there is similarity between credit default and incidence of claims. The construction generally is as follows. Consider an n-dimensional continuous random vector $\mathbf{X}$ with continuous marginal distributions that do represent the so-called latent variables. Let the vector $\mathbf{D} = (D_1, ..., D_n)'$ represent the deterministic cut-off levels. We call $(\mathbf{X}, \mathbf{D})$ the latent variable model for the Bernoulli random vector $\mathbf{I}$ satisfying the following relationships:

$$I_k = 1 \text{ if and only if } X_k \leq D_k \text{ for all } k \in \{1, ..., n\}.$$

Such a construction provides for richer models of dependence as one for example can specify multivariate joint distributions of the continuous vector $\mathbf{X}$ via copulas. Copulas are the joint multivariate distribution functions of random vectors whose marginals are standard Uniform. According to Sklar's Theorem, we can write this joint distribution as

$$F_{\mathbf{X}}(x_1, ..., x_n) = P(X_1 \leq x_1, ..., X_n \leq x_n) = C(F_1(x_1), ..., F_n(x_n)) \tag{3}$$

for some so-called copula function C and where $F_1, ..., F_n$ denotes the marginal distribution functions. The copula function contains the parameter or vector of parameters that describes the dependence, and the copula function is unique in the case of continuous marginals. For latent variable models, the random vector $\mathbf{X}$ is usually one with continuous marginal distribution functions. Furthermore, in the continuous case, we can easily extract the unique copula function C from the joint multivariate distribution using

$$C\left(u_1, ..., u_n\right) = F_{\mathbf{X}}\left(F_1^{-1}(x_1), ..., F_n^{-1}(x_n)\right) \tag{4}$$

where $F_1^{-1}, ..., F_n^{-1}$ are the inverses (quantiles) of the continuous distribution functions. There are now many textbooks that provide good fundamental instructions on copulas, among them include Joe (1997) and Nelsen (1998). Some references that provide nice introduction with actuarial applications include the work of Frees and Valdez (1998) and Wang (2000).■

Example 2.7: *Time-until-claim and common shock models*

The construction for this type of dependence model relies on the random variable which refers to the time-until-default. Li (2000) for example, uses this time-until-default concept to construct dependencies. Here, as a special case, we discuss common shock models discussed by Marshall and Olkin (1967). Consider the bivariate construction where we begin with two dependent random variables T_1 and T_2 and we assume that the shocks follow three independent Poisson processes with parameters λ_1, λ_2 and λ_{12}, assumed to be all non-negative. These parameters, respectively indicate whether the shocks affect only the first variable, the second variable, or both. It follows that the times X_1, X_2 and X_{12} of occurrence of these shocks have independent exponential distribution so that one can write

$$P\left(T_1 > t_1, T_2 > t_2\right) = P\left(X_1 > t_1\right) P\left(X_2 > t_2\right) P\left(X_{12} > \max\left(t_1, t_2\right)\right).$$

One can also write $T_1 = \min\left(X_1, X_{12}\right)$ and $T_2 = \min\left(X_2, X_{12}\right)$, and derive the above probability. The copula representation, as well as multivariate extension, of the Marshall and Olkin common shock construction can also be found in Embrechts, Lindskog and McNeil (2001). More detailed construction of Common Poisson shock models with applications to insurance and credit risk modeling can be found in Lindskog and McNeil (2003).■

3 Dependence through mixing

This section discusses the general framework that we adopt in our data calibration to empirically validate the presence (or absence) of dependence in a portfolio of insurance risks. As stated in the introduction, mixture models have appeared in the credit risk literature and

I found the textbook by McNeil, Frey and Embrechts (2005) to provide very useful details in the context of credit risks.

3.1 The generic construction

Consider once more the multivariate random vector $\mathbf{I}$ and supposedly that conditional on an unobservable random vector $\mathbf{Z} = (Z_1, ..., Z_p)'$, the random variables

$$I_{k|\mathbf{Z}} = I_k \left| \mathbf{Z} \right. \text{ for all } k \in \{1, 2, ..., n\}$$

are independently distributed, but not necessarily identically distributed. In the case of identical distribution, we say that the random variables are exchangeable. Generally the dimension p of $\mathbf{Z}$ is smaller than the dimension in the portfolio n, i.e. $p < n$. Additionally, we assume there are functions $Q_k : \mathbf{R}^p \rightarrow [0, 1]$ for $k \in \{1, 2, ..., n\}$ such that

$$P\left(I_k = 1 \left| \mathbf{Z} \right.\right) = p_{k|\mathbf{Z}}\left(1 \left| \mathbf{z} \right.\right) = Q_k\left(\mathbf{Z}\right). \tag{5}$$

It follows that

$$P\left(I_1 = i_1, ..., I_n = i_n \left| \mathbf{Z} \right.\right) = \prod_{k=1}^{n} p_{k|\mathbf{Z}}\left(i_k \left| \mathbf{z} \right.\right) = \prod_{k=1}^{n} Q_k\left(\mathbf{Z}\right)^{i_k}\left(1 - Q_k\left(\mathbf{Z}\right)\right)^{1-i_k}$$

where we have $i_k \in \{0, 1\}$.

We assume that $\mathbf{Z}$ has cumulative distribution function denoted by $F_{\mathbf{Z}}$ with support that is a subset of $\mathbf{R}^p$. We can then re-express the unconditional marginal probability as

$$q_k = P\left(I_k = 1\right) = \int Q_k\left(\mathbf{Z}\right) dF_{\mathbf{Z}}\left(\mathbf{z}\right) = \mathrm{E}_{\mathbf{Z}}\left[Q_k\left(\mathbf{Z}\right)\right]$$

where the integration is over the support of $F_{\mathbf{Z}}$. Furthermore, we can describe the joint (unconditional) probability distribution of $\mathbf{I}$ by

$$\begin{aligned} p_{\mathbf{I}}\left(i_1, ..., i_n\right) &= P\left(I_1 = i_1, ..., I_n = i_n\right) \\ &= \int \left[\prod_{k=1}^{n} Q_k\left(\mathbf{Z}\right)^{i_k}\left(1 - Q_k\left(\mathbf{Z}\right)\right)^{1-i_k}\right] dF_{\mathbf{Z}}\left(\mathbf{z}\right) \end{aligned} \tag{6}$$

The unobservable random variable $\mathbf{Z}$ understandably induces the dependence among the responses $I_1, ..., I_n$ and in the special case where $\mathbf{Z}$ is degenerate, so that say $\mathbf{Z} = \mathbf{z}_0$ with probability one and $\mathbf{z}_0$ is a vector with fixed constants, then the model in (6) yields the case of independence. This is immediately seen from the fact that then the joint probability in (6) can be written as simply the product of functions of each marginal probabilities alone. Notice that in the case where this mixing variable has a degenerate distribution, then the model in (6) becomes the case of independence:

$$p_{\mathbf{I}}\left(i_1, ..., i_n\right) = P\left(I_1 = i_1\right) \times \cdots \times P\left(I_n = i_n\right). \tag{7}$$

The unobservable $\mathbf{Z}$ is called a mixing variable and the resulting model is sometimes called a Bernoulli mixture model. See for example, Koyluoglu and Hickman (1998) and Gordy (2000).

Notice that the probability of a simultaneous occurrence of claims in the portfolio can be expressed as

$$p_{\mathbf{I}}\left(1, ..., 1\right) = \int \left[\prod_{k=1}^{n} Q_k\left(\mathbf{Z}\right)\right] dF_{\mathbf{Z}}\left(\mathbf{z}\right) = \mathrm{E}_{\mathbf{Z}}\left[\prod_{k=1}^{n} Q_k\left(\mathbf{Z}\right)\right].$$

For any $k \neq k^* \in \{1, 2, ..., n\}$, we have bivariate joint probabilities for any pairs given by

$$P\left(I_k = 1, I_{k^*} = 1\right) = \mathrm{E}_{\mathbf{Z}}\left[Q_k\left(\mathbf{Z}\right) Q_{k^*}\left(\mathbf{Z}\right)\right].$$

The covariance for any pairs can be expressed as

$$Cov\left(I_k, I_{k^*}\right) = \mathrm{E}_{\mathbf{Z}}\left[Q_k\left(\mathbf{Z}\right) Q_{k^*}\left(\mathbf{Z}\right)\right] - \mathrm{E}_{\mathbf{Z}}\left[Q_k\left(\mathbf{Z}\right)\right] \mathrm{E}_{\mathbf{Z}}\left[Q_{k^*}\left(\mathbf{Z}\right)\right].$$

Finally, the correlation coefficient for any pairs is thus given by the following expression:

$$\begin{aligned} \rho\left(I_k, I_{k^*}\right) &= \frac{Cov\left(I_k, I_{k^*}\right)}{\sqrt{Var\left(I_k\right) Var\left(I_{k^*}\right)}} \\ &= \frac{\mathrm{E}_{\mathbf{Z}}\left[Q_k\left(\mathbf{Z}\right) Q_{k^*}\left(\mathbf{Z}\right)\right] - \mathrm{E}_{\mathbf{Z}}\left[Q_k\left(\mathbf{Z}\right)\right] \mathrm{E}_{\mathbf{Z}}\left[Q_{k^*}\left(\mathbf{Z}\right)\right]}{\sqrt{\mathrm{E}_{\mathbf{Z}}\left[Q_k\left(\mathbf{Z}\right)\right] \mathrm{E}_{\mathbf{Z}}\left[Q_{k^*}\left(\mathbf{Z}\right)\right]\left(1 - \mathrm{E}_{\mathbf{Z}}\left[Q_k\left(\mathbf{Z}\right)\right]\right)\left(1 - \mathrm{E}_{\mathbf{Z}}\left[Q_{k^*}\left(\mathbf{Z}\right)\right]\right)}}. \end{aligned}$$

Note that when the variables concerned are Bernoulli, the correlation may not be the best way to describe the dependence. Although still a useful measure of dependence, the linear dependence that it reveals often for continuous random variables is less understood. As often used in survival modeling, the ratio of relative risk provides a more intuitive interpretation. Here we ask: "how much does one insurance risk induce another insurance risk to go on claim?" This is the important idea of dependence in this context, and the relative risk is defined by

$$\delta\left(I_k, I_{k^*}\right) = \frac{P\left(I_k = 1 \left| I_{k^*} = 1\right.\right)}{P\left(I_k = 0 \left| I_{k^*} = 1\right.\right)}. \tag{8}$$

Using the mixture model framework, we can deduce that

$$P\left(I_k = 1 \left| I_{k^*} = 1\right.\right) = \frac{P\left(I_k = 1, I_{k^*} = 1\right)}{P\left(I_{k^*} = 1\right)} = \frac{\mathrm{E}_{\mathbf{Z}}\left[Q_k\left(\mathbf{Z}\right) Q_{k^*}\left(\mathbf{Z}\right)\right]}{q_{k^*}}$$

and similarly, we can deduce that

$$P\left(I_k = 0 \left| I_{k^*} = 1\right.\right) = \frac{P\left(I_k = 0, I_{k^*} = 1\right)}{P\left(I_{k^*} = 1\right)} = \frac{\mathrm{E}_{\mathbf{Z}}\left[\left(1 - Q_k\left(\mathbf{Z}\right)\right) Q_{k^*}\left(\mathbf{Z}\right)\right]}{q_{k^*}}$$

so that it is clear that the relative risk in (8) has the equivalent representation

$$\delta\left(I_k, I_{k^*}\right) = \frac{\mathrm{E}_{\mathbf{Z}}\left[Q_k\left(\mathbf{Z}\right) Q_{k^*}\left(\mathbf{Z}\right)\right]}{\mathrm{E}_{\mathbf{Z}}\left[\left(1 - Q_k\left(\mathbf{Z}\right)\right) Q_{k^*}\left(\mathbf{Z}\right)\right]}. \tag{9}$$

A number of interpretations giving rise to the dependencies of claims occurrences among the insurance risks can always be given to the mixing variable $\mathbf{Z}$. For example, it could describe a set of common factors, possibly interpreted as unobservable variables that could be driving the dependencies such as weather conditions or poor driving conditions as described in the introduction. Conditional then on these common factors, the claim occurrences of the various risks become independent. A similar type of model can also be used to describe the often unobserved heterogeneity of the insurance risks in the portfolio, so much so that this unobservable random variable can be interpreted as risk characteristics of the individual contract that may be unobservable.

Example 3.1: *Gamma-distributed mixing variable*

Consider the special case where $\mathbf{Z}$ is a vector of independent "standard" Gamma distributed random variables with each having marginal density function

$$f_{Z_j}\left(x\right) = \frac{1}{\Gamma\left(\alpha\right)} x^{\alpha-1} e^{-x} \quad \text{for } x > 0,\ j \in \{1, ..., p\}$$

where α is a positive parameter. Furthermore, the form of the conditional distribution is as follows

$$Q_k\left(\mathbf{Z}\right) = 1 - \exp\left(-\sum_{j=1}^{p} w_{j,k} Z_j\right) = 1 - \exp\left(-\mathbf{w}_j' \mathbf{Z}\right).$$

It has been noted, for example, in Gordy (2000) that this representation leads to the CreditRisk$^+$ model as documented in Credit-Suisse First Boston (1997). However, the CreditRisk$^+$ model is often not expressed as a Bernoulli with a Gamma mixture model, as done here, but rather as a Poisson mixture model.

3.2 Mixture models with covariates

The mixture models described in the previous subsections can be specified with covariates or explanatory variables. Covariates are often introduced into the model to capture the non-homogeneity in the portfolio and are used to understand how they influence the probability of a claim.

To illustrate, suppose we have a set of deterministic r covariates given for each risk in the insurance portfolio. Denote the observed values of these r covariates by the vector $\mathbf{x}_k = (x_{1k}, ..., x_{rk})'$ and these covariates will enter into the model specification via the conditional probability of default as follows

$$P\left(I_k = 1 \left|\mathbf{Z}; \mathbf{x}_k\right.\right) = p_{k|\mathbf{Z}}\left(1 \left|\mathbf{z}; \mathbf{x}_k\right.\right) = Q\left(\mathbf{Z}; \mathbf{x}_k\right) \tag{10}$$

with usually a specified functional form as

$$Q\left(\mathbf{Z}; \mathbf{x}_k\right) = g\left(\mathbf{x}_i'\boldsymbol{\beta} + \boldsymbol{\sigma}'\mathbf{Z}\right) \tag{11}$$

where $g : \mathbf{R} \rightarrow [0, 1]$ is some increasing function such as a distribution function (e.g. $g = \Phi$, the standard Normal c.d.f.), $\boldsymbol{\beta} = (\beta_1, ..., \beta_r)'$ and $\boldsymbol{\sigma} = (\sigma_1, ..., \sigma_p)'$. Here, $\boldsymbol{\beta}$ and $\boldsymbol{\sigma}$ are parameter vectors satisfying also $\boldsymbol{\sigma}' > \mathbf{0}$. The g function is typically known as the link function, particularly in the case of Generalized Linear Models. In a generalized linear mixed model (GLMM) which is a special case of our construction of Bernoulli mixture models with covariates, the random vector $\mathbf{I}$, conditionally on an unobservable random vector $\mathbf{Z}$, are independent with identical distributions coming from the same exponential family (e.g. Bernoulli, Poisson, Gamma, Normal) with mean $E\left(I_k \left|\mathbf{Z}\right.\right) = \mu_k$ and variance $Var\left(I_k \left|\mathbf{Z}\right.\right) = c_k\nu\left(\mu_k\right)$ for some positive constant c_k and variance function $\nu\left(\cdot\right)$. The link function is typically expressed as $\mu_k = g\left(\mathbf{x}_i'\boldsymbol{\beta} + \boldsymbol{\sigma}'\mathbf{Z}\right)$. The term $\mathbf{x}_i'\boldsymbol{\beta}$ usually refers to the "fixed effects" while the term $\boldsymbol{\sigma}'\mathbf{Z}$, the "random effects". Such models are easily implementable in SAS. For further discussion of this type of model specification, we refer to the book by McCulloch and Searle (2001).

4 Results of empirical investigation

This section discusses the results of our empirical investigation. First, in subsection 3.1, we describe the data used in the investigation. To simplify the model, we assume exchangeability and use a univariate mixing variable and in subsection 3.2, we describe the three different mixture models used for comparison. These are the logit-Normal, probit-Normal, and the Beta-Binomail mixture models. Subsection 3.3 gives the results of our model estimation while subsection 3.4 extends these estimation introducing premium as covariates.

4.1 Data description

Our data consists of policy exposure and claims experience for the 1993-2001 period from a portfolio of automobile insurance contracts drawn from a single insurance company in Singapore. The records consisted of exposure and experience at the individual registered and insured vehicle level, and also contained details about driver and automobile risk characteristics that can be used as covariates in explaining differences in claim dependencies.

Table 1 provides a summary of the frequency of claims we observe from our portfolio broken down by calendar year of policy exposure. According to this table, we have a total of 19,148 claims during the period of 1993-2001 for which there were 210,446 policies issued. The overall frequency of claims during the entire period was about 9.1%, and there appears to be slight variability of the frequency of claims by calendar year of exposure. Indeed, later when we also investigated the effect of calendar year as a covariate, we find no evidence that claim dependencies vary by calendar year of exposure.

The data that is used in our empirical investigation are disaggregated by risk i, denoting the insured vehicle in the portfolio, and over time t, denoting the respective calendar year. For each observational unit then of (i, t), the incidence of claim observed consists of I_{it} which denotes the binary or indicator variable that the insured vehicle i has been observed to claim in year t. A value of 1 indicates a claim has been made, and a value of 0 otherwise.

Table 1

Claims Frequency and Policy Exposure by Calendar Year

Year	Number of claims	Policy exposed	Claims frequency
1993	840	12,157	6.9%
1994	1,739	15,389	11.3%
1995	869	8,074	10.8%
1996	736	7,556	9.7%
1997	1,760	16,216	10.9%
1998	2,455	23,691	10.4%
1999	3,630	36,647	9.9%
2000	3,770	45,806	8.2%
2001	3,349	44,910	7.5%
Total	19,148	210,446	9.1%

In the database, we also have the active exposure e_{it}, measured in (a fraction of) a year, which gives the length of time throughout the calendar year for which the insured risk has been exposed. This can be a crucial information in the calibration, as this contributes to the variability of the incidence of claim. The various driver and vehicle characteristics are described by the vector $\mathbf{x}_{it}$ and will serve as explanatory variables, or covariates, in our analysis. Covariates help to distinguish the additional non-homogeneity that may be present in the portfolio, and these characteristics are best described in a later section.

For notational convenience, we write the observable data consisting of the available information as follows:

$$\{I_{it}, e_{it}, \mathbf{x}_{it},\ t = 1, ..., T_i,\ i = 1, 2, ..., m\}.$$

In other words, during the observation period, we have m insured registered vehicles for which each vehicle is observed T_i (yearly) times. The maximum value of T_i is 9 calendar years because the data consists of annual portfolios for the period from 1993 up until 2001, inclusive. As indicated earlier, there are a total of 210,446 recorded observations in the files. These claims and exposures were appropriately drawn just from a portfolio of a single insurer in our larger database.

4.2 Selection of mixture models

There were three mixture models considered in our model estimation. Note that we also examined the Gamma mixing distribution, but during the fitting, convergence in the estimation process failed, giving an indication of a poor fit, and therefore this has been discarded in our comparison.

Model 1: *Logit-Normal mixing distribution*
In this case, we choose the form

$$Q(Z) = 1/\left(1 + \exp\left(-\mu - \sigma Z\right)\right) \tag{12}$$

where Z is again standard Normal random variable, and μ and σ are parameters with $\sigma > 0$. In statistics, this choice of a link function is often referred to as a logit-Normal type model.

Model 2: *Probit-Normal mixing distribution*
In this case, we choose

$$Q(Z) = \Phi\left(\mu + \sigma Z\right) \tag{13}$$

where Z is standard Normal random variable, and μ and σ are parameters with $\sigma > 0$. Here Φ denotes the distribution function of a standard Normal. In statistics, this choice of a link function is often referred to as a probit-Normal type model.

Model 3: *Beta mixing variable*
In this case, we choose Q to be Beta(a, b) distribution. Assume its density function has the form $f_Z(z) = z^{a-1}(1-z)^{b-1}/\beta(a,b)$ where $\beta(a,b)$ is the complete Beta function which can be expressed in terms of the gamma function as $\beta(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$. The appendix provides calculation details resulting from a Beta-Binomial model.

In the estimation of the parameters in the three different models described above, one has the choice of either a method-of-moment or the maximum likelihood estimation. We have chosen the maximum likelihood estimation procedure. Indeed, as demonstrated by McNeil, Frey and Embrechts (2005), the maximum likelihood estimation procedure outperforms the method-of-moment estimation procedure. These authors used simulation to make this demonstration.

In implementing maximum likelihood procedures, we suppose that we have observed an n-dimensional random vector of

$$\mathbf{i} = (i_1, i_2, ..., i_n)$$

consisting of either 1's (when a claim occurs) or 0's (when a claim does not occur). The likelihood function associated with this observed value can then be written as

$$\begin{aligned} L(\theta; \mathbf{i}) &= p_{\mathbf{I}}(i_1, ..., i_n) = P(I_1 = i_1, ..., I_n = i_n) \\ &= \int \left[\prod_{k=1}^{n} Q(Z)^{i_k} (1 - Q(Z))^{1-i_k} \right] dF_Z(z) \qquad (14) \end{aligned}$$

Here the parameter vector θ consists of the parameters of the mixing distribution.

In the case of the logit-Normal mixture model, the choice of the Q function is given in (12) and the parameter vector consists of $\theta = (\mu, \sigma)$. Similarly for the Probit-Normal mixture model, the choice of the Q function has the expression given in (13) and the parameter vector consists of $\theta = (\mu, \sigma)$. In contrast, for the Beta-Binomial mixture model, the parameter vector is $\theta = (a, b)$, or if parameterization in (18) is used instead, the vector is $\theta = (q, \gamma)$.

4.3 Estimation results

Results of the maximum likelihood estimation for the three different models (Logit-Normal, Probit-Normal, and the Beta-Binomial) are presented in Table 2. The estimation has been done using procedures NLMIXED and IML in SAS, and because the likelihood function requires integration as indicated in the likelihood function (14), the estimation routine takes approximately half-hour to run using all 210,446 records in the dataset. This dataset contains a total of 91,728 different registered vehicles over the nine-year period under investigation. In the estimation routines, we have restricted the σ in both the Logit-Normal and the Probit-Normal mixture models to be strictly positive. Similarly for the Beta-Binomial model, the parameters have been restricted to $a > 0$ and $b > 0$.

When the mixing variable is based on the Normal distribution, the parameters estimates for μ are -2.418 and -1.398, and are 0.696 and 0.342 for σ, respectively, using the logit and probit functions as link functions. In examining whether the presence of the mixing variable is significant, one can test whether the coefficient σ of Z is significantly different from zero. One can do a rough hypothesis test by examining the resulting standard errors of these estimates. In the Logit-Normal mixture model, for example, the estimate for σ is approximately 32 standard errors above 0. In a similar fashion, the Probit-Normal mixture model produces an estimate of σ that is approximately 28 standard errors above 0. Both models indicate the strong presence of the mixing variable.

In contrast, the parameter estimates for the Beta-Binomial are 13.6 and 132.0 respectively for the parameters a and b. Both estimates produced standard errors that indicate significant differences from zero.

Table 2

Summary of Maximum Likelihood Estimates

Mixture model	parms	estimates	standard errors	Neg log likelihood $-\log L(\theta; \mathbf{k})$	AIC	BIC
Logit-Normal	μ	-2.418	0.014 *			
	σ	0.696	0.022 *	65,195.1	130,394	130,391
Probit-Normal	μ	-1.398	0.007 *			
	σ	0.342	0.012 *	65,154.8	130,314	130,310
Beta-Binomial	a	13.6	0.125 *			
	b	132.0	3.205 *	65,292.4	130,589	130,586

* indicates significant at the 5% level.

All three models indicate significant presence of a mixing variable, and in comparing the performance of these three different models, and because they are non-nested, a measure such as the Aitken Information Criterion (AIC) and/or the Bayesian Information Criterion (BIC) might be suitable. See, for example, the appendix of Frees (2004) for a short discussion about these information criteria. The AIC (or BIC) is a function of the negative of the log-likelihood at the optimum parameter estimates and the number of model parameters to estimate. In all three models, we have each two different parameters to estimate, and hence it suffices to compare the negative of the log-likelihood maximized value which for all three models are reported also in Table 2. The smallest of the negative log-likelihood maximized value provides an indication of being the “best” model, and so in this case, the Probit-Normal mixture model appears to give the most suitable model in this case. However, its negative log-likelihood value is not very much different from that of the Logit-Normal mixture model, giving an indication that the Logit-Normal mixture model is not far from being an inadequate model for the dependencies of claims.

In order to understand the presence of dependencies in the incidence of claims, we focus on two measures of degree of dependence. The first one is the well-known measure of pair-wise correlation defined by

$$\rho = \frac{\mathrm{E}_Z\left(Q\left(Z\right)^2\right) - \left[\mathrm{E}_Z\left(Q\left(Z\right)\right)\right]^2}{\mathrm{E}_Z\left(Q\left(Z\right)\right)\mathrm{E}_Z\left(1 - Q\left(Z\right)\right)}. \tag{15}$$

The other is the relative risk measure defined by

$$\delta = \frac{\mathrm{E}_Z\left(Q\left(Z\right)^2\right)}{\mathrm{E}_Z\left(Q\left(Z\right)\right) - \mathrm{E}_Z\left(Q\left(Z\right)^2\right)} \tag{16}$$

which provides a measure of the degree to which one insurance risk induces another insurance risk to go on claim.

Calculation of the asymptotic variances of the maximum likelihood estimates of these dependence measures are explained in the appendix.

All three models roughly give estimates of the probability of claim of approximately 9.3%, and this value appears to be reasonably in line with our empirically observed value of approximately 9.1%, as given in Table 1. The estimates for the dependence measures given in Table 3 are self-explanatory. However, if we focus on the relative risk measure, the estimated relative risk under the Logit-Normal mixture model is 14.46%. This means that, roughly, odds are 15 to 100 that one vehicle getting on claim will induce another vehicle to go on claim. This rough estimate is not very much different for the Probit-Normal model. For the Beta-Binomial model, the relative risk measure is about 11%, which is closer to the odds ratio when there is pair-wise independence (not surprisingly so because the correlation is closer to 0 in this case).

Table 3

Estimates of Various Degrees of Claim Dependencies

Mixture model	dependence measures	estimates	standard errors
Logit-Normal	q	0.0931	0.0007 *
	δ	0.1446	0.0032 *
	ρ	0.0367	0.0025 *
Probit-Normal	q	0.0934	0.0007 *
	δ	0.1477	0.0034 *
	ρ	0.0390	0.0027 *
Beta-Binomial	q	0.0936	0.0030 *
	δ	0.1108	0.0037 *
	ρ	0.0068	0.0002 *

* indicates significant at the 5% level.

For completeness purpose, we also provide in Table 4 a summary of the estimated covariances of the parameter estimates, for all three models. These covariance estimates were used to compute the standard errors in Table 3 associated with the probability of claim, relative risk, and correlation measures. These standard errors were computed using the formulas for the asymptotic variances that were discussed in the appendix.

Table 4

Estimated Covariances of Parameter Estimates

Mixture model	parameters	covariance estimates
Logit-Normal	$\begin{pmatrix} \mu \\ \sigma \end{pmatrix}$	$\begin{pmatrix} 0.0001895 & -0.000247 \\ -0.000247 & 0.0004888 \end{pmatrix}$
Probit-Normal	$\begin{pmatrix} \mu \\ \sigma \end{pmatrix}$	$\begin{pmatrix} 0.000044 & -0.00006 \\ -0.00006 & 0.000145 \end{pmatrix}$
Beta-Binomial	$\begin{pmatrix} a \\ b \end{pmatrix}$	$\begin{pmatrix} 0.01565 & 1.69871 \\ 1.69871 & 10.26959 \end{pmatrix}$

4.4 Premium as a covariate

We also considered mixture models specified with covariate or explanatory variables. Covariates available in our dataset included driver characteristics such as age and sex of driver as well as vehicle characteristics that include types and models of vehicles registered and insured. Information about insurance coverage such as coverage type (e.g. comprehensive), premiums as well as NCD (No Claims Discount ranging from 0% to 50%) are also available in our dataset. We injected these covariates into our mixture models through a process of elimination. In the end, we find that only the premium information provided significant explanatory variable, not surprisingly so, because the premium provides a measure of the riskiness of the insured according to the actuarial valuation. We also investigated the effect of introducing calendar year as a covariate; there were no significant differences in the claim dependencies across calendar years.

Now in reporting the results in this documentation, we focus on the Logit-Normal and the Probit-Normal mixture models. The premium is introduced as a covariate through the μ parameter by specifying that

$$\mu = \beta_0 + \beta_1 \log(\text{Premium}/1000), \tag{17}$$

where β_0 provides the intercept while the β_1 is the slope of the $\log(\text{Premium}/1000)$. The comparison of the maximum likelihood estimates for the Logit-Normal and the Probit-Normal with this regression specification is outlined in Table 5.

According to Table 5, under the Logit-Normal model, the intercept estimate is -1.804 with a standard error of 0.014, and the estimate of the coefficient of the $\log(\text{Premium}/1000)$ is 0.528 with a standard error of 0.009. Both sets of estimates and standard errors provide indication of the significance of regression equation. The σ parameter is 0.521 which is 22 standard errors above zero.

On the other hand, under the Probit-Normal model, the intercept estimate is -1.080 with a standard error of 0.007. The estimate of the coefficient of the $\log(\text{Premium}/1000)$ is 0.261 with a standard error of 0.004. These small standard errors provide a similar indication of the significance of the regression equation. Furthermore, the σ parameter is 0.289 which is also approximately 22 standard errors above zero.

Table 5

Maximum Likelihood Estimates of Normal Mixtures with Covariates

Mixture model	parms	estimates	standard errors	Neg log likelihood $-\log L(\theta; \mathbf{k})$	AIC	BIC
Logit-Normal	β_0	-1.804	0.014 *			
	β_1	0.528	0.009 *			
	σ	0.521	0.024 *	63,043.6	126,093	126,088
Probit-Normal	β_0	-1.080	0.007 *			
	β_1	0.261	0.004 *			
	σ	0.289	0.013 *	63,074.1	126,154	126,149

* indicates significant at the 5% level.

To visualize the effects of the level of premium on the probability of claim, the relative risk, and the correlation, we display Figures 1 to 3 which also provide for a comparison of the resulting differences on the effects between the Logit-Normal and the Probit-Normal models. According to all three figures, increasing the level of the premium has the effect of increasing the chances of claims, the relative risk, as well as the correlation.

5 Concluding Remarks

In this article, we summarize the results of our empirical investigation of estimating the presence of claim dependencies in a portfolio of insurance contracts. We have also provided a general description about the flexibility of the mixture model as a way to model the dependencies. In calibrating these mixture models, we have extracted a portfolio from a randomly selected company that has been drawn from a large data base consisting of portfolios of motor (or automobile) insurance policies provided to us by the General Insurance Association of Singapore. This association, which consists of membership of general insurers in Singapore, has similar function to the Insurance Services office (I.S.O.) of the United States.

In the estimation of the mixture model parameters, we have employed maximum likelihood method and exploited the capabilities of the NLMIXED and IML procedures in SAS.

Anyone interested can e-mail the author for a copy of the SAS codes used.

Our investigation provides a clear indication of the presence of claim dependencies. On a correlation basis, the dependence estimate is around the neighborhood of 4%. But correlation is not the best way to provide for the understanding of the dependencies in claims. Rather, as a good proxy, we use the relative risk measure which provides the degree to which one insurance risk induces another insurance risk to go on claim. Our estimates for the relative risk in our portfolio are in the neighborhood of 14%.

We also find that of all the covariates that were investigated, the premium has the most influence on the level of claim dependencies. We chose to focus on the Logit-Normal and the Probit-Normal to demonstrate how to inject covariate information into the mixture models. This is because both of these models outperform the Beta-Binomial model. However, the Beta-Binomial is also flexible enough to accommodate inclusion of covariates. See for example Prentice (1986, 1988) on how covariates can enter into the Beta-Binomial model.

There are several advantages to using the mixture framework. We list here some of the advantages and flexibility provided by the use of mixture models in estimating for the presence of dependencies:

- Mixture models reduce the dimensionality of the problem. Instead of specifying a multivariate distribution or a copula corresponding to a random vector of dimension equal to the total number of insurance risk exposed, it allows for a reduction in dimension in the sense that only the distribution of the unobservable has to be estimated. Parametric model construction for the unobservable is typical, and there are standard models usually suggested in the literature. Here, we considered the Normal random variable and the Beta variable as potential mixing variables.
- Because the likelihood function requires integrating out the effect of the mixing variable, estimation routines can become cumbersome for mixture models. However, as demonstrated in this paper, SAS has procedure called NLMIXED that allow for this and that can accommodate covariates as well. However, at the moment, the only mixing variable that can be used in NLMIXED is a Normal random variable. As alternative, we also have the Beta-Binomial mixture model which is flexible enough because one

can get explicit forms after integrating out the Beta mixing variable.

- A third advantage of the mixture model is some possible mathematical tractability. We can therefore estimate correlations that describe the dependencies of claim occurrences. We did mention that correlations are not the best way to understand the dependencies of Bernoulli events. Alternatively, we propose to use the relative risk concept which describes how one insurance claim occurrence induces another claim occurrence.
- Finally, the unobservable variable in the Bernoulli mixture model has the advantage of providing a natural interpretation to the resulting model. It also can be interpreted as a way to model the presence of heterogeneity in the insurance risk portfolio.

Appendix A

The Beta-Binomial model

Using the Beta as a mixing random variable, this leads us to the so-called Beta-Binomial model for the random sum

$$S = \sum\nolimits_{k=1}^{n} I_k$$

which refers to the total number of claims. To find the distribution of S, first notice that

$$\begin{aligned} q_k &= \frac{1}{\beta(a,b)} \int_0^1 z z^{a-1} (1-z)^{b-1} \, dz \\ &= \frac{\beta(a+1,b)}{\beta(a,b)} \int_0^1 \frac{z^{a+1-1}(1-z)^{b-1}}{\beta(a+1,b)} dz \\ &= \frac{\beta(a+1,b)}{\beta(a,b)} = \frac{a}{a+b}. \end{aligned}$$

By conditional independence assumption, we know that conditional on $Q = z$, the total number of claims S has then a Binomial distribution with n as total number of trials and probability of success z. This conditional probability then is equal to

$$P(S = s \,|Q = z) = \binom{n}{s} z^s (1-z)^{n-s} \text{ for } s = 0, 1, ..., n.$$

The unconditional probability is then obtained by integrating overall the range of values of z, and thus we have

$$\begin{aligned} P(S = s) &= \int_0^1 P(S = s \,|Q = z) \, dF_Z(z) \\ &= \binom{n}{s} \frac{1}{\beta(a,b)} \int_0^1 z^s (1-z)^{n-s} z^{a-1} (1-z)^{b-1} \, dz \\ &= \binom{n}{s} \frac{\beta(s+a, n-s+b)}{\beta(a,b)} \int_0^1 \frac{z^{s+a-1}(1-z)^{n-s+b-1}}{\beta(s+a, n-s+b)} dz \\ &= \binom{n}{s} \frac{\beta(s+a, n-s+b)}{\beta(a,b)}. \end{aligned}$$

Indeed, using the relationship between the beta and gamma functions, we can write

$$\begin{aligned} \frac{\beta(s+a, n-s+b)}{\beta(a,b)} &= \frac{\Gamma(a+s)}{\Gamma(a)} \frac{\Gamma(b+n-s)}{\Gamma(b)} \Bigg/ \frac{\Gamma(a+b+n)}{\Gamma(a+b)} \\ &= \frac{\prod_{j=0}^{s-1}(a+j) \prod_{j=0}^{n=s-1}(b+j)}{\prod_{j=0}^{n-1}(a+b+j)} \end{aligned}$$

and dividing both numerator and denominator by $(a+b)$, we have

$$P(S=s)=\binom{n}{s}\frac{\beta(s+a,n-s+b)}{\beta(a,b)}=\binom{n}{s}\frac{\prod_{j=0}^{s-1}(q+\gamma j)\prod_{j=0}^{n-s-1}(p+\gamma j)}{\prod_{j=0}^{n-1}(1+\gamma j)}, \tag{18}$$

where $q=a/(a+b)$, $p=1-q=b/(a+b)$, and $\gamma=1/(a+b)$. This parameterization is exactly the form that has been considered in Prentice (1986, 1988). This parameterization can also be readily introduced with covariates.

Asymptotic variances

In this appendix, we derive expressions for the asymptotic variances of the correlation and relative risk measure estimates (based on maximum likelihood). All the mixture models we consider in this paper have two parameters, and to generalize the situation, denote this vector of two parameters by $\theta=(\theta_1,\theta_2)'$. Then, the correlation and relative risk measures given in (15) and (16) are therefore thought of as functions of θ. We therefore conveniently denote them by $\rho(\theta)$ and $\delta(\theta)$, respectively. If $\widehat{\theta}$ then denotes the maximum likelihood estimates, then the asymptotic variances are respectively given by

$$\left(\frac{\partial\rho}{\partial\theta}\right)' Var\left(\widehat{\theta}\right)\left(\frac{\partial\rho}{\partial\theta}\right)$$

where $\partial\rho/\partial\theta=(\partial\rho/\partial\theta_1,\partial\rho/\partial\theta_2)'$, and

$$\left(\frac{\partial\delta}{\partial\theta}\right)' Var\left(\widehat{\theta}\right)\left(\frac{\partial\delta}{\partial\theta}\right)$$

where $\partial\delta/\partial\theta=(\partial\delta/\partial\theta_1,\partial\delta/\partial\theta_2)'$.

For the correlation measure, by letting the quantities

$$\rho_n=\mathrm{E}_Z\left(Q(Z)^2\right)-\left[\mathrm{E}_Z(Q(Z))\right]^2$$

and

$$\rho_d=\mathrm{E}_Z(Q(Z))-\mathrm{E}_Z\left(Q(Z)^2\right)$$

which corresponds to the numerator and denominator of the expression, we can write the partial derivatives as:

$$\frac{\partial \rho_n}{\partial \theta_j} = 2 \times \left[\mathrm{E}_Z \left(Q\left(Z\right) \frac{\partial Q\left(Z\right)}{\partial \theta_j} \right) - \mathrm{E}_Z \left(Q\left(Z\right)\right) \mathrm{E}_Z \left(\frac{\partial Q\left(Z\right)}{\partial \theta_j} \right) \right]$$

and

$$\frac{\partial \rho_d}{\partial \theta_j} = \mathrm{E}_Z \left(\frac{\partial Q\left(Z\right)}{\partial \theta_j} \right) - 2 \times \mathrm{E}_Z \left(Q\left(Z\right)\right) \mathrm{E}_Z \left(\frac{\partial Q\left(Z\right)}{\partial \theta_j} \right)$$

for $j = 1, 2$. Thus, we have

$$\frac{\partial \rho}{\partial \theta_j} = \left(\rho_d \frac{\partial \rho_n}{\partial \theta_j} - \rho_n \frac{\partial \rho_d}{\partial \theta_j} \right) \Big/ \rho_d^2 \tag{19}$$

for $j = 1, 2$.

For the relative risk measure, first we write it as

$$\delta = \frac{\mathrm{E}_Z \left(Q\left(Z\right)^2 \right)}{\mathrm{E}_Z \left(Q\left(Z\right)\right) - \mathrm{E}_Z \left(Q\left(Z\right)^2 \right)} = \left(\frac{\mathrm{E}_Z \left(Q\left(Z\right)\right)}{\mathrm{E}_Z \left(Q\left(Z\right)^2 \right)} - 1 \right)^{-1}.$$

Then we can compute the partial derivatives as

$$\begin{aligned} \frac{\partial \delta}{\partial \theta_j} &= -\left(\frac{\mathrm{E}_Z \left(Q\left(Z\right)\right)}{\mathrm{E}_Z \left(Q\left(Z\right)^2 \right)} - 1 \right)^{-2} \times \\ &\quad \left(\mathrm{E}_Z \left(Q\left(Z\right)^2 \right) \mathrm{E}_Z \left(\frac{\partial Q\left(Z\right)}{\partial \theta_j} \right) - 2\mathrm{E}_Z \left(Q\left(Z\right)\right) \mathrm{E}_Z \left(Q\left(Z\right) \frac{\partial Q\left(Z\right)}{\partial \theta_j} \right) \right) \Big/ \left[\mathrm{E}_Z \left(Q\left(Z\right)^2 \right) \right]^2 \end{aligned}$$

which further simplifies to

$$\frac{\partial \delta}{\partial \theta_j} = \left(\frac{\delta}{\mathrm{E}_Z \left(Q\left(Z\right)^2 \right)} \right)^2 \times \left(2\mathrm{E}_Z \left(Q\left(Z\right)\right) \mathrm{E}_Z \left(Q\left(Z\right) \frac{\partial Q\left(Z\right)}{\partial \theta_j} \right) - \mathrm{E}_Z \left(Q\left(Z\right)^2 \right) \mathrm{E}_Z \left(\frac{\partial Q\left(Z\right)}{\partial \theta_j} \right) \right). \tag{20}$$

Now to specialize the situation in the case of the Logit-Normal mixture model, we can write

$$\frac{\partial Q\left(Z\right)}{\partial \mu} = \frac{e^{-\mu-\sigma Z}}{\left(1 + e^{-\mu-\sigma Z}\right)^2} = Q\left(Z\right)\left(1 - Q\left(Z\right)\right) \tag{21}$$

and

$$\frac{\partial Q\left(Z\right)}{\partial \sigma} = \frac{Z e^{-\mu-\sigma Z}}{\left(1 + e^{-\mu-\sigma Z}\right)^2} = ZQ\left(Z\right)\left(1 - Q\left(Z\right)\right) \tag{22}$$

where we have replaced (θ_1, θ_2) with the parameter vector (μ, σ).

In the special case of the Probit-Normal mixture model, we write

$$\frac{\partial Q\left(Z\right)}{\partial \mu} = \phi\left(\mu + \sigma Z\right) \tag{23}$$

and

$$\frac{\partial Q\left(Z\right)}{\partial \sigma} = Z\phi\left(\mu + \sigma Z\right) \tag{24}$$

where $\phi\left(\cdot\right)$ denotes the density function of a standard Normal distribution.

Now finally, in the special case of the Beta-Binomial mixture model, we have the parameter vector $\theta = (a, b)'$ and the explicit formulas

$$q = a/\left(a + b\right)$$

for the probability of a claim,

$$\delta = \left(a + 1\right)/b$$

for the relative risk measure, and

$$\rho = \left(a + b + 1\right)^{-1}$$

for the correlation. It is straightforward to show that the partial derivatives are respectively

$$\left(\frac{\partial q}{\partial \theta}\right)' = \left(b\left(a + b\right)^{-2}, -a\left(a + b\right)^{-2}\right),$$

$$\left(\frac{\partial \delta}{\partial \theta}\right)' = \left(b^{-1}, -\left(a + 1\right)b^{-2}\right),$$

and

$$\left(\frac{\partial \rho}{\partial \theta}\right)' = \left(-\left(a + b + 1\right)^{-2}, -\left(a + b + 1\right)^{-2}\right).$$

References

[1] Albrecher, H., Kantor, J. (2002) "Simulation of Ruin Probabilities for Risk Processes of Markovian Type," *Monte Carlo Methods and Applications* 8: 111-127.

[2] Cossette, H., Gaillardetz, P., Marceau, E., Rioux, J. (2002) "On Two Dependent Individual Risk Models," *Insurance: Mathematics & Economics* 30: 153-166.

[3] Credit-Suisse First Boston (1997) *CreditRisk$^+$: A Credit Risk Management Framework*, Technical Document, avaliable at <http://www.csfb.com/creditrisk>

[4] Denuit, M., Lefevre, C., Utev, S. (2002) "Measuring the impact of dependence between claims occurrences," *Insurance: Mathematics & Economics* 30: 1-19.

[5] Dhaene, J., Goovaerts, M. (1997) "On the Dependency of Risks in the Individual Life Model," *Insurance: Mathematics & Economics* 19: 243-253.

[6] Dhaene, J., Denuit, M., Goovaerts, M., Kaas, R., Vyncke, D. (2002a) "The Concept of Comonotonicity in Actuarial Science and Finance: Theory," *Insurance: Mathematics & Economics* 31: 3-33.

[7] Dhaene, J., Denuit, M., Goovaerts, M., Kaas, R., Vyncke, D. (2002b) "The Concept of Comonotonicity in Actuarial Science and Finance: Applications," *Insurance: Mathematics & Economics* 31: 133-161.

[8] Embrechts, P., Lindskog, F., McNeil, A. (2001) "Modelling Dependence with Copulas and Applications to Risk Management," working paper, ETHZ.

[9] Frey, R., McNeil, A.J. (2003) "Dependent Defaults in Models of Portfolio Credit Risk," *Journal of Risk* 6: 59-92.

[10] Frees, E.W. (Jed) (2004). *Longitudinal and Panel Data: Analysis and Applications in the Social Sciences.* Cambridge University Press, Cambridage, UK.

[11] Frees, E.W., Valdez, E.A. (1998) "Understanding Relationships using Copulas," *North American Actuarial Journal* 2: 1-25.

[12] Frees, E.W., Wang, P. (2005) "Credibility using Copulas," *North American Actuarial Journal* 9: 31-48.

[13] Genest, C., Marceau, E., Mesfioui, M. (2003) "Compound Poisson Approximations for Individual Models with Dependent Risks," *Insurance: Mathematics & Economics* 32: 73-91.

[14] Gordy, M. (2000) "A Comparative Anatomy of Credit Risk Models," *Journal of Banking and Finance* 24: 119-149.

[15] Heilmann, W.R. (1986) "On the Impact of Independence of Risks on Stop-Loss Transforms," *Insurance: Mathematics & Economics* 5: 197-199.

[16] Hu, T. Wu, Z. (1999) "On Dependence of Risks and Stop-Loss Premiums, " *Insurance: Mathematics & Economics* 24: 323-332.

[17] Hurlimann, W. (1993) “Bivariate Distributions with Atomic Conditionals and Stop-Loss Transforms of Random Sums,” *Statistics & Probability Letters* 17: 329-335.

[18] Joe, H. (1997), *Multivariate Models and Dependence Concepts* London: Chapman & Hall.

[19] Koyluoglu, U., Hickman, A. (1998) “Reconciling the differences,” *Risk* 11: 56-62.

[20] Li, D. (2001) “On Default Correlation: A Copula Function Approach,” *Journal of Fixed Income* 9: 43-54.

[21] Lindskog, F., McNeil, A. (2003) “Common Poisson Shock Models: Applications to Insurance and Credit Risk Modelling,” *ASTIN Bulletin* 33: 209-238.

[22] Marceau, E., Cossette, H., Gaillardetz, P., Rioux, J. (1999) “Dependence in the Individual Risk Model ,” working paper, Laval University.

[23] Marshall, A.W., Olkin, I. (1967) “A Multivariate Exponential Distribution,” *Journal of the American Statistical Association* 62: 30-44.

[24] McCulloch, C.E., Searle, S.R. (2001) *Generalized Linear and Mixed Models* New York: Wiley.

[25] McNeil, A.J., Frey,R., Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press, Princeton.

[26] Muller, A. (1997) “Stop-Loss Order for Portfolios of Dependent Risks,” *Insurance: Mathematics & Economics* 21: 219-223.

[27] Nelsen, R. (1999) *An Introduction to Copulas* New York: Springer.

[28] Prentice, R.L. (1986) “Binary Regression Using an Extended Beta-Binomial Distribution, with discussion of Correlation Induced by Covariate Measurement Errors,” *Journal of the American Statistical Association* **81**: 321-327.

[29] Prentice, R.L. (1988) “Correlated Binary Regression with Covariates Specific to Each Binary Observation,” *Biometrics* **44**: 1033-1048.

[30] Purcaru, O., Denuit, M. (2002) “On the Dependence Induced by Frequency Credibility Models” *Belgian Actuarial Bulletin* 2: 73-79.

[31] Purcaru, O., Denuit, M. (2003). “Dependence in Dynamic Claim Frequency Credibility Models” *ASTIN Bulletin* 33: 23-40.

[32] Valdez, E.A., Mo, K. (2002) “Ruin Probabilities with Dependent Claims,” UNSW working paper.

[33] Wang, S. (1998) “Aggregation of Correlated Risk Portfolios: Models and Algorithms,” *Proceedings of the Casualty Actuarial Society 85: 848-939.*

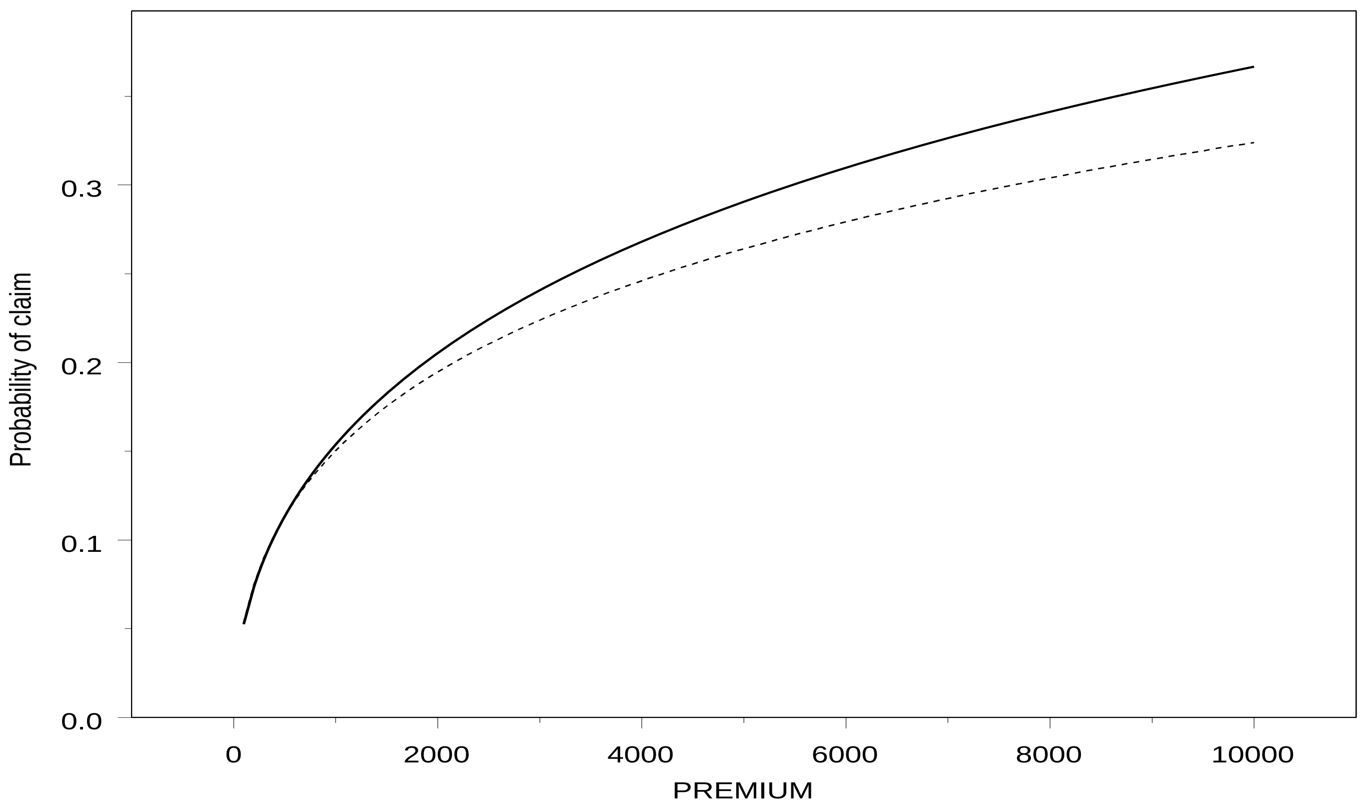

Figure 1: Graphical display of probability of claim as a function of Premium. The smooth curve refers to the Logit-Normal Model while the broken curve refers to the Probit-Normal Model.

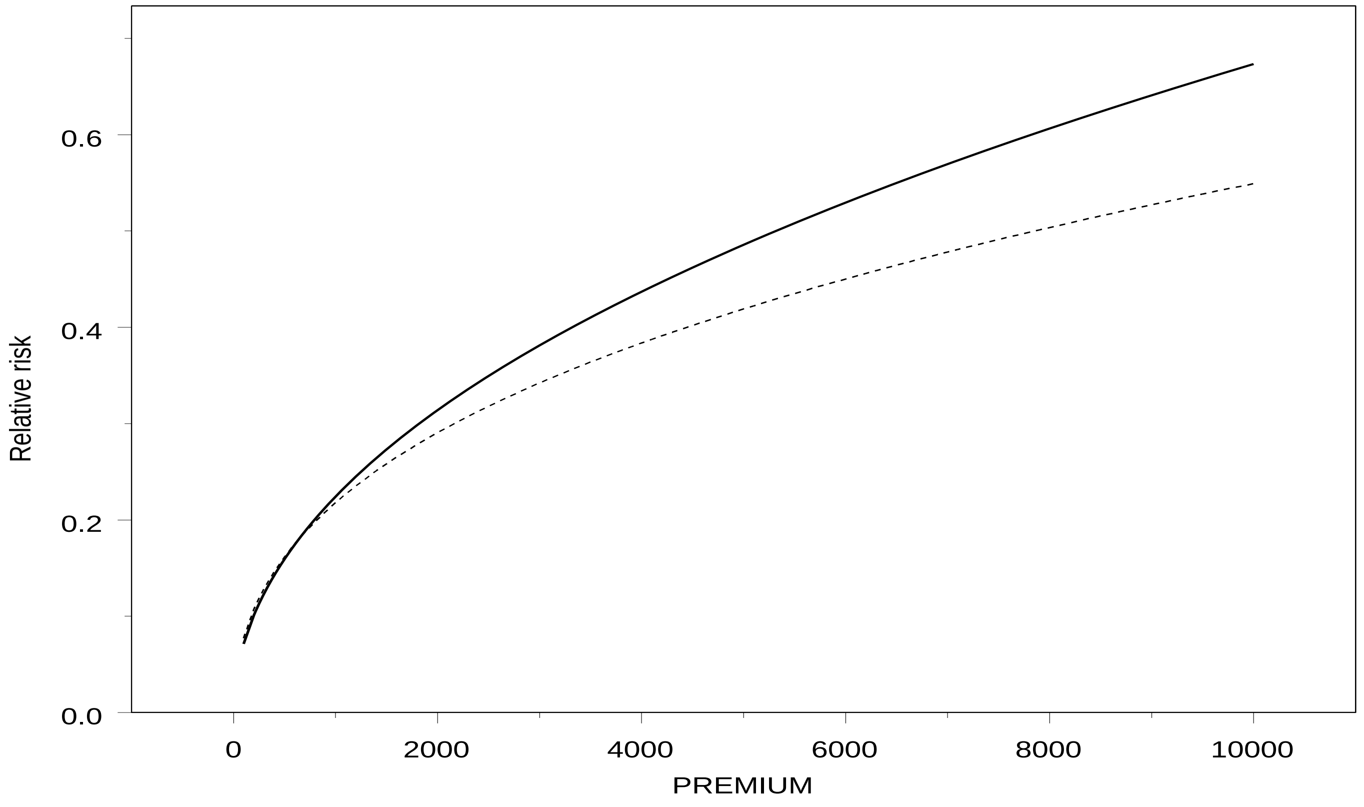

Figure 2: Graphical display of relative risk as a function of Premium. The smooth curve refers to the Logit-Normal Model while the broken curve refers to the Probit-Normal Model.

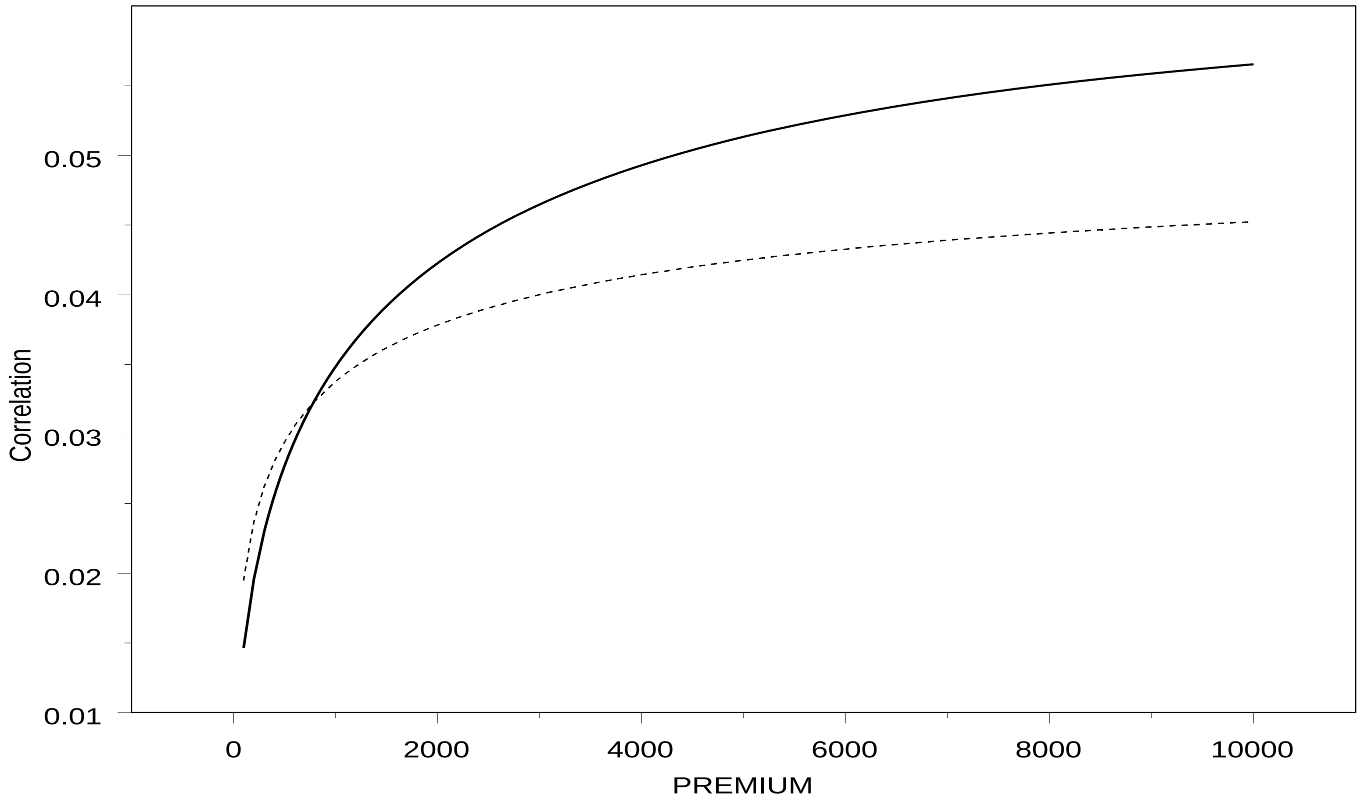

Figure 3: Graphical display of correlation as a function of Premium. The smooth curve refers to the Logit-Normal Model while the broken curve refers to the Probit-Normal Model.