

Article

# Multivariate Frequency-Severity Regression Models in Insurance

Edward W. Frees <sup>1,\*</sup>, Gee Lee <sup>1</sup> and Lu Yang <sup>2</sup>

<sup>1</sup> School of Business, University of Wisconsin-Madison, 975 University Avenue, Madison, WI 53706, USA; gee.lee@wisc.edu

<sup>2</sup> Department of Statistics, University of Wisconsin-Madison, 1300 University Avenue, Madison, WI 53706, USA; luyang@stat.wisc.edu

\* Correspondence: jfrees@bus.wisc.edu; Tel.: +1-608-262-0429

Academic Editor: Montserrat Guillén

Received: 16 November 2015; Accepted: 15 February 2016; Published: 25 February 2016

**Abstract:** In insurance and related industries including healthcare, it is common to have several outcome measures that the analyst wishes to understand using explanatory variables. For example, in automobile insurance, an accident may result in payments for damage to one's own vehicle, damage to another party's vehicle, or personal injury. It is also common to be interested in the frequency of accidents in addition to the severity of the claim amounts. This paper synthesizes and extends the literature on multivariate frequency-severity regression modeling with a focus on insurance industry applications. Regression models for understanding the distribution of each outcome continue to be developed yet there now exists a solid body of literature for the marginal outcomes. This paper contributes to this body of literature by focusing on the use of a copula for modeling the dependence among these outcomes; a major advantage of this tool is that it preserves the body of work established for marginal models. We illustrate this approach using data from the Wisconsin Local Government Property Insurance Fund. This fund offers insurance protection for (i) property; (ii) motor vehicle; and (iii) contractors' equipment claims. In addition to several claim types and frequency-severity components, outcomes can be further categorized by time and space, requiring complex dependency modeling. We find significant dependencies for these data; specifically, we find that dependencies among lines are stronger than the dependencies between the frequency and average severity within each line.

**Keywords:** tweedie distribution; copula regression; government insurance; dependency modeling; inflated count model

## 1. Introduction and Motivation

Many insurance data sets feature information about how often claims arise, the frequency, in addition to the claim size, the severity. Observable responses can include:

- $N$ , the number of claims (events),
- $y_k$ ,  $k = 1, \dots, N$ , the amount of each claim (loss), and
- $S = y_1 + \dots + y_N$ , the aggregate claim amount.

By convention, the set  $\{y_j\}$  is empty when  $N = 0$ .

**Importance of Modeling Frequency.** The aggregate claim amount  $S$  is the key element for an insurer's balance sheet, as it represents the amount of money paid on claims. So, why do insurance companies regularly track the frequency of claims as well as the claim amounts? As in an earlier review [1], we can segment these reasons into four categories: (i) features of contracts; (ii) policyholder behavior and risk mitigation; (iii) databases that insurers maintain; and (iv) regulatory requirements.

1. Contractually, it is common for insurers to impose deductibles and policy limits on a per occurrence and on a per contract basis. Knowing only the aggregate claim amount for each policy limits any insights one can get into the impact of these contract features.
2. Covariates that help explain insurance outcomes can differ dramatically between frequency and severity. For example, in healthcare, the decision to utilize healthcare by individuals (the frequency) is related primarily to personal characteristics whereas the cost per user (the severity) may be more related to characteristics of the healthcare provider (such as the physician). Covariates may also be used to represent risk mitigation activities whose impact varies by frequency and severity. For example, in fire insurance, lightning rods help to prevent an accident (frequency) whereas fire extinguishers help to reduce the impact of damage (severity).
3. Many insurers keep data files that suggest developing separate frequency and severity models. For example, insurers maintain a “policyholder” file that is established when a policy is written. A separate file, often known as the “claims” file, records details of the claim against the insurer, including the amount. These separate databases facilitate separate modeling of frequency and severity.
4. Insurance is a closely monitored industry sector. Regulators routinely require the reporting of claims numbers as well as amounts. Moreover, insurers often utilize different administrative systems for handling small, frequently occurring, reimbursable losses, e.g., prescription drugs, versus rare occurrence, high impact events, e.g., inland marine. Every insurance claim means that the insurer incurs additional expenses suggesting that claims frequency is an important determinant of expenses.

**Importance of Including Covariates.** In this work, we assume that the interest is in the joint modeling of frequency and severity of claims. In actuarial science, there is a long history of studying frequency, severity and the aggregate claim for homogeneous portfolios; that is, identically and independently distributed realizations of random variables. See any introductory actuarial text, such as [2], for an introduction to this rich literature.

In contrast, the focus of this review is to assume that explanatory variables (covariates, predictors) are available to the analyst. Historically, this additional information has been available from a policyholder’s application form, where various characteristics of the policyholder were supplied to the insurer. For example, in motor vehicle insurance, classic rating variables include the age and sex of the driver, type of the vehicle, region in which the vehicle was driven, and so forth. The current industry trend is towards taking advantage of “big data”, with attempts being made to capture additional information about policyholders not available from traditional underwriting sources. An important example is the inclusion of personal credit scores, developed and used in the industry to assess the quality of personal loans, that turn out to also be important predictors of motor vehicle claims experience. Moreover, many insurers are now experimenting with global positioning systems combined with wireless communication to yield real-time policyholder usage data and much more. Through such systems, they gather micro data such as the time of day that the car is driven, sudden changes in acceleration, and so forth. This foray into detailed information is known as “telematics”. See, for example, [3] for further discussion.

**Importance of Multivariate Modeling.** To summarize reasons for examining insurance outcomes on a multivariate basis, we utilize an earlier review in [4]. In that paper, frequencies were restricted to binary outcomes, corresponding to a claim or no claim, known as “two-part” modeling. In contrast, this paper describes more general frequency modeling, although the motivation for examining multivariate outcomes are similar. Analysts and managers gain useful insights by studying the joint behavior of insurance risks, *i.e.*, a multivariate approach:

- For some products, insurers must track payments separately by component to meet contractual obligations. For example, in motor vehicle coverage, deductibles and limits depend on the

coverage type, e.g., bodily injury, damage to one's own vehicle, or damage to another party; is natural for the insurer to track claims by coverage type.

- For other products, there may be no contractual reasons to decompose an obligation by components and yet the insurer does so to help better understand the overall risk. For example, many insurers interested in pricing homeowners insurance are now decomposing the risk by "peril", or cause of loss. Homeowners is typically sold as an all-risk policy, which covers all causes of loss except those specifically excluded. By decomposing losses into homogenous categories of risk, actuaries seek to get a better understanding of the determinants of each component, resulting in a better overall predictor of losses.
- It is natural to follow the experience of a policyholder over time, resulting in a vector of observations for each policyholder. This special case of multivariate analysis is known as "panel data", see, for example, [5].
- In the same fashion, policy experience can be organized through other hierarchies. For example, it is common to organize experience geographically and analyze spatial relationships.
- Multivariate models in insurance need not be restricted to only insurance losses. For example, a study of term and whole life insurance ownership is in [6]. As an example in customer retention, both [7,8] advocate for putting the customer at the center of the analysis, meaning that we need to think about the several products that a customer owns simultaneously.

An insurer has a collection of multivariate risks and the interest is managing the distribution of outcomes. Typically, insurers have a collection of tools that can then be used for portfolio management including deductibles, coinsurance, policy limits, renewal underwriting, and reinsurance arrangements. Although pricing of risks can often focus on the mean, with allowances for expenses, profit, and "risk loadings", understanding capital requirements and firm solvency requires understanding of the portfolio distribution. For this purpose, it is important to treat risks as multivariate in order to get an accurate picture of their dependencies.

**Dependence and Contagion.** We have seen in the above discussion that dependencies arise naturally when modeling insurance data. As a first approximation, we typically think about risks in a portfolio as being independent from one another and rely upon risk pooling to diversify portfolio risk. However, in some cases, risks share common elements such as an epidemic in a population, a natural disaster such as a hurricane that affects many policyholders simultaneously, or an interest rate environment shared by policies with investment elements. These common (pandemic) elements, often known as "contagion", induce dependencies that can affect a portfolio's distribution significantly.

Thus, one approach is to model risks as univariate outcomes but to incorporate dependencies through unobserved "latent" risk factors that are common to risks within a portfolio. This approach is viable in some applications of interest. However, one can also incorporate contagion effects into a more general multivariate approach that we adopt this view in this paper. We will also consider situations where data are available to identify models and so we will be able to use the data to guide our decisions when formulating dependence models.

Modeling dependencies is important for many reasons. These include:

1. Dependencies may impact the statistical significance of parameter estimates.
2. When we examine the distribution of one variable conditional on another, dependencies are important.
3. For prediction, the degree of dependency affects the degree of reliability of our predictions.
4. Insurers want to construct products that do not expose them to extreme variation. They want to understand the distribution of a product that has many identifiable components; to understand the distribution of the overall product, one strategy is to describe the distribution of each product and a relationship among the distributions.

A recent review paper [3] provides additional discussion.

**Plan for the Paper.** The following is a plan to introduce readers further to the topic. Section 2 gives a brief overview of univariate models, that is, regression models with a single outcome for a response. This section sets the tone and notation for the rest of the paper. Section 3 provides an overview of multivariate modeling, focusing on the “copula” regression approach described here. This section discusses continuous, discrete, and mixed (Tweedie) outcomes. For our regression applications, the focus is mainly on a family of copulas known as “elliptical”, because of their flexibility of modeling pairwise dependence and wide usages in multivariate analysis. Section 3 also summarizes a modeling strategy for the empirical approach of copula regression.

Section 4 reviews other recent work on multivariate frequency-severity model and describes the benefits of diversification, particularly important in an insurance context. To illustrate our ideas and approach, Sections 5 and 6 provide our analysis using data from the Wisconsin Local Government Property Insurance Fund. Section 7 concludes with a few closing remarks.

## 2. Univariate Foundations

For notation, define  $N$  for the random number of claims,  $S$  for the aggregate claim amount, and  $\bar{S} = S/N$  for the average claim amount (defined to be 0 when  $N = 0$ ). To model these outcomes, we use a collection of covariates  $\mathbf{x}$ , some of which may be useful for frequency modeling whereas others will be useful for severity modeling. The dependent variables  $N$ ,  $S$ , and  $\bar{S}$  as well as covariates  $\mathbf{x}$  vary by the risk  $i = 1, \dots, n$ . For each risk, we also are interested in multivariate outcomes indexed by  $j = 1, \dots, p$ . So, for example,  $\mathbf{N}_i = (N_{i1}, \dots, N_{ip})'$  represents the vector of  $p$  claim outcomes from the  $i$ th risk.

This section summarizes modeling approaches for a single outcome ( $p = 1$ ). A more detailed review can be found in [1].

### 2.1. Frequency-Severity

For modeling the joint outcome  $(N, S)$  (or equivalently,  $(N, \bar{S})$ ), it is customary to first condition on the frequency and then modeling the severity. Suppressing the  $\{i\}$  subscript, we decompose the distribution of the dependent variables as:

$$\begin{aligned} f(N, S) &= f(N) \times f(S|N) \\ \text{joint} &= \text{frequency} \times \text{conditional severity}, \end{aligned} \quad (1)$$

where  $f(N, S)$  denotes the joint distribution of  $(N, S)$ . Through this decomposition, we do *not* require independence of the frequency and severity components.

There are many ways to model dependence when considering the joint distribution  $f(N, S)$  in Equation (1). For example, one may use a latent variable that affects both frequency  $N$  and loss amounts  $S$ , thus inducing a positive association. Copulas are another tool used regularly by actuaries to model non-linear associations and will be described in subsequent Section 4. The conditional probability framework is a natural method of allowing for potential dependencies and provides a good starting platform for empirical work.

### 2.2. Modeling Frequency Using GLMs

It has become routine for actuarial analysts to model the frequency  $N_i$  based on covariates  $\mathbf{x}_i$  using generalized linear models, GLMs, *cf.*, [9]. For binary outcomes, logit and probit forms are most commonly used, *cf.*, [10]. For count outcomes, one begins with a Poisson or negative binomial distribution. Moreover, to handle the excessive number of zeros relative to that implied by these distributions, analysts routinely examine zero-inflated models, as described in [11].

A strength of GLMs relative to other non-linear models is that one can express the mean as a simple function of linear combinations of the covariates. In insurance, it is common to use a “logarithmic link” for this function and so express the mean as  $\mu_i = E N_i = \exp(\mathbf{x}_i' \boldsymbol{\beta})$ , where  $\boldsymbol{\beta}$  is a vector of

parameters associated with the covariates. This function is used because it yields desirable parameter interpretations, seems to fit data reasonably well, and ties well with other approaches traditionally used in actuarial ratemaking applications [12].

It is also common to identify one of the covariates as an “exposure” that is used to calibrate the size of a potential outcome variable. In frequency modeling, the mean is assumed to vary proportionally with  $E_i$ , for exposure. To incorporate exposures, we specify one of the explanatory variables to be  $\ln E_i$  and restrict the corresponding regression coefficient to be 1; this term is known as an *offset*. With this convention, we have

$$\ln \mu_i = \ln E_i + \mathbf{x}'_i \boldsymbol{\beta} \Leftrightarrow \frac{\mu_i}{E_i} = \exp(\mathbf{x}'_i \boldsymbol{\beta}).$$

Since there are inflated numbers of 0 s and 1 s in our data, a "zero-one-inflated" model is introduced. As an extension of the zero-inflated method, a zero-one-inflated model employs two generating processes. The first process is governed by a multinomial distribution that generates structural zeros and ones. The second process is governed by a Poisson or negative binomial distribution that generates counts, some of which may be zero or one.

Denote the latent variable in the first process as  $I_i, i = 1, \dots, n$ , which follows a multinomial distribution with possible values 0, 1 and 2 with corresponding probabilities  $\pi_{0,i}, \pi_{1,i}, \pi_{2,i} = 1 - \pi_{0,i} - \pi_{1,i}$ . Here,  $N_i$  is frequency.

$$N_i \sim \begin{cases} 0 & I_i = 0 \\ 1 & I_i = 1 \\ P_i & I_i = 2. \end{cases}$$

Here,  $P_i$  may be a Poisson or negative binomial distribution. With this, the probability mass function of  $N_i$  is

$$f_{N_i}(n) = \pi_{0,i} I_{\{n=0\}} + \pi_{1,i} I_{\{n=1\}} + \pi_{2,i} P_i(n).$$

A logit specification is used to parameterize the probabilities for the latent variable  $I_i$ . Denote the covariates associated with  $I_i$  as  $\mathbf{z}_i$ . A logit specification is used to parameterize the probabilities for the latent variable  $I_i$ . Using level 2 as a reference, the specification is

$$\log \frac{\pi_{j,i}}{\pi_{2,i}} = \mathbf{z}'_i \boldsymbol{\gamma}_j, j = 0, 1.$$

Correspondingly,

$$\pi_{j,i} = \frac{\exp(\mathbf{z}'_i \boldsymbol{\gamma}_j)}{1 + \exp(\mathbf{z}'_i \boldsymbol{\gamma}_j) + \exp(\mathbf{z}'_i \boldsymbol{\gamma}_j)}, j = 0, 1.$$

$$\pi_{2,i} = 1 - \pi_{0,i} - \pi_{1,i}$$

Maximum likelihood estimation is used to fit the parameters.

### 2.3. Modeling Severity

**Modeling Severity Using GLMs.** For insurance analysts, one strength of the GLM approach is that the same set of routines can be used for continuous as well as discrete outcomes. For severities, it is common to use a gamma or inverse Gaussian distribution, often with a logarithmic link (primarily for parameter interpretability).

One strength of the linear exponential family that forms the basis of GLMs is that a sample average of outcomes comes from the same distribution as the outcomes. Specifically, suppose that we have  $m$  independent variables from the same distribution with location parameter  $\theta$  and scale parameter  $\phi$ . Then, the sample average comes from the same distributional family with location parameter  $\theta$  and scale parameter  $\phi/m$ . This result is helpful as insurance analysts regularly face grouped data as well as individual data. For example, [1] provides a demonstration of this basic property.

To illustrate, in the aggregate claims model, if individual losses have a gamma distribution with mean  $\mu_i$  and scale parameter  $\phi$ , then, conditional on observing  $N_i$  losses, the average aggregate loss  $\bar{S}_i$  has a gamma distribution with mean  $\mu_i$  and scale parameter  $\phi/N_i$ .

**Modeling Severity Using GB2.** The GLM is the workhorse for industry analysts interested in analyzing the severity of claims. Naturally, because of the importance of claims severity, a number of alternative approaches have been explored, *cf.*, [13] for an introduction. In this review, we focus on a specific alternative, using a distribution family known as the “generalized beta of the second kind”, or GB2, for short.

A random variable with a GB2 distribution can be written as

$$e^\mu \left( \frac{G_1}{G_2} \right)^\sigma = C_1 e^\mu F^\sigma = e^\mu \left( \frac{Z}{1-Z} \right)^\sigma,$$

where the constant  $C_1 = (\alpha_1/\alpha_2)^\sigma$ ,  $G_1$  and  $G_2$  are independent gamma random variables with scale parameter 1 and shape parameters  $\alpha_1$  and  $\alpha_2$ , respectively. Further, the random variable  $F$  has an  $F$ -distribution with degrees of freedom  $2\alpha_1$  and  $2\alpha_2$ , and the random variable  $Z$  has a beta distribution with parameters  $\alpha_1$  and  $\alpha_2$ .

Thus, the GB2 family has four parameters ( $\alpha_1, \alpha_2, \mu$  and  $\sigma$ ), where  $\mu$  is the location parameter. Including limiting distributions, the GB2 encompasses the “generalized gamma” (by allowing  $\alpha_2 \rightarrow \infty$ ) and hence the exponential, Weibull, and so forth. It also encompasses the “Burr Type 12” (by allowing  $\alpha_1 = 1$ ), as well as other families of interest, including the Pareto distributions. The GB2 is a flexible distribution that accommodates positive or negative skewness, as well as heavy tails. See, for example, [2] for an introduction to these distributions.

For incorporating covariates, it is straightforward to show that the regression function is of the form

$$E(y|\mathbf{x}) = C_2 \exp(\mu(\mathbf{x})) = C_2 e^{\mathbf{x}'\boldsymbol{\beta}},$$

where the constant  $C_2$  can be calculated with other (non-location) model parameters. Under the most commonly used way of parametrization for the GB2, where  $\mu$  is associated with covariates, if  $-\alpha_1 < \sigma < \alpha_2$ , then we have

$$C_2 = \frac{B(\alpha_1 + \sigma, \alpha_2 - \sigma)}{B(\alpha_1, \alpha_2)}$$

where  $B(\alpha_1, \alpha_2) = \Gamma(\alpha_1)\Gamma(\alpha_2)/\Gamma(\alpha_1 + \alpha_2)$ .

Thus, one can interpret the regression coefficients in terms of a proportional change. That is,  $\partial [\ln E(y)] / \partial x_k = \beta_k$ . In principle, one could allow for any distribution parameter to be a function of the covariates. However, following this principle would lead to a large number of parameters; this typically yields computational difficulties as well as problems of interpretations, [14]. In this paper,  $\mu$  is used to incorporate covariates. An alternative parametrization, as described in Appendix A.1, is introduced as an extension of the GLM framework.

#### 2.4. Tweedie Model

Frequency-severity modeling is widely used in insurance applications. However, for simplicity, it is also common to use only the aggregate loss  $S$  as a dependent variable in a regression. Because the distribution of  $S$  typically contains a positive mass at zero representing no claims, and a continuous component for positive values representing the amount of a claim, a widely used mixture is the Tweedie (1984) distribution. The Tweedie distribution is defined as a Poisson sum of gamma random variables. Specifically, suppose that  $N$  has a Poisson distribution with mean  $\lambda$ , representing the number of claims. Let  $y_j$  be an i.i.d. sequence, independent of  $N$ , with each  $y_j$  having a gamma distribution with parameters  $\alpha$  and  $\beta$ , representing the amount of a claim. Note,  $\beta$  is standard notation for this parameter used in loss-model textbooks, and the reader should understand it is different from the

bold-faced  $\beta$ , as the latter is a symbol we will use for the coefficients corresponding to explanatory variables. Then,  $S = y_1 + \dots + y_N$  is a Poisson sum of gammas.

To understand the mixture aspect of the Tweedie distribution, first note that it is straightforward to compute the probability of zero claims as  $\Pr(S = 0) = \Pr(N = 0) = e^{-\lambda}$ . The distribution function can be computed using conditional expectations,

$$\Pr(S \leq y) = e^{-\lambda} + \sum_{n=1}^{\infty} \Pr(N = n) \Pr(S_n \leq y), \quad y \geq 0.$$

Because the sum of i.i.d. gammas is a gamma,  $S_n = y_1 + \dots + y_n$  (not  $S$ ) has a gamma distribution with parameters  $n\alpha$  and  $\beta$ . For  $y > 0$ , the density of the Tweedie distribution is

$$f_S(y) = \sum_{n=1}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} \frac{\beta^{n\alpha}}{\Gamma(n\alpha)} y^{n\alpha-1} e^{-y\beta}.$$

From this, straight-forward calculations show that the Tweedie distribution is a member of the linear exponential family. Now, define a new set of parameters  $\mu, \phi, P$  through the relations

$$\lambda = \frac{\mu^{2-P}}{\phi(2-P)}, \quad \alpha = \frac{2-P}{P-1} \quad \text{and} \quad \frac{1}{\beta} = \phi(P-1)\mu^{P-1}.$$

Easy calculations show that

$$E S = \mu \quad \text{and} \quad \text{Var } S = \phi\mu^P,$$

where  $1 < P < 2$ . The Tweedie distribution can also be viewed as a choice that is intermediate between the Poisson and the gamma distributions.

In the basic form of the Tweedie regression model, the scale (or dispersion) parameter  $\phi$  is constant. However, if one begins with the frequency-severity structure, calculations show that  $\phi$  depends on the risk characteristics (*i. cf.*, [1]). Because of this and the varying dispersion (heteroscedasticity) displayed by many data sets, researchers have devised ways of accommodating and/or estimating this structure. The most common way is the so-called “double GLM” procedure proposed in [15] that models the dispersion as a known function of a linear combination of covariates (as for the mean, hence the name “double GLM”).

### 3. Multivariate Models and Methods

#### 3.1. Copula Regression

Copulas have been applied with GLMs in the biomedical literature since the mid-1990s ([16–18]). In the actuarial literature, the *t*-copula and the Gaussian copula with GLMs as marginal distributions were used to develop credibility predictions in [19]. In more general cases, [20] provides a detailed introduction of copula regression that focuses on the Gaussian copula and [21] surveys copula regression applications.

**Introducing Copulas.** Specifically, a *copula* is a multivariate distribution function with uniform marginals. Let  $U_1, \dots, U_p$  be  $p$  uniform random variables on  $(0, 1)$ . Their joint distribution function

$$C(u_1, \dots, u_p) = \Pr(U_1 \leq u_1, \dots, U_p \leq u_p) \tag{2}$$

is a copula.

Of course, we seek to use copulas in applications that are based on more than just uniformly distributed data. Thus, consider arbitrary marginal distribution functions  $F_1(y_1), \dots, F_p(y_p)$ . Then, we can define a multivariate distribution function using the copula such that

$$F(y_1, \dots, y_p) = C(F_1(y_1), \dots, F_p(y_p)). \quad (3)$$

If outcomes are continuous, then we can differentiate the distribution functions and write the density function as

$$f(y_1, \dots, y_p) = c(F_1(y_1), \dots, F_p(y_p)) \prod_{j=1}^p f_j(y_j), \quad (4)$$

where  $f_j$  is the density of the marginal distribution  $F_j$  and  $c$  is the copula density function.

It is easy to check from the construction in Equation (3) that  $F(\cdot)$  is a multivariate distribution function. Sklar established the converse in [22]. He showed that any multivariate distribution function  $F(\cdot)$  can be written in the form of Equation (3), that is, using a copula representation. Sklar also showed that, if the marginal distributions are continuous, then there is a unique copula representation. See, for example, the introductory book to copulas [23,24] for an introduction to copulas from an insurance perspective and [25] for a comprehensive modern treatment.

**Regression with Copulas.** In a regression context, we assume that there are covariates  $\mathbf{x}$  associated with outcomes  $\mathbf{y} = (y_1, \dots, y_p)'$ . In a parametric context, we can incorporate covariates by allowing them to be functions of the distributional parameters.

Specifically, we assume that there are  $n$  independent risks and  $p$  outcomes for each risk  $i$ ,  $i = 1, \dots, n$ . For this section, consider an outcome  $\mathbf{y}_i = (y_{i1}, \dots, y_{ip})'$  and  $K \times 1$  vector of covariates  $\mathbf{x}_i$ , where  $K$  is the number of covariates. The marginal distribution of  $y_{ij}$  is a function of  $\mathbf{x}_{ij}$ ,  $\beta_j$  and  $\theta_j$ . Here,  $\mathbf{x}_{ij}$  is a  $K_j \times 1$  vector of explanatory variables for risk  $i$  and outcome type  $j$ , a subset of  $\mathbf{x}_i$ , and  $\beta_j$  is a  $K_j \times 1$  vector of marginal parameters to be estimated. The systematic component  $\mathbf{x}'_{ij}\beta_j$  determines the location parameter. The vector  $\theta_j$  summarizes additional parameters of the marginal distribution that determine the scale and shape. Let  $F_{ij} = F(y_{ij}; \mathbf{x}_{ij}, \beta_j, \theta_j)$  denote the marginal distribution function.

This describes a classical approach to regression modeling, treating explanatory variances/covariates as non-stochastic (“fixed”) variables. An alternative is to think of the covariates themselves as random and perform statistical inference conditional on them. Some advantages of this alternative approach are that one can model the time-changing behavior of covariates, as in [26], or investigate non-parametric alternatives, as in [27]. These represent excellent future steps in copula regression modeling that are not addressed further in this article.

### 3.2. Multivariate Severity

For continuous severity outcomes, we may consider the density function  $f_{ij} = f(y_{ij}; \mathbf{x}_{ij}, \beta_j, \theta_j)$  associated with the distribution function  $F_{ij}$  and  $c$  the copula density function with parameter vector  $\alpha$ . With this, using Equation (4), the log-likelihood function of the  $i$ th risk is written as

$$l_i(\beta, \theta, \alpha) = \sum_{j=1}^p \ln f_{ij} + c(F_{i1}, \dots, F_{ip}; \alpha), \quad (5)$$

where  $\beta = (\beta_1, \dots, \beta_p)$  and  $\theta = (\theta_1, \dots, \theta_p)$  are collections of parameters over the  $p$  outcomes. This is a fully parametric set-up; the usual maximum likelihood techniques enjoy certain optimality properties and are the preferred estimation method.

If we consider only a single outcome, say  $y_{i1}$ , then the associated log-likelihood is  $\ln f_{i1}$ . Thus, the set of outcomes  $y_{i1}, \dots, y_{in1}$  allows for the usual “root- $n$ ” consistent estimator of  $\beta_1$ , and similarly for the other outcomes  $y_{ij}$ ,  $j = 2, \dots, p$ . By considering each outcome in isolation of the others, we can

get desirable estimators of the regression coefficients  $\beta_j, j = 1, \dots, p$ . These provide excellent starting values to calculate the fully efficient maximum likelihood estimators using the log-likelihood from Equation (5). Joe coined the phrase “inference for margins”, sometimes known by the acronym *IFM* in [28], to describe this approach to estimation.

In the same way, one can consider any pair of outcomes,  $(y_{ij}, y_{ik})$  for  $j \neq k$ . This permits consistent estimation of the marginal regression parameters as well as the association parameters between the  $j$ th and  $k$ th outcomes. As with the IFM, this technique provides excellent starting values of a fully efficient maximum likelihood estimation recursion. Moreover, they provide the basis for an alternative estimation method known as “composite likelihood”, *cf.*, [29] or [30], for a description in a copula regression context.

### 3.3. Multivariate Frequency

If outcomes are discrete, then one can take differences of the distribution function in Equation (3) to write the probability mass function as

$$f(y_1, \dots, y_p) = \sum_{j_1=1}^2 \cdots \sum_{j_p=1}^2 (-1)^{j_1 + \dots + j_p} C(u_{1,j_1}, \dots, u_{p,j_p}). \quad (6)$$

Here,  $u_{j,1} = F_j(y_{j-})$  and  $u_{j,2} = F_j(y_j)$  are the left- and right-hand limits of  $F_j$  at  $y_j$ , respectively. For example, when  $p = 2$ , we have

$$\begin{aligned} f(y_1, y_2) = & C(F_1(y_1), F_2(y_2)) - C(F_1(y_1-), F_2(y_2)) \\ & - C(F_1(y_1), F_2(y_2-)) + C(F_1(y_1-), F_2(y_2-)). \end{aligned}$$

It is straightforward in principle to estimate parameters using Equation (6) and standard maximum likelihood theory.

In practice, two caveats should be mentioned. The first is that the result of [22] only guarantees that the copula is unique over the range of the outcomes, a point emphasized with several interesting examples in [31]. In a regression context, this non-identifiability is less likely to be a concern, as noted in [25,29,30,32]. Moreover, the latter reference emphasizes that the Gaussian copula with binary data has been used for decades by researchers as this is just another form for the commonly used multivariate probit.

The second issue is computational. As can be seen in Equation (6), likelihood inference involves the computation of multidimensional rectangle probabilities. The review article [30] describes several variations of maximum likelihood that can be useful as the dimension  $p$  increases, see also [33]. As in [34], the composite likelihood method is used for computation. For large values of  $p$ , the pair (also known as “vine”) copula approach described in [35] for discrete outcomes seems to be a promising approach.

### 3.4. Multivariate Tweedie

As emphasized in [25] (p. 226), in copula regression it is possible to have outcomes that are combinations of continuous, discrete, and mixture distributions. One case of special interest in insurance modeling is the multivariate Tweedie, where each marginal distribution is a Tweedie and the margins are joined by a copula. Specifically, Shi considers different types of insurance coverages with Tweedie margins in [36].

To illustrate the general principles, consider the bivariate case ( $p = 2$ ). Suppressing the  $i$  index and covariate notation, the joint distribution is

$$f(y_1, y_2) = \begin{cases} C(F_1(0), F_2(0)) & y_1 = 0, y_2 = 0 \\ f(y_1)\partial_1 C(F_1(y_1), F_2(0)) & y_1 > 0, y_2 = 0 \\ f(y_2)\partial_2 C(F_1(0), F_2(y_2)) & y_1 = 0, y_2 > 0 \\ f(y_1)f(y_2)c(F_1(y_1), F_2(y_2)) & y_1 > 0, y_2 > 0 \end{cases},$$

where  $\partial_j C$  denotes the partial derivative of copula with respect to  $j$ th component.

See [36] for additional details of this estimation where he also described a double GLM approach to accommodate varying dispersion parameters.

### 3.5. Association Structures and Elliptical Copulas

First consider an outline of the evolution of multivariate regression modeling.

1. The multivariate normal (Gaussian) distribution has provided a foundation for multivariate data analysis, including regression. By permitting a flexible structure for the mean, one can readily incorporate complex mean structures including high order polynomials, categorical variables, interactions, semi-parametric additive structures, and so forth. Moreover, the variance structure readily permits incorporating time series patterns in panel data, variance components in longitudinal data, spatial patterns, and so forth. One way to get a feel for the breadth of variance structures readily accommodated is to examine options in standard statistical software packages such as PROC Mixed in [37] (for example, the TYPE switch in the RANDOM statement permits the choice of over 40 variance patterns).
2. In many applications, appropriately modeling the mean and second moment structure (variances and covariances) suffices. However, for other applications, it is important to recognize the underlying outcome distribution and this is where copulas come into play. As we have seen, copulas are available for any distribution function and thus readily accommodate binary, count, and long-tail distributions that cannot be adequately approximated with a normal distribution. Moreover, marginal distributions need not be the same, e.g., the first outcome may be a count Poisson distribution and the second may be a long-tail gamma.
3. Pair copulas (*cf.*, [25]) may well represent the next step in the evolution of regression modeling. A copula imposes the same dependence structure on all  $p$  outcomes whereas a pair copula has the flexibility to allow the dependence structure itself to vary in a disciplined way. This is done by focusing on the relationship between pairs of outcomes and examining conditional structures to form the dependence of the entire vector of outcomes. This approach is useful for high dimensional outcomes (where  $p$  is large), an important developing area of statistics. This represents an excellent future step in copula regression modeling that is not addressed further in this article.

As described in [25], there is a host of copulas available depending on the interests of the analyst and the scientific purpose of the investigation. Considerations for the choice of a copula may include computational convenience, interpretability of coefficients, a latent structure for interpretability, and a wide range of dependence, allowing both positive and negative associations.

For our applications of regression modeling, we typically begin with the elliptical copula family. This family is based on the family of elliptical distributions that includes the multivariate normal and  $t$ -distributions, see more in [38].

This family has most of the desirable traits that one would seek in a copula family. From our perspective, the most important feature is that it permits the same family association matrices found in the multivariate Gaussian distribution. This not only allows the analyst to investigate a wide degree of association patterns, but also allows estimation to be accomplished in a familiar way using the same structure as in the Gaussian family, e.g., [37].

For example, if the  $i$ th risk evolves over time, we might use a familiar time series model to represent associations, e.g.,

$$\Sigma_{AR1}(\rho) = \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{pmatrix}$$

such as an autoregressive of order 1 (AR1). See [19] for an actuarial application.

For a more complex example, suppose that  $\mathbf{y}_i = (\mathbf{y}_{i1}, \mathbf{y}_{i2}, \mathbf{y}_{i3})'$  represents three types of expenses for the  $i$ th company observed over 4 time periods. Then, we might use the following dependence structure

$$\Sigma = \begin{pmatrix} \Sigma_{AR1}(\rho_1) & \sigma_{12}\Sigma_{12} & \sigma_{13}\Sigma_{13} \\ \sigma_{12}\Sigma'_{12} & \Sigma_{AR1}(\rho_2) & \sigma_{23}\Sigma_{23} \\ \sigma_{13}\Sigma'_{13} & \sigma_{23}\Sigma'_{23} & \Sigma_{AR1}(\rho_3) \end{pmatrix},$$

as in [39]. This is a commonly used specification in models of several time series in econometrics where  $\sigma_{jk}$  represents a cross-sectional association between  $\mathbf{y}_{ij}$  and  $\mathbf{y}_{ik}$  and  $\Sigma_{jk}$  represents cross-associations with time lags.

### 3.6. Assessing Dependence

Dependence can be assessed at all stages of the model fitting process:

- Copula identification begins after marginal models have been fit. Then, use the “Cox-Snell” residuals from these models to check for association. Create simple correlation statistics (Spearman, polychoric) as well as plots ( $pp$  and tail dependence plots) to look for dependence structures and identify a parametric copula.
- After a model identification, estimate the model and examine how well the model fits. Examine the residuals to search for additional patterns using, for example, correlation statistics and  $t$ -plot (for elliptical copulas). Examine the statistical significance of fitted association parameters to seek a simpler fit that captures the important tendencies of the data.
- Compare the fitted model to alternatives. Use overall goodness of fit statistics for comparisons, including AIC and BIC, as well as cross-validation techniques. For nested models, compare via the likelihood ratio test and use Vuong’s procedure for comparing non-nested alternative specifications.
- Compare the models based on a held-out sample. Use statistical measures and economically meaningful alternatives such as the Gini statistic.

In the first and second step of this process, a variety of hypothesis tests and graphical methods can be employed to identify the specific type of copula (e.g., Archmedian, elliptical, extreme value, and so forth) that corresponds to the given data. Researchers have developed a graphical tool called the Kendall plot, or the K-plot for short, to detect dependence. See [40]. To determine whether a joint distribution corresponds to an Archimedean copula or a specific extreme-value copula, goodness-of-fit tests developed by [41–44] can be helpful. The reader may also reference [25] for a comprehensive coverage of the various assessment methods for dependence.

### 3.7. Frequency-Severity Modeling Strategy

Multivariate frequency-severity modeling strategies are a subset of the usual regression and copula identification and inference strategies. In absence of a compelling theory to suggest the appropriate covariates and predictors (which is generally the case for insurance applications), the modeling strategy consists of model identification, estimation, and inference. Typically, this is done in

a recursive fashion where one sets aside a random portion of the data for identification and estimation (the “training” sample), and one proceeds to validate and conduct inference on another portion (the “test” sample). See, for example, [45] for a description of this and many procedures for variable selection, mainly in a cross-sectional regression context.

You can think about the identification and estimation procedures as three components in a copula regression model:

1. Fit the mean structure. Historically, this is the most important aspect. One can apply robust standard error procedures to get consistent and approximately normally distributed coefficients, assuming a correct mean structure.
2. Fit the variance structure with a selected distribution. In GLMs, the choice of the distribution dictates the variance structure that can be over-ruled with a separately specified variance, e.g., a “double GLM.”
3. Fit the dependence structure with a choice of copula.

For frequency-severity modeling, there are two mean and variance structures to work with, one for the frequency and one for the severity.

### 3.7.1. Identification and Estimation

Although the estimation of parametric copulas is fairly established, the literature on identification of copulas is still in the early stage of development. As described in Section 3.2, maximum likelihood is the usual choice with an inference for margins and/or composite likelihood approach for starting values of the iterations. A description of composite likelihood in the context of copula modeling can be found in [25]. As noted here, composite likelihood may be particularly useful for multivariate discrete data when univariate margins have common parameters (p. 233, [25]). Another variation of maximum likelihood in copula regression is the “maximization by parts” method, as described in [20] and utilized in [46]. In the context of copula regressions, the idea behind this is to split the likelihood into two pieces, an easier part corresponding to the marginals and a more difficult part corresponding to the copula. The estimation routine takes advantage of these differing levels of difficulty in the calculations.

Identification of copula models used in regression typically starts with residuals from marginal fits. For severities, the idea is to estimate a parametric fit to the marginal distribution, such as normal or gamma regression. Then, one applies the distribution function (that depends on covariates) to the observation. Using notation, we can write this as  $F_i(y_i) = \hat{\epsilon}_i$ . This is known as the “probability integral transformation.” If the model is correctly specified, then the  $\hat{\epsilon}_i$  has a uniform (0,1) distribution. This is an idea that dates back to works by Cox and Snell in [47] and so these are often known as “Cox-Snell” residuals. For copula identification, it is recommended in [25] to take an inverse normal distribution transform (*i.e.*,  $\Phi^{-1}(\hat{\epsilon}_i)$ , for a standard normal distribution function  $\Phi$ ) to produce “normal scores.”

Because of the discreteness with frequencies, these residuals are not uniformly distributed even if the model is correctly specified. In this case, one can “jitter” the residuals. Specifically, define a modified distribution function  $F_i(y, \lambda) = \Pr(Y_i < y) + \lambda \Pr(Y_i = y)$  and let  $V$  be a uniform random number that is independent of  $Y_i$ . Then, we can define the jittered residual to be  $F_i(y_i, V) = \tilde{\epsilon}_i$ . If the model is correctly specified, then jittered residuals have a uniform (0,1) distribution, *cf.*, [48].

Compared to classical residuals, residuals from probability integral transforms have less ability to guide model development—we can only tell if the marginal models are approximately correct. The main advantage of this residual is that it is applicable to all (parametric) distributions. If you are working with a distribution that supports other definitions of residuals, then these are likely to be more useful because they may tell you how to improve your model specification, not whether or not it is approximately correct. If the marginal model fit is adequate, then we can think of the residuals as approximate realizations from a uniform distribution and use standard techniques from copula theory to identify a copula. We refer to [25] for a summary of this literature.

### 3.7.2. Model Validation

After identification and estimation, it is customary to compare a number of alternative models based on the training and on the test samples. For the training sample, the “in-sample” comparisons are typically based on the significance of the coefficients, overall goodness of fit measures (including information criteria such as *AIC* and *BIC*), cross-validation, as well as likelihood ratio comparisons for nested models.

For comparisons among non-nested parametric models, it is now common in the literature to cite a statistic due to Vuong [49]. For this statistic, one calculates the contribution to the logarithmic likelihood such as in Equation (5) for two models, say,  $l_i^{(1)}$  and  $l_i^{(2)}$ . One prefers Model (1) compared to Model (2) if the average difference,  $\bar{D} = m^{-1} \sum_{i=1}^m D_i$ , is positive, where  $D_i = l_i^{(1)} - l_i^{(2)}$  and  $m$  is the size of the validation sample. To assess the significance of this difference, one can apply approximate normality with approximate standard errors given as  $SD_D / \sqrt{m}$  where  $SD_D^2 = (m - 1)^{-1} \sum_{i=1}^m (D_i - \bar{D})^2$ . In a copula context, see (p. 257, [25]) for a detailed description of this procedure, where sample size adjustments similar to those used in *AIC* and *BIC* are also introduced.

Comparison among models using test data, or “out-of-sample” comparisons are also important in insurance because many of these models are used for predictive purposes such as setting rates for new customers. Out-of-sample measures compare held-out observations to those predicted by the model. Traditionally, absolute values and squared differences have been used to summarize differences between these two. However, for many insurance data sets, there are large masses at zero, meaning that these traditional metrics are less helpful. To address this problem, a newer measure is developed in [50] that they call the “Gini index.” In this context, the Gini index is twice the average covariance between the predicted outcome and the rank of the predictor. In order to compare models, Theorem 5 of [50] provides standard errors for the difference of two Gini indices.

## 4. Frequency Severity Dependency Models

In traditional models of insurance data, the claim frequency is assumed to be independent of claim severity. We emphasize in Appendix A.4 that the average severity may depend on frequency, even when this classical assumption holds.

One way of modeling the dependence is through the conditioning argument developed in Section 2.1. An advantage of this approach is that the frequency can be used as a covariate to model the average severity. See [51] for a healthcare application of this approach. For another application, a Bayesian approach for modeling claim frequency and size was proposed in [52], with both covariates as well as spatial random effects taken into account. The frequency was incorporated into the severity model as covariate. In addition, they checked both individual and average claim modeling and found the results were similar in their application.

As an alternative approach, copulas are widely used for frequency severity dependence modeling. In [46], Czado *et al.* fit Gaussian copula on Poisson frequency and gamma severity and used an optimization by parts method from [53] to do the estimation. They derived the conditional distribution of frequency given severity. In [54], the distribution of policy loss is derived without the independence assumption between frequency and severity. They also showed that the ignoring of dependence can lead to underestimation of loss. A Vuong test was adopted to select the copula.

To see how the copula approach works, recall that  $\bar{S}$  represents average severity of claims and  $N$  denotes frequency. Using a copula, we can express the likelihood as

$$f_{\bar{S}, N}(s, n) = \begin{cases} f_{\bar{S}, N}(s, n | N > 0) P(N > 0) & \text{for } n > 0 \\ P(N = 0) & \text{for } n = s = 0 \end{cases}$$

Denote

$$D_1(u, v) = \frac{\partial}{\partial u} C(u, v) = P(V \leq v | U = u).$$

With this,

$$\begin{aligned} f_{\bar{s},N}(s, n|N > 0) &= \frac{\partial}{\partial s} P(\bar{S} \leq s, N \leq n|N > 0) \\ &= \frac{\partial}{\partial s} C(F_{\bar{S}}(s), F_N(n|N > 0)) \\ &= f_{\bar{S}}(s) D_1(F_{\bar{S}}(s), F_N(n|N > 0)). \end{aligned}$$

This yields the following expression for the likelihood

$$f_{\bar{s},N}(s, n) = \begin{cases} f_{\bar{S}}(s)P(N > 0)(D_1(F_{\bar{S}}(s), F_N(n|N > 0)) \\ \quad - D_1(F_{\bar{S}}(s), F_N(n-1|N > 0))) & \text{for } s > 0, n \geq 1 \\ P(N = 0) & \text{for } s = 0, N = 0. \end{cases}$$

For another approach, Shi *et al.* also built a dependence model between frequency and severity in [55]. They used an extra indicator variable for occurrence of claim to deal with the zero-inflated part, and built a dependence model between frequency and severity conditional on positive claim. The two approaches described previously were compared; one approach using frequency as a covariate for the severity model, and the other using copulas. They used a zero-truncated negative binomial for positive frequency and the GG model for severity. In [56], a mixed copula regression based on GGS copula (see [25] for an explanation of this copula) was applied on a medical expenditure panel survey (MEPS) dataset. In this way, the negative tail dependence between frequency and average severity can be captured.

Brechmann *et al.* applied the idea of the dependence between frequency and severity to the modeling of losses from operational risks in [57]. For each risk class, they considered the dependence between aggregate loss and the presence of loss. Another application of this methodology in operational risk aggregation can be found in [58]. Li *et al.* focused on two dependence models; one for the dependence of frequencies across different business lines, and another for the aggregate losses. They applied the method on Chinese banking data and found significant difference between these two methods.

## 5. LGPIF Case Study

We demonstrate the multivariate frequency severity modeling approach using a data set from the Wisconsin Local Government Property Insurance Fund (LGPIF). The LGPIF was established to provide property insurance for local government entities that include counties, cities, towns, villages, school districts, fire departments, and other miscellaneous entities, and is administered by the Wisconsin Office of the Insurance Commissioner. Properties covered under this fund include government buildings, vehicles, and equipment. For example, a county entity may need coverage for its snow plowing trucks, in addition to its building and contents [59]. These data provide a good example of a typical multi-line insurance company encountered in practice. More details about the project may be found at the Local Government Property Insurance Fund project website [60].

### 5.1. Data / Problem Description

The data consist of six coverage groups; building and content (BC), contractor's equipment (IM), comprehensive new (PN), comprehensive old (PO), collision new (CN), collision old (CO) coverage. The data are longitudinal, and Tables 1 and 2 provide summary statistics for the frequencies and severities of claims within the in-sample years 2006 to 2010, and the validation sample 2011.

**Table 1.** Data Summary by Coverage, 2006–2010 (Training Sample).

	Average Frequency	Average Severity	Annual Claims in Each Year	Average Coverage (Million)	Number of Claims	Number of Observations
BC	0.879	9868	17,143	37.050	4992	5660
IM	0.056	624	766	0.848	318	4622
PN	0.159	197	466	0.158	902	1638
PO	0.103	311	504	0.614	587	2138
CN	0.127	374	744	0.096	720	1529
CO	0.120	538	951	0.305	680	2013

**Table 2.** Data Summary by Coverage, 2011 (Validation Sample).

	Average Frequency	Average Severity	Annual Claims in Year	Average Coverage (Million)	Number of Claims	Number of Observations
BC	0.945	8352	20,334	42.348	1038	1095
IM	0.076	382	645	0.972	83	904
PN	0.224	307	634	0.172	246	287
PO	0.128	220	312	0.690	140	394
CN	0.125	248	473	0.093	137	268
CO	0.081	404	656	0.375	89	375

Table 3 describes each coverage group. Automobile coverage is subdivided into four subcategories, which correspond to combinations for collision versus comprehensive and for new versus old cars.

**Table 3.** Description of Coverage Groups.

Code	Name of Coverage	Description
BC	Building and Contents	This coverage provides insurance for buildings and the properties within. In case the policyholder has purchased a rider, claims in this group may reflect additional amounts covered under endorsements.
IM	Contractor's Equipment	IM, an abbreviation for "inland marine" is used as the coverage code for equipments coverage, which originally belong to contractors.
C	Collision	This provides coverage for impact of a vehicle with an object, impact of vehicle with an attached vehicle, or overturn of a vehicle.
P	Comprehensive	Direct and accidental loss or damage to motor vehicle, including breakage of glass, loss caused by missiles, falling objects, fire, theft, explosion, earthquake, windstorm, hail, water, flood, malicious mischief or vandalism, riot or civil common, or colliding with a bird or animal.
N	New	This code is used as an indication that the coverage is for vehicles of current model year, or 1~2 years prior to the current model year.
O	Old	This code is used as an indication that the coverage is for vehicles three or more years prior to the current model year.

From Table 3, there are collision and comprehensive coverages, each for new and old vehicles of the entity. Hence, an entity can potentially have collision coverage for new vehicles (CN), collision coverage for old vehicles (CO), comprehensive coverage for new vehicles (PN), and comprehensive coverage for old vehicles (PO). Hence, in our analysis, we consider these sub-coverages as individual lines of businesses, and work with six separate lines, including building and contents (BC), and contractor's equipment (IM) as separate lines also.

Preliminary dependence measures for discrete claim frequencies and continuous average severities can be obtained using polychoric and polyserial correlations. These dependence measures both assume latent normal variables, whose values fall within the cut-points of the discrete variables. The polychoric correlation is the inferred latent correlation between two ordered categorical variables;

the polyserial correlation is the inferred latent correlation between a continuous variable and an ordered categorical variable, *cf.* [25].

Table 4 shows the polychoric correlation among the frequencies of the six coverage groups. Note that these dependencies in Table 4 are measured before controlling for the effects of explanatory variables on the frequencies. As Table 4 shows, there is evidence of correlation across different lines, however these cross-sectional dependencies may be due to correlations in the exposure amounts or, in other words, the sizes of the entities.

**Table 4.** Polychoric Correlation among Frequencies of Claims.

	BC	IM	PN	PO	CN
IM	0.506				
PN	0.465	0.584			
PO	0.490	0.590	0.771		
CN	0.492	0.541	0.679	0.566	
CO	0.559	0.601	0.642	0.668	0.646

The dependence between frequencies and average claim severities is often of interest to modelers. In Appendix A.4 we show that average severity may depend on frequency, even when the classical assumption, independence of frequency and individual severities, holds. Our data are consistent with this result. The diagonal entries of Table 5 show the polyserial correlations between the frequency and severity of each coverage group.

**Table 5.** Polyserial Correlation between Frequencies and Severities.

	BC Frequency	IM Frequency	PN Frequency	PO Frequency	CN Frequency	CO Frequency
BC Severity	−0.033	0.029	−0.063	−0.069	0.020	−0.050
IM Severity	−0.033	−0.078	0.110	0.249	0.159	0.225
PN Severity	0.074	0.275	−0.146	−0.216	0.119	0.143
PO Severity	0.111	0.171	−0.161	−0.119	0.258	0.137
CN Severity	−0.112	−0.174	−0.003	0.135	0.032	−0.175
CO Severity	−0.099	−0.079	−0.055	−0.083	−0.068	−0.032

According to Table 5, the observed correlation between frequency and severity is small. For the CN line, a positive correlation can be observed although very small (0.032, while the other correlations between frequency and severity are negative). Again, these numbers only provide a rough idea of the dependency. Table 6 shows the Spearman correlation between the average severities, for those observations with at least one positive claim. The correlation among the severities of new and old car comprehensive coverage is high.

**Table 6.** Correlation among Average Severities.

	BC	IM	PN	PO	CN
IM	0.220				
PN	0.098	0.095			
PO	0.229	0.118	0.415		
CN	0.084	0.237	0.166	0.200	
CO	0.132	0.261	0.075	0.140	0.244

In summary, these summary statistics show that there are potentially interesting dependencies among the response variables.

## Explanatory Variables

Table 7 shows the number of observations available in the data set, for years 2006–2010.

**Table 7.** Number of Observations.

	BC	IM	PN	PO	CN	CO
Coverage > 0	5660	4622	1638	2138	1529	2013
Average Severity > 0	1684	236	315	263	370	362

Explanatory variables used are summarized in Table 8. The marginal analyses for each line are performed on the subset for which the coverage amounts shown in Table 7 are positive.

**Table 8.** Summary of Explanatory Variables.

Variable Name	Description	Mean
lnCoverageBC	Log of the building and content coverage amount.	37.050
lnCoverageIM	Log of the contractor’s equipment coverage amount.	0.848
lnCoveragePN	Log of the comprehensive coverage amount for new vehicles.	0.158
lnCoveragePO	Log of the comprehensive coverage amount for old vehicles.	0.614
lnCoverageCN	Log of the collision coverage amount for new vehicles.	0.096
lnCoverageCO	Log of the collision coverage amount for old vehicles.	0.305
NoClaimCreditBC	Indicator for no building and content claims in prior year.	0.328
NoClaimCreditIM	Indicator for no contractor’s equipment claims in prior year.	0.421
NoClaimCreditPN	Indicator for no comprehensive claims for new cars in prior year.	0.110
NoClaimCreditPO	Indicator for no comprehensive claims for old cars in prior year.	0.170
NoClaimCreditCN	Indicator for no collision claims for new cars in prior year.	0.090
NoClaimCreditCO	Indicator for no collision claims for old cars in prior year.	0.140
EntityType	City, County, Misc, School, Town (Categorical)	
lnDeductBC	Log of the BC deductible level, chosen by the entity.	7.137
lnDeductIM	Log of the IM deductible level, chosen by the entity.	5.340

## 5.2. Marginal Model Fitting—Zero/One Frequency, GB2 Severity

For each coverage type, a frequency-severity model is fit marginally.

### 5.2.1. BC (Building and Contents) Frequency Modeling

In the frequency part, we fit several commonly employed count models: Poisson, negative binomial (NB), zero-inflated Poisson (zeroinflPoisson), zero-inflated negative binomial (zeroinflNB). Our data not only exhibit a large mass at 0, as with many other insurance claims data, but also an inflated number of 1 s. For BC, there are 997 policies with 1 claim. This can be compared to the expected number under zero-inflated Poisson, 754, and under the zero-inflated negative binomial, 791. (See Table 9 for details). These zero-inflated models underestimate the point mass at 1 due to the shrinkage to 0. Thus, alternative “zero-one-inflated” models are introduced in Section 2.2.

Table 9 shows the expected count for each frequency value under different models and the empirical values from the data. A Poisson distribution underestimates the zero proportions while zero-inflated and negative binomial models underestimate the proportion of 1 s. The zero-one inflated models do provide the best fits for simultaneously estimating the probability of a zero and a one.

Chi-square goodness of fit statistics can be used to compare different models. Table 10 shows the result. It is calculated depending on Table 9. The zero-one-inflated negative binomial is significantly better than other methods.

**Table 9.** Comparison between Empirical Values and Expected Values for the building and contents (BC) Line.

	Empical	ZeroinflPoisson	ZeroonePoisson	Poisson	NB	ZeroinflNB	ZerooneNB
0	3976	4038.125	3975.403	3709.985	4075.368	4093.699	3996.906
1	997	754.384	1024.219	1012.267	809.077	791.424	1003.169
2	333	355.925	276.082	417.334	313.359	314.618	280.600
3	136	187.897	146.962	202.288	155.741	157.282	136.758
4	76	106.780	82.052	106.874	88.866	89.615	75.822
5	31	63.841	48.426	60.160	55.484	55.697	46.021
6	19	39.850	30.212	36.540	36.919	36.845	29.854
7	19	26.082	19.850	24.261	25.765	25.553	20.379
8	16	18.025	13.670	17.440	18.663	18.395	14.482
9	5	13.165	9.808	13.222	13.932	13.652	10.632
10	7	10.087	7.269	10.305	10.664	10.393	8.016
11	2	8.007	5.505	8.124	8.336	8.084	6.180
12	4	6.505	4.219	6.427	6.636	6.406	4.855
13	5	5.357	3.248	5.086	5.367	5.159	3.875
14	5	4.441	2.502	4.024	4.401	4.214	3.136
15	2	3.690	1.925	3.182	3.653	3.485	2.569
16	4	3.062	1.479	2.519	3.066	2.914	2.127
17	3	2.530	1.134	1.999	2.598	2.460	1.777
18	1	2.077	0.867	1.597	2.221	2.095	1.498
≥ 19	19	10.168	5.167	16.366	19.876	18.004	11.343
0 proportion	0.702	0.713	0.702	0.655	0.720	0.723	0.706
1 proportion	0.176	0.133	0.181	0.179	0.143	0.140	0.177

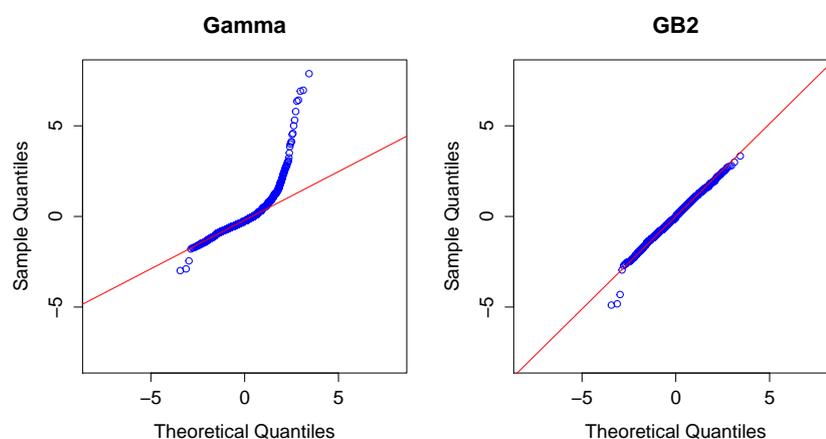
**Table 10.** Goodness of Fit Statistics for BC Line.

ZeroinflPoisson	ZeroonePoisson	Poisson	NB	ZeroinflNB	ZerooneNB
154.573	77.064	105.201	88.086	98.400	34.515

5.2.2. BC (Building and Contents) Severity Modeling

In the average severity part, the most commonly used distribution, gamma, is fit and compared with the GB2 model. To do the goodness of fit test, the quantiles of normal Cox-Snell residuals are compared with normal quantiles.

Figure 1 shows the residual plot of severity fitted with gamma and GB2. Clearly, the gamma does not fit well especially in the tail part.



**Figure 1.** QQ Plot for Residuals of Gamma and GB2 Distribution for BC.

### 5.2.3. Building and Contents Model Summary

Table 11 shows the coefficients for the fitted marginal models. Here, coefficients of GB2, NB and the zero-one-inflated parts are provided.

**Table 11.** Coefficients of Marginal Models for BC Line.

	Variable Name	Coef.	Standard Error	
GB2	(Intercept)	5.620	0.199	***
	lnCoverageBC	0.136	0.029	***
	NoClaimCreditBC	0.143	0.076	.
	lnDeductBC	0.321	0.034	***
	EntityType: City	−0.121	0.090	
	EntityType: County	−.059	0.112	
	EntityType: Misc	0.052	0.142	
	EntityType: School	0.182	0.092	*
	EntityType: Town	−0.206	0.141	
	$\sigma$	0.343	0.070	
	$\alpha_1$	0.486	0.119	
	$\alpha_2$	0.349	0.083	
NB	(Intercept)	−0.798	0.198	***
	lnCoverageBC	0.853	0.033	***
	NoClaimCreditBC	−0.400	0.132	**
	lnDeductBC	−0.232	0.035	***
	EntityType: City	−0.074	0.090	
	EntityType: County	0.015	0.117	
	EntityType: Misc	−0.513	0.188	**
	EntityType: School	−1.056	0.094	***
	EntityType: Town	−0.016	0.160	
	log(size)	0.370	0.115	
Zero	(Intercept)	−6.928	0.840	***
	CoverageBC	−0.408	0.135	**
	lnDeductBC	0.880	0.108	***
	NoClaimCreditBC	0.954	0.459	*
One	(Intercept)	−5.466	0.965	***
	CoverageBC	0.142	0.117	
	lnDeductBC	0.323	0.137	*
	NoClaimCreditBC	0.669	0.447	

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1.

### 5.2.4. Marginal Models for Other Lines

Appendix A.2 provides the model selection and marginal model results for lines other than building and contents.

### 5.3. Copula Identification and Fitting

Dependence is fit at two levels. The first is between frequency and average severity within each line. The second is among different lines.

#### 5.3.1. Frequency Severity Dependence

Vuong’s test, as described in Section 3.7.2, is used for copula selection. Specifically, we consider two models  $M^{(1)}$  and  $M^{(2)}$ , in our example,  $M^{(1)}$  is Gaussian copula while  $M^{(2)}$  is  $t$  copula. Let  $\Delta_{12}$

be the difference in divergence from models  $M^{(1)}$  and  $M^{(2)}$ . When the true density is  $g$ , this can be written as

$$\Delta_{12} = n^{-1} \sum_i \left\{ E_g[\log f^{(2)}(Y_i; x_i, \theta^{(2)})] - E_g[\log f^{(1)}(Y_i; x_i, \theta^{(1)})] \right\}.$$

A large sample 95% confidence interval for  $\Delta_{12}$ ,  $\bar{D} \pm 1.96 \times n^{-1/2}SD_D$ , is provided in Table 12. Table 12 shows the comparison of Gaussian copula against  $t$  copula with commonly used degrees of freedom for frequency and severity dependence in BC line. An interval completely below 0 indicates that copula 1 is significantly better than copula 2. Thus, the Gaussian copula is preferred.

**Table 12.** Vuong Test of Copulas for BC Frequency and Severity Dependence.

Copula 1	Copula 2	95% Interval	
Gaussian	$t(df = 3)$	-0.0307	-0.0122
Gaussian	$t(df = 4)$	-0.0202	-0.0065
Gaussian	$t(df = 5)$	-0.0147	-0.0038
Gaussian	$t(df = 6)$	-0.0114	-0.0023

Maximum likelihood estimation with the full multivariate likelihood, which estimates parameters in marginal and copula models simultaneously, is fit here. Table 13 shows parameters of BC line with the full likelihood method. Here the marginal dispersion parameters are fixed from marginal models. By comparing Tables 11 and 13, it can be seen that the coefficients are close. As pointed out in [25], inference functions for marginals, with the results in Table 11, is efficient and can provide a good starting point for the full likelihood method, as in Table 13.

**Table 13.** Coefficients of Total Likelihood for BC Line.

	Variable Name	Coef.	Standard Error	
GB2	(Intercept)	5.629	0.195	***
	lnCoverageBC	0.144	0.029	***
	NoClaimCreditBC	0.222	0.076	**
	lnDeductBC	0.320	0.031	***
	EntityType: City	-0.148	0.090	.
	EntityType: County	-0.043	0.111	.
	EntityType: Misc	0.158	0.143	.
	EntityType: School	0.225	0.092	*
	EntityType: Town	-0.218	0.141	.
		$\sigma$	0.343	0.070
	$\alpha_1$	0.486	0.119	.
	$\alpha_2$	0.349	0.083	.
NB	(Intercept)	-0.789	0.083	***
	lnCoverageBC	1.003	0.001	***
	NoClaimCreditBC	-0.297	0.172	.
	lnDeductBC	-0.230	0.001	***
	EntityType: City	-0.068	0.097	.
	EntityType: County	-0.489	0.109	***
	EntityType: Misc	-0.468	0.202	*
	EntityType: School	-0.645	0.083	***
	EntityType: Town	0.267	0.166	.
		log(Size)	0.370	0.115

Table 13. Cont.

	Variable Name	Coef.	Standard Error	
Zero	(Intercept)	−6.246	0.364	***
	lnCoverageBC	−0.338	0.047	***
	lnDeductBC	0.910	0.050	***
	NoClaimCreditBC	0.888	0.355	*
One	(Intercept)	−5.361	0.022	***
	lnCoverageBC	0.345	0.013	***
	lnDeductBC	0.335	0.010	***
	NoClaimCreditBC	0.556	0.431	
$\rho$	Dependence	−0.132	0.033	***

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1

For other lines, the results of the full likelihood method are summarized in Table 14. As described in Appendix A.2, the other lines use a negative binomial model for claim frequencies, not the 0–1 inflated model introduced in Section 2.2. For the CO line severity,  $\frac{1}{\sigma}$  is fitted for the purpose of computation. Model selection and marginal model results can be found in Appendix A.2.

Table 14. Coefficients of Total Likelihood for Other Lines.

	IM		PN		PO		CN		CO							
	Coef.	Std. Error	Coef.	Std. Error	Coef.	Std. Error	Coef.	Std. Error	Coef.	Std. Error						
GB2	(Intercept)	8.153	0.823	***	7.918	0.046	***	7.554	0.092	***	6.773	0.059	***	9.334	0.000	***
	lnCoverage	0.304	0.065	***	0.078	0.045		0.081	0.057		0.137	0.039	***	0.161	0.000	***
	NoClaimCredit	0.190	0.202		0.021	0.209		0.695	0.194	***	0.140	0.144		−0.296	0.001	***
	lnDeduct	0.028	0.125													
	$\sigma$	0.955	0.365		0.047	0.043		0.100	0.130		0.863	0.513		40.193	31.080	
	$\alpha_1$	1.171	0.630		0.054	0.050		0.102	0.137		4.932	6.441		0.038	0.030	
	$\alpha_2$	1.337	0.856		0.076	0.068		0.108	0.145		1.279	1.131		0.025	0.019	
	NB	(Intercept)	−1.331	0.594	*	−2.160	0.284	***	−2.664	0.297	***	−0.467	0.158	**	−1.746	0.187
Coverage		0.796	0.077	***	0.239	0.065	***	0.490	0.067	***	0.487	0.054	***	0.782	0.056	***
NoClaimCredit		−0.371	0.141	**	−0.588	0.194	**	−0.612	0.177	***	−0.668	0.157	***	−0.324	0.139	*
lnDeduct		−0.140	0.085													
EntityType: City		−0.306	0.235		0.574	0.330		0.411	0.376		0.433	0.186	*	0.680	0.232	**
EntityType: County		0.139	0.274		3.083	0.294	***	2.477	0.329	***	1.131	0.172	***	1.284	0.211	***
EntityType: Misc		−2.195	1.024	*	−0.060	0.642		−0.508	0.709		−0.323	0.456		0.486	0.442	
EntityType: School		−0.032	0.292		0.389	0.297		0.926	0.327	**	−0.192	0.185		1.350	0.208	***
EntityType: Town size		−0.405	0.277		−0.579	0.481		−1.022	0.650		−1.529	0.385	***	−0.450	0.355	
$\rho$	Dependence	−0.109	0.097		−0.154	0.064	*	−0.166	0.073	*	0.171	0.064	**	0.009	0.045	

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1

Table 13 shows significantly strong negative association between frequency and average severity for the building and contents (BC) line. In contrast, the results are mixed for other lines. Table 14 shows no significant relationships for the CO and IM lines, mild negative relationships for the PN and PO lines, and a strong positive relationship for the CN line. For the BC and CN lines, these results are consistent with the polyserial correlations in Table 5, calculated without covariates.

### 5.3.2. Dependence between Different Lines

The second level of dependence lies between different lines. In this section, the dependence model for frequencies, severities and aggregate loss with Tweedie margins, as in Section 3.4, are fit. Here, we use marginal results from the inference functions for margins method. In principle, full likelihood can be used. As mentioned previously in this section, in our case, the results of inference functions for margins are close to full likelihood estimation.

Tables 15 and 16 show the dependence parameters of copula models for frequencies and severities, respectively. A Gaussian copula is applied and the composite likelihood method is used for computation. Comparing Tables 4 and 15, it can be seen that frequency dependence parameters decrease substantially. This is due to controlling for the effects of explanatory variables. In contrast, comparing Tables 6 and 16, there appears to be little change in the dependence parameters. This may be due to the smaller impact that explanatory variables have on the severity modeling when compared to frequency modeling.

**Table 15.** Dependence Parameters for Frequency.

	BC	IM	PN	PO	CN
IM	0.190				
PN	0.141	0.162			
PO	0.054	0.206	0.379		
CN	0.101	0.149	0.271	0.081	
CO	0.116	0.213	0.151	0.231	0.297

**Table 16.** Dependence Parameters for Severity.

	BC	IM	PN	PO	CN
IM	0.145				
PN	0.134	0.051			
PO	0.298	0.099	0.498		
CN	0.062	0.110	0.156	0.168	
CO	0.106	0.215	0.083	0.080	0.210

Table 17 shows the result of dependence parameters for different lines with Tweedie margins. The coefficients of marginal models are in Appendix A.3.

**Table 17.** Dependence Parameters for Tweedies.

	BC	IM	PN	PO	CN
IM	0.210				
PN	0.279	0.367			
PO	0.358	0.412	0.559		
CN	0.265	0.266	0.553	0.328	
CO	0.417	0.359	0.496	0.562	0.573

## 6. Out-of-Sample Validation

For out-of-sample validation, the coefficient estimates from the marginals and the dependence parameters are used to obtain the predicted claims with the held-out 2011 data. We compare the independent case, dependent frequency-severity model, and the dependent pure premium approach.

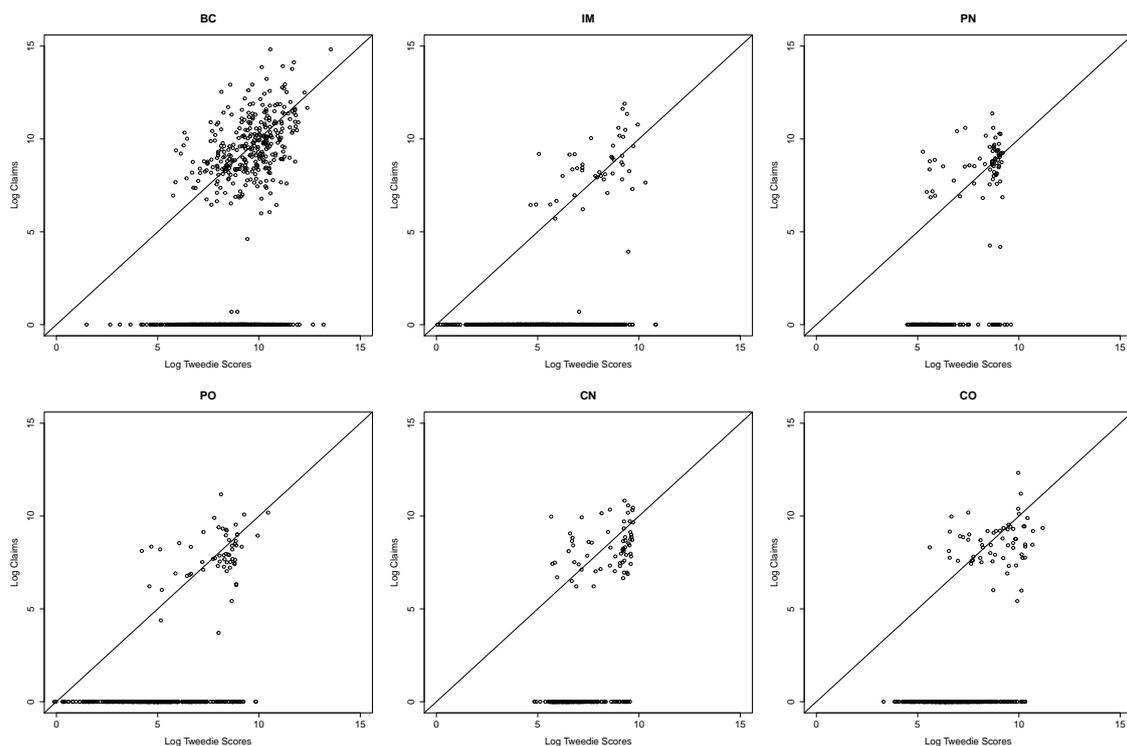
### 6.1. Spearman Correlation

The out of sample validation is performed on the 2011 held-out data, where there are 1098 observations. The claim scores for the pure premium approach are obtained using the conditional mean of the Tweedie distribution for each policyholder. For the frequency-severity approach, the conditional mean for the zero-one-inflated negative binomial distribution is multiplied to the first moment of the GB2 severity distribution for the policyholder. Claim scores for the dependent pure premium approach and the dependent frequency-severity approach are computed using a Monte Carlo simulation of the normal copula.

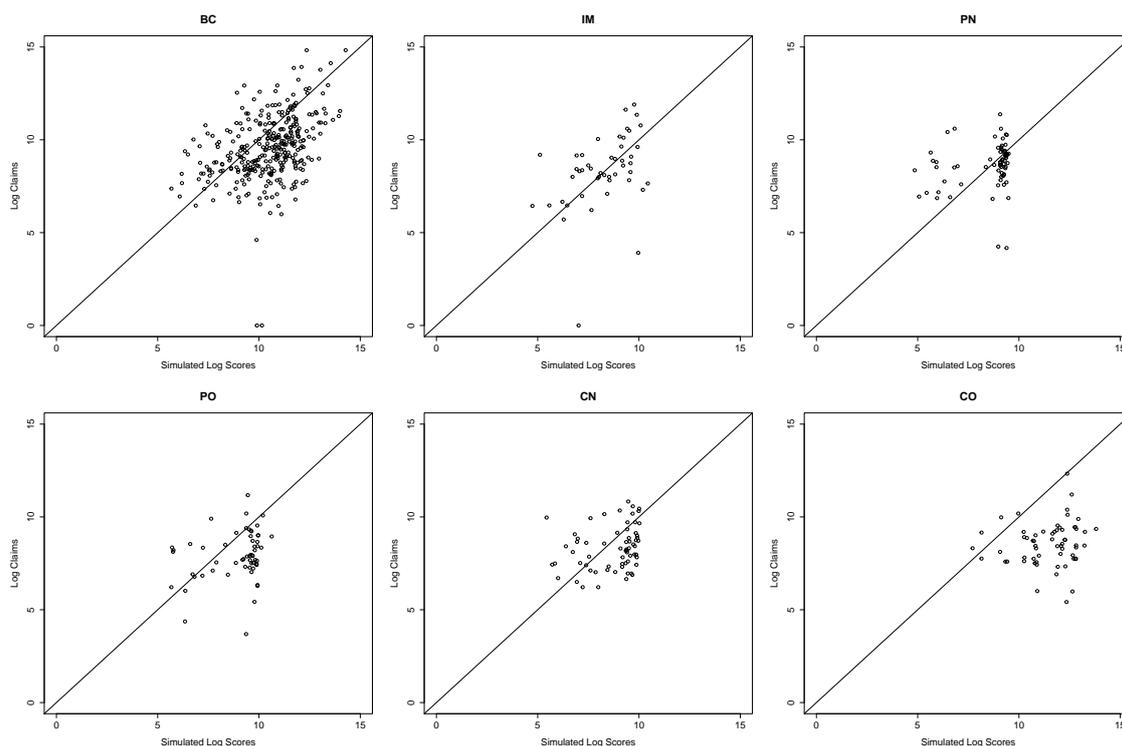
We first consider the nonparametric Spearman correlation between model predictions and the held-out claims. Four models are considered: the frequency-severity and pure premium (Tweedie) model, assuming independence among lines, and assuming a Gaussian copula among lines. As can be seen from Table 18, the predicted claims are about the same whether dependence is considered or not. The interesting question is how much improvement the zero-one-inflated negative binomial model, and the long-tail distribution (*GB2*) marginals bring. We observe that the long-tail nature of the severity distribution sometimes results in a large predicted claim. We found that prediction of the mean, using the first moment, can be numerically sensitive. Figure 2 shows a plot of the predicted claims against the out-of-sample claims, for the independent pure premium approach. Figure 3 shows the dependent frequency-severity approach.

**Table 18.** Out-of-Sample Correlation.

	BC	IM	PN	PO	CN	CO	Total
Independent Tweedie	0.410	0.304	0.602	0.461	0.512	0.482	0.500
Dependent Tweedie (Monte Carlo)	0.412	0.305	0.601	0.462	0.511	0.481	0.501
Independent Frequency-Severity	0.440	0.308	0.590	0.475	0.525	0.469	0.498
Dependent Frequency-Severity (Monte Carlo)	0.435	0.308	0.590	0.477	0.525	0.485	0.521



**Figure 2.** Out-of-Sample Validation for Independent Tweedie. (In these plots, the conditional mean for each policyholder is plotted against the claims.)



**Figure 3.** Out-of-Sample Validation for Dependent Frequency-Severity. (In these plots, the claim scores for each line is simulated from the frequency-severity model with dependence, using a Monte Carlo approach with  $B = 50,000$  samples from the normal copula. The model with 01-NB and GB2 marginals show clear improvement for the BC line, in particular for the upper tail prediction. For other lines such as CO, the GB2 marginal results in miss-scaling).

## 6.2. Gini Index

To further validate our results, we use the Gini index to measure the satisfaction of the fund manager with each score. The Gini index is calculated using relativities computed with the actual premium collected by the LGPIF in 2011 as the denominator, with the scores predicted by each model as numerator. This means we are looking for improvements over the original premium scores used by the LGPIF. We expect the Gini index to be higher with the frequency-severity approach, as the fit for the upper tail is better. Figure 4 compares the independent Tweedie approach, and the dependent frequency-severity approach. For the dependent frequency-severity approach, a random 6-dimensional vector is sampled from the normal copula, and the quantiles are converted to frequencies and severities.

For the BC line scores calculated using the Tweedie model, we obtain  $-1.74\%$  Gini index, meaning this model does not improve the existing premium scores used by the fund. Note that in [59], where the interest is more in the regularization problem, a constant premium is used as denominator for assessing the relativity. Here, the denominator used is the original premiums, which means in order for the index to be positive, there must be an improvement over the original premiums. The dependent frequency-severity scores with  $B = 50,000$ , normal copula, and zero-one-inflated NB and GB2 marginals results in a Gini index of  $22.77\%$ , meaning a clear improvement from the original premium scores. As a side note, the Spearman correlations are: original BC premiums  $42.59\%$ , Tweedie model  $40.97\%$ , and Frequency-severity model  $43.52\%$ , with the out-of-sample claims. Also the reader may observe from Table 18 that the improvement is mostly due to the better marginal model fit, instead of the dependence modeling.

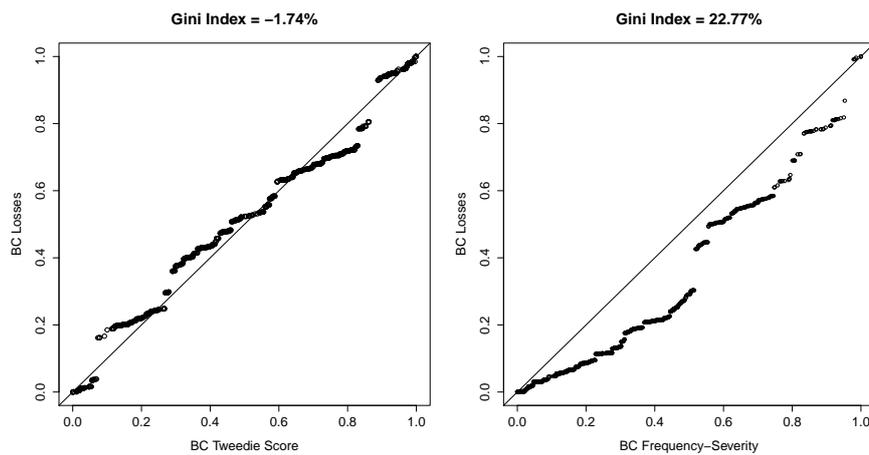


Figure 4. Ordered Lorenz Curves for BC.

### 7. Concluding Remarks

This study shows standard procedures for dependence modeling for a multiple lines insurance company. We have demonstrated that sophisticated marginal models may improve claim score calculations, and further demonstrated how multivariate claim distributions may be estimated using copulas. Our study also verifies that dependence modeling has little influence on the claim scores, but rather is a potentially useful tool for assessing risk measures of liabilities when losses are dependent on one another. A potentially interesting study would be to analyze the difference in risk measures associated with the correlation in liabilities carried by a multiple lines insurer, with and without dependence modeling. We leave this study for future work.

An interesting question is how to predict claims that are large relative to the rest of the distribution. For example, when simulating the GB2 distribution, we regularly generated large predicted values (that our software converted into “infinite values”). This suggests that more sophisticated consideration of the upper limits (possibly due to policy limits of the LGPIF) may be necessary to model the claim severities using long-tail distributions.

**Acknowledgments:** This work was partially funded by a Society of Actuaries CAE Grant. The first author acknowledges support from the University of Wisconsin-Madison’s Hickman-Larson Chair in Actuarial Science. The authors are grateful to two reviewers for insightful comments leading to an improved article.

**Author Contributions:** All authors contributed substantially to this work.

**Conflicts of Interest:** The authors declare no conflict of interest.

### A. Appendix

#### A.1. Alternative Way of Choosing Location Parameters for GB2

An alternative way to choose the location parameter in GB2 is through log linear model in [61]. The density of  $GB2(\sigma, \mu, \alpha_1, \alpha_2)$  is

$$f(y; \mu, \sigma, \alpha_1, \alpha_2) = \frac{[\exp(z)]^{\alpha_1}}{y\sigma B(\alpha_1, \alpha_2)[1 + \exp(z)]^{\alpha_1 + \alpha_2}}$$

where  $z = \frac{\ln(y) - \mu}{\sigma}$ .

As pointed out in [62], if

$$Y \sim GB2(\sigma, \mu, \alpha_1, \alpha_2),$$

$$\log(Y) = \mu + \sigma(\log \alpha_1 - \log \alpha_2) + \sigma \log F(2\alpha_1, 2\alpha_2).$$

This is actually the log linear model used with errors following the log-F distribution. Thus  $\mu + \sigma(\log \alpha_1 - \log \alpha_2)$  can be used as the location parameter associated with covariates.

As a special case of GB2, the location parameter of GG can be derived based on GB2. The density of GG( $a, b, \alpha_1$ ) is

$$GG(y; a, b, \alpha_1) = \frac{a}{\Gamma(\alpha_1)y} (y/b)^{a\alpha_1} e^{-(y/b)^a}.$$

Reparametrizing the GB2( $a, b, \alpha_1, \alpha_2$ ) with  $a = \frac{1}{\sigma}$ ,  $b = \exp(\mu)$ , we have

$$GG(a, b, \alpha_1) = \lim_{\alpha_2 \rightarrow \infty} GB2(a, b\alpha_2^{1/a}, \alpha_1, \alpha_2).$$

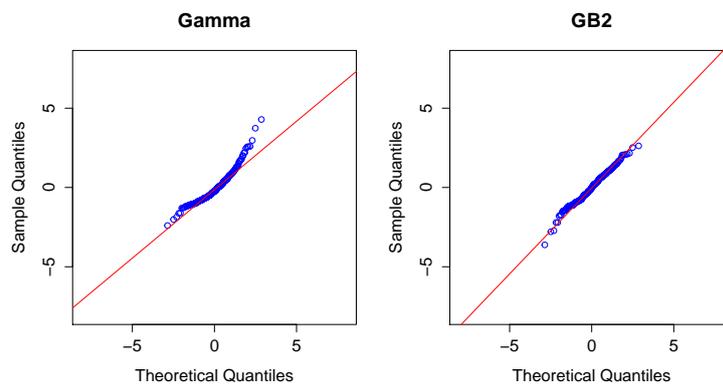
The location parameter for GG( $a, b, \alpha_1$ ) should be  $\log(b) + \sigma \log(\alpha_2) + \sigma(\log(\alpha_1) - \log(\alpha_2)) = \log(b) + \sigma \log(\alpha_1)$ . This is consistent with the results in [63].

When  $a = 1$ , the GG distribution becomes the gamma distribution with shape parameter  $\alpha_1$  and scale parameter  $b$ .  $\log(b) + \log(\alpha_1)$  is the location parameter, which is the log-mean of the gamma distribution, and is hence consistent with the GLM framework.

### A.2. Other Lines

#### A.2.1. IM (Contractor’s Equipment)

The property fund uses IM as a symbol to denote contractor’s equipment, and we follow this notation. Figure A1 shows the residual plot of severity fitted with gamma and GB2 in the IM line. Based on the plot, the GB2 is chosen.



**Figure A1.** QQ Plot for Residuals of Gamma and GB2 Distribution for contractor’s equipment (IM).

Table A1 shows the expected count for each frequency value under different models and empirical values from the data. The proportion of 1 s for the IM line is not high, and hence most models were able to capture this.

**Table A1.** Comparison between Empirical Values and Expected Values for IM line.

	Empical	ZeroinflPoisson	ZeroonePoisson	Poisson	NB	ZeroinflNB	ZerooneNB
0	4386	4381.660	4383.572	4351.363	4383.527	4384.718	4384.736
1	182	189.282	184.517	233.111	191.214	188.278	188.252
2	40	35.986	37.159	29.992	31.187	32.558	32.560
3	6	10.383	11.428	5.794	9.313	9.716	9.719
4	4	3.237	3.662	1.311	3.555	3.668	3.669
5	2	1.009	1.155	0.324	1.548	1.564	1.565
6	2	0.311	0.357	0.081	0.740	0.724	0.724
≥7	0	0.132	0.151	0.024	0.889	0.763	0.764
0 proportion	0.949	0.948	0.948	0.941	0.948	0.949	0.949
1 proportion	0.039	0.041	0.040	0.050	0.041	0.041	0.041

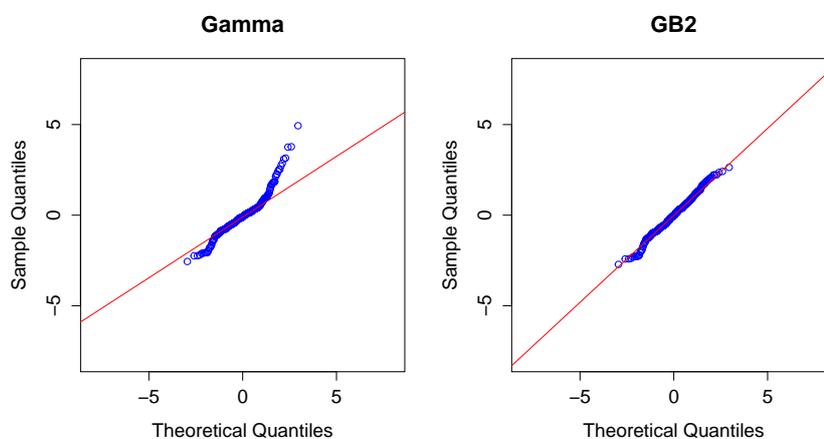
**Table A2.** Goodness of Fit Statistics for IM Line.

<b>ZeroinflPoisson</b>	<b>ZeroonePoisson</b>	<b>Poisson</b>	<b>NB</b>	<b>ZeroinflNB</b>	<b>ZerooneNB</b>
13.046	11.204	74.788	7.335	6.497	6.493

Table A2 shows goodness of fit tests result. It was calculated using the results in Table A1. The parsimonious model, negative binomial, is preferred.

A.2.2. PN (Comprehensive New)

The property fund uses P for comprehensive, and N to denote new vehicles. Hence PN would mean comprehensive coverage for new vehicles. Figure A2 shows the residual plot of severity, fitted with gamma and GB2 for PN. Based on the plot, the GB2 is chosen for the PN line.



**Figure A2.** QQ Plot for Residuals of Gamma and GB2 Distribution for comprehensive new (PN).

Table A3 shows the expected count for each frequency value under different models, and the empirical values from the data.

**Table A3.** Comparison between Empirical Values and Expected Values for PN Line.

	<b>Empical</b>	<b>ZeroinflPoisson</b>	<b>ZeroonePoisson</b>	<b>Poisson</b>	<b>NB</b>	<b>ZeroinflNB</b>	<b>ZerooneNB</b>
0	1323	1318.274	1310.359	1260.404	1324.292	1327.832	1323.999
1	154	116.320	161.105	153.445	148.051	141.023	149.927
2	50	53.333	28.759	78.837	52.433	53.831	50.552
3	33	49.295	30.636	62.798	32.670	33.472	32.031
4	19	41.465	32.238	41.674	22.728	23.348	22.648
5	16	28.661	28.095	22.970	16.145	16.650	16.341
6	13	16.614	20.560	10.853	11.546	11.920	11.827
7	7	8.282	12.953	4.510	8.291	8.537	8.557
8	4	3.623	7.169	1.683	5.972	6.111	6.185
9	4	1.413	3.540	0.573	4.314	4.371	4.466
10	3	0.497	1.579	0.180	3.124	3.124	3.221
11	1	0.159	0.643	0.053	2.267	2.232	2.320
12	2	0.047	0.241	0.015	1.649	1.593	1.670
13	4	0.013	0.084	0.004	1.202	1.137	1.201
14	2	0.003	0.027	0.001	0.878	0.811	0.863
15	1	0.001	0.008	0	0.643	0.578	0.620
16	1	0	0.002	0	0.471	0.412	0.445
≥17	1	0	0.001	0	0.788	0.651	0.712
0 proportion	0.808	0.805	0.800	0.769	0.809	0.811	0.809
1 proportion	0.094	0.071	0.098	0.094	0.090	0.086	0.092

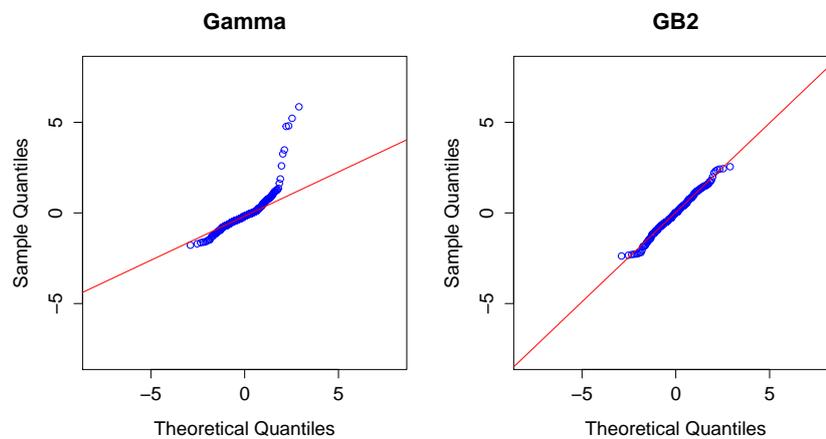
Table A4 shows goodness of fit tests result. It was calculated using the results in Table A3. The simpler model, negative binomial, is preferred.

**Table A4.** Goodness of Fit Statistics for PN Line.

ZeroinflPoisson	ZeroonePoisson	Poisson	NB	ZeroinflNB	ZerooneNB
31,776.507	2113.085	93,179.199	11.609	14.537	11.853

A.2.3. PO (Comprehensive Old)

The property fund uses symbol O to denote old, hence PO would be comprehensive coverage for old vehicles. Figure A3 shows the residual plot of severity fitted with gamma and GB2, for the PO line.



**Figure A3.** QQ Plot for Residuals of Gamma and GB2 Distribution for comprehensive old (PO).

Table A5 shows the expected count for each frequency value under different models and empirical values from the data.

**Table A5.** Comparison between Empirical Values and Expected Values for PO Line.

	Empical	ZeroinflPoisson	ZeroonePoisson	Poisson	NB	ZeroinflNB	ZerooneNB
0	1875	1873.859	1867.234	1811.952	1879.908	1880.884	1879.944
1	155	121.646	153.552	180.597	142.801	139.962	144.827
2	42	54.712	34.118	76.153	45.728	45.669	43.381
3	26	38.944	28.731	40.240	24.579	24.945	23.977
4	12	24.618	21.921	18.234	15.068	15.469	15.010
5	8	13.332	14.536	7.152	9.671	10.002	9.780
6	7	6.366	8.598	2.512	6.361	6.609	6.509
7	4	2.757	4.662	0.814	4.256	4.434	4.397
8	2	1.109	2.380	0.248	2.886	3.010	3.006
9	2	0.421	1.173	0.072	1.979	2.065	2.076
10	1	0.153	0.570	0.020	1.371	1.429	1.447
11	1	0.054	0.276	0.005	0.958	0.998	1.017
≥12	3	0.026	0.250	0.002	2.339	2.426	2.521
0 proportion	0.877	0.876	0.873	0.847	0.879	0.880	0.879
1 proportion	0.072	0.057	0.072	0.084	0.067	0.065	0.068

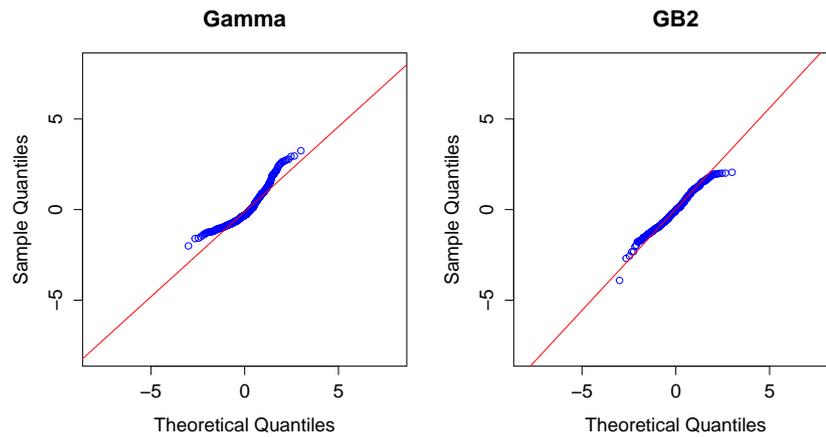
Table A6 shows goodness of fit tests result. It was calculated using the results in Table A5. Negative binomial model is selected, based on the test results.

**Table A6.** Goodness of Fit Statistics for PO Line.

ZeroinflPoisson	ZeroonePoisson	Poisson	NB	ZeroinflNB	ZerooneNB
387.671	43.127	5365.758	2.995	3.824	2.512

A.2.4. CN (Collision New)

Figure A4 shows the residual plot of severity fitted with gamma and GB2 for the CN line.



**Figure A4.** QQ Plot for Residuals of Gamma and GB2 Distribution for collision new (CN).

Table A7 shows the expected count for each frequency value under different models, and the empirical values from the data.

**Table A7.** Comparison between Empirical Values and Expected Values for CN Line.

	Empical	ZeroinflPoisson	ZeroonePoisson	Poisson	NB	ZeroinflNB	ZerooneNB
0	1159	1157.476	1153.549	1090.320	1168.332	1169.317	1163.521
1	228	201.586	229.228	274.394	210.934	209.282	226.545
2	74	79.485	56.894	95.634	69.231	69.270	60.183
3	26	43.915	36.167	40.251	33.137	33.325	30.364
4	16	24.613	23.909	17.118	18.478	18.644	17.630
5	9	12.536	14.538	6.987	10.951	11.066	10.846
6	3	5.718	7.937	2.723	6.677	6.749	6.866
7	3	2.352	3.904	1.018	4.137	4.178	4.422
8	4	0.881	1.745	0.366	2.589	2.612	2.881
9	1	0.303	0.716	0.127	1.633	1.645	1.895
10	0	0.097	0.272	0.042	1.036	1.043	1.255
11	3	0.029	0.096	0.014	0.661	0.664	0.837
12	0	0.008	0.032	0.004	0.424	0.425	0.561
13	1	0.002	0.010	0.001	0.273	0.273	0.378
14	1	0.001	0.003	0	0.176	0.176	0.256
15	1	0	0.001	0	0.114	0.114	0.174
≥16	0	0	0	0	0.176	0.175	0.296
0 proportion	0.758	0.757	0.754	0.713	0.764	0.765	0.761
1 proportion	0.149	0.132	0.150	0.179	0.138	0.137	0.148

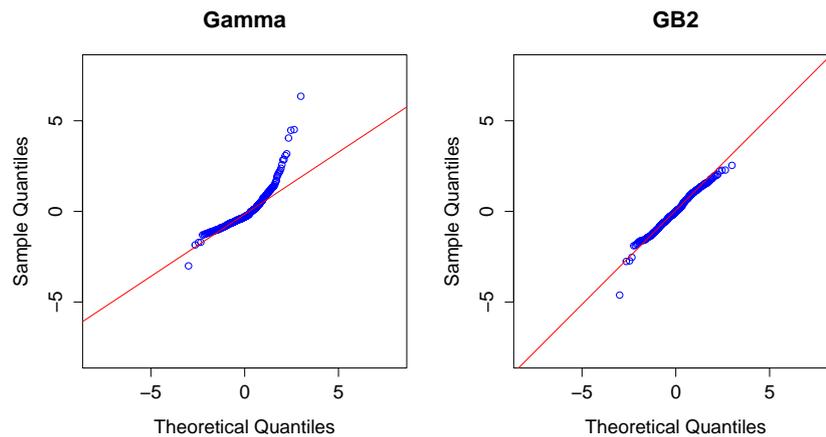
Table A8 shows the goodness of fit tests result. It was calculated using the values in Table A7. The parsimonious model, negative binomial, is selected.

**Table A8.** Goodness of Fit Statistics for CN Line.

ZeroinflPoisson	ZeroonePoisson	Poisson	NB	ZeroinflNB	ZerooneNB
10,932.035	1791.868	15,221.056	29.911	30.378	22.574

A.2.5. CO (Collision, Old)

Figure A5 shows the residual plot of severity fitted with Gamma and GB2 for the CO line. GB2 is preferred. Note, here  $\frac{1}{\sigma}$  instead of  $\sigma$  is fitted for computational stability.



**Figure A5.** QQ Plot for Residuals of Gamma and GB2 Distribution for collision old (CO).

Table A9 shows the expected count for each frequency value under different models, and the empirical values from the data.

**Table A9.** Comparison between Empirical Values and Expected Values for CO Line.

	Empical	ZeroinflPoisson	ZeroonePoisson	Poisson	NB	ZeroinflNB	ZerooneNB
0	1651	1649.151	1647.342	1600.173	1654.590	1656.293	1653.325
1	224	197.854	220.281	262.576	218.632	212.240	220.052
2	63	85.200	63.848	84.408	66.972	70.829	67.211
3	34	42.854	37.325	37.334	30.322	32.286	31.127
4	22	21.236	21.693	16.911	16.287	16.989	16.606
5	5	9.765	11.729	7.164	9.512	9.603	9.511
6	2	4.137	5.815	2.811	5.824	5.650	5.671
7	5	1.647	2.658	1.040	3.673	3.408	3.468
8	3	0.645	1.144	0.373	2.365	2.091	2.158
9	3	0.265	0.488	0.133	1.547	1.301	1.361
≥10	1	0.230	0.516	0.076	2.495	1.883	2.024
0 proportion	0.820	0.819	0.818	0.795	0.822	0.823	0.822
1 proportion	0.111	0.098	0.109	0.130	0.109	0.105	0.109

Table A10 shows the goodness of fit tests result. It was calculated using the results in Table A9. The parsimonious model, negative binomial, is selected.

**Table A10.** Goodness of Fit Statistics for CO Line.

ZeroinflPoisson	ZeroonePoisson	Poisson	NB	ZeroinflNB	ZerooneNB
60.691	25.206	121.987	10.387	11.440	10.370

A.3. Tweedie Margins

Table A11 shows marginal coefficients for each line.

Table A11. Marginal Coefficients of Tweedie Model.

Variable Name	BC		IM		PN				
	Estimate	Standard Error	Estimate	Standard Error	Estimate	Standard Error			
(Intercept)	5.855	0.969	***	8.404	1.081	***	6.284	0.437	***
lnCoverage	0.758	0.155	***	1.022	0.134	***	0.395	0.107	***
lnDeduct	0.147	0.148		-0.277	0.154	.			
NoClaimCredit	-0.272	0.371		-0.330	0.244		-0.570	0.296	.
EntityType: City	0.264	0.574		0.223	0.406		0.930	0.497	.
EntityType: County	0.204	0.719		0.671	0.501		2.550	0.462	***
EntityType: Misc	-0.380	0.729		-1.945	1.098	.	-0.010	0.942	
EntityType: School	0.072	0.521		-0.340	0.520		0.036	0.474	
EntityType: Town	0.940	0.658		-0.487	0.476		0.185	0.586	
$\phi$	165.814			849.530			376.190		
$P$	1.669			1.461			1.418		

Variable Name	PO		CN		CO				
	Estimate	Standard Error	Estimate	Standard Error	Estimate	Standard Error			
(Intercept)	5.868	0.489	***	8.263	0.294	***	7.889	0.340	***
lnCoverage	0.860	0.119	***	0.474	0.098	***	0.841	0.117	***
lnDeduct									
NoClaimCredit	0.155	0.319		-0.369	0.253		-1.025	0.331	**
EntityType: City	0.747	0.612		0.169	0.347		-0.723	0.540	
EntityType: County	1.414	0.577	*	1.112	0.325	***	0.863	0.434	*
EntityType: Misc	0.033	0.925		-0.596	0.744		-0.579	0.939	
EntityType: School	0.989	0.544	.	-0.631	0.316	*	0.477	0.399	
EntityType: Town	-2.482	1.123	*	-1.537	0.499	**	-0.628	0.564	
$\phi$	322.662			336.297			302.556		
$P$	1.508			1.467			1.527		

Notes:  $\phi$ : dispersion parameter,  $P$ : power parameter,  $1 < P < 2$ . Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1

Figure A6 shows the cdf plot of jittered aggregate losses, as described in Section 3.7. For most lines, the plots do not show a uniform trend. This tells us that the Tweedie model may not be ideal for such cases.

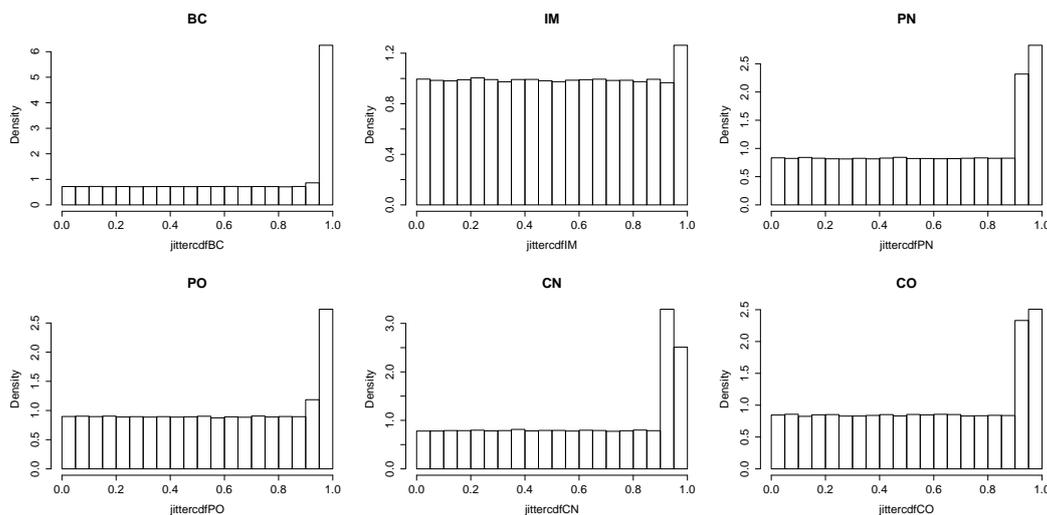


Figure A6. Jittering Plot of Tweedie.

A.4. Dependence of Frequency and Severity

To motivate this problem, let us think about a classical “aggregate loss” model insurance. In this model, we have a count random variable  $N$  representing the number of claims of a policyholder.

The positive claim amounts are denoted as  $Y_1, Y_2, \dots$  which are independent and identically distributed. During a specified time interval, the policyholder incurs  $N$  claims  $Y_1, \dots, Y_N$ , that sum to

$$S = \sum_{j=1}^N Y_j,$$

known as the *aggregate loss*. In the classical model, the claims frequency distribution  $N$  is assumed independent of the amount distribution  $Y$ .

#### A.4.1. Moments

The random variable  $S$  is said to have a “compound distribution.” Its moments are readily computable in terms of the frequency and severity moments. Using the law of iterated expectations, we have

$$E S = E (E(S|N)) = E (\mu_Y N) = \mu_Y \mu_N$$

and

$$\begin{aligned} E S^2 &= E \left\{ E(S^2|N) \right\} = E \left\{ N E(Y^2) + N(N-1) \mu_Y^2 \right\} \\ &= \mu_N (\sigma_Y^2 + \mu_Y^2) + (\sigma_N^2 + \mu_N^2 - \mu_N) \mu_Y^2 \\ &= \mu_N \sigma_Y^2 + (\sigma_N^2 + \mu_N^2) \mu_Y^2 \end{aligned}$$

so

$$\begin{aligned} \text{Var } S &= \mu_N \sigma_Y^2 + \sigma_N^2 \mu_Y^2 \\ &= \text{Poisson } \mu_N E(Y^2). \end{aligned}$$

#### A.4.2. Average Severity

These are calculations basic to the actuarial curriculum. Less common is an expression for the average severity, defined as  $\bar{S} = S/N$ . Note that when  $N = 0$ , we define  $\bar{S} = 0$ . Also, use the notation  $p_n = \Pr(N = n)$ . Again, using the rule of iterated expectations, we have

$$\begin{aligned} E \bar{S} &= 0 \times p_0 + \sum_{n=1}^{\infty} \left( E \left\{ \frac{1}{n} \sum_{i=1}^n Y_i \right\} | N = n \right) p_n \\ &= \sum_{n=1}^{\infty} \{ \mu_Y \} p_n \\ &= \mu_Y (1 - p_0) \end{aligned}$$

and

$$\begin{aligned} E \bar{S}^2 &= 0^2 \times p_0 + \sum_{n=1}^{\infty} \frac{1}{n^2} \left( n E Y^2 + n(n-1) \mu_Y^2 \right) p_n \\ &= \sum_{n=1}^{\infty} \left( \mu_Y^2 + \frac{E Y^2 - \mu_Y^2}{n} \right) p_n \\ &= \mu_Y^2 (1 - p_0) + \sigma_Y^2 \sum_{n=1}^{\infty} \frac{p_n}{n}. \end{aligned}$$

Thus,

$$\text{Var } \bar{S} = p_0(1 - p_0)\mu_Y^2 + \sigma_Y^2 \sum_{n=1}^{\infty} \frac{p_n}{n}.$$

For zero truncated distributions with  $p_0 = 0$ , we have  $E \bar{S} = \mu_Y$  and  $\text{Var } \bar{S} = \sigma_Y^2 E \frac{1}{N}$ .

#### A.4.3. Correlation and Dependence

The random variables  $S$  and  $N$  are clearly related. To demonstrate this, we have

$$\begin{aligned} \text{Cov}(S, N) &= E \{E(SN|N)\} - E(S)E(N) \\ &= E \left\{ \mu_Y N^2 \right\} - \mu_Y \mu_N^2 \\ &= \mu_Y \left\{ \sigma_N^2 + \mu_N^2 \right\} - \mu_Y \mu_N^2 \\ &= \mu_Y \sigma_N^2 \geq 0. \end{aligned}$$

However, the case for the average severity and frequency is not so clear.

$$\begin{aligned} \text{Cov}(\bar{S}, N) &= E(S) - E(\bar{S})E(N) \\ &= \mu_Y \mu_N - (1 - p_0)\mu_Y \mu_N \\ &= p_0 \mu_Y \mu_N. \end{aligned}$$

Thus,  $\bar{S}$  and  $N$  are uncorrelated when  $p_0 = 0$ .

Are  $\bar{S}$  and  $N$  independent when  $p_0 = 0$ ? Basic calculations show that this is not the case. To this end, define the  $n$ -fold convolution of the  $Y$  random variables,  $F^{*n}(x) = \Pr(Y_1 + \dots + Y_n \leq x)$ . With this notation, we have

$$\Pr(\bar{S} \leq s | N = n) = \Pr(Y_1 + \dots + Y_n \leq sn | N = n) = F^{*n}(sn)$$

and

$$\begin{aligned} \Pr(\bar{S} \leq s) &= \sum_{n=0}^{\infty} \Pr(\bar{S} \leq s | N = n) p_n = \sum_{n=0}^{\infty} F^{*n}(sn) p_n \\ &\neq F^{*n}(sn) = \Pr(\bar{S} \leq s | N = n). \end{aligned}$$

So,  $\bar{S}$  and  $N$  provide a nice example of two random variables that are uncorrelated but not independent.

#### References

1. Frees, E.W. Frequency and severity models. In *Predictive Modeling Applications in Actuarial Science*; Frees, E.W., Meyers, G., Derrig, R.A., Eds.; Cambridge University Press: Cambridge, UK, 2014.
2. Klugman, S.A.; Panjer, H.H.; Willmot, G.E. *Loss Models: From Data to Decisions*; John Wiley & Sons: Hoboken, New Jersey, 2012; Volume 715.
3. Frees, E.W. Analytics of insurance markets. *Annu. Rev. Financ. Econ.* **2015**, *7*. Available online: <http://www.annualreviews.org/loi/financial> (accessed on 20 February 2016).
4. Frees, E.W.; Jin, X.; Lin, X. Actuarial applications of multivariate two-part regression models. *Ann. Actuar. Sci.* **2013**, *7*, 258–287.
5. Frees, E.W. *Longitudinal and Panel Data: Analysis and Applications in the Social Sciences*; Cambridge University Press: New York, NY, USA, 2004.
6. Frees, E.W.; Sun, Y. Household life insurance demand: A multivariate two-part model. *N. Am. Actuar. J.* **2010**, *14*, 338–354.

7. Brockett, P.L.; Golden, L.L.; Guillén, M.; Nielsen, J.P.; Parner, J.; Pérez-Marín, A.M. Survival analysis of a household portfolio of insurance policies: How much time do you have to stop total customer defection? *J. Risk Insur.* **2008**, *75*, 713–737.
8. Guillén, M.; Nielsen, J.P.; Pérez-Marín, A.M. The need to monitor customer loyalty and business risk in the European insurance industry. *Geneva Pap. Risk Insur.-Issues Pract.* **2008**, *33*, 207–218.
9. De Jong, P.; Heller, G.Z. *Generalized Linear Models for Insurance Data*; Cambridge University Press: Cambridge, UK, 2008.
10. Guillén, M. Regression with categorical dependent variables. In *Predictive Modeling Applications in Actuarial Science*; Frees, E.W., Meyers, G., Derrig, R.A., Eds.; Cambridge University Press: Cambridge, UK, 2014.
11. Boucher, J.P. Regression with count dependent variables. In *Predictive Modeling Applications in Actuarial Science*; Frees, E.W., Meyers, G., Derrig, R.A., Eds.; Cambridge University Press: Cambridge, UK, 2014.
12. Mildenhall, S.J. A systematic relationship between minimum bias and generalized linear models. *Proc. Casualty Actuar. Soc.* **1999**, *86*, 393–487.
13. Shi, P. Fat-tailed regression models. In *Predictive Modeling Applications in Actuarial Science*; Frees, E.W., Meyers, G., Derrig, R.A., Eds.; Cambridge University Press: Cambridge, UK, 2014.
14. Sun, J.; Frees, E.W.; Rosenberg, M.A. Heavy-tailed longitudinal data modeling using copulas. *Insur. Math. Econ.* **2008**, *42*, 817–830.
15. Smyth, G.K. Generalized linear models with varying dispersion. *J. R. Stat. Soc. Ser. B* **1989**, *51*, 47–60.
16. Meester, S.G.; Mackay, J. A parametric model for cluster correlated categorical data. *Biometrics* **1994**, *50*, 954–963.
17. Lambert, P. Modelling irregularly sampled profiles of non-negative dog triglyceride responses under different distributional assumptions. *Stat. Med.* **1996**, *15*, 1695–1708.
18. Song, X.K.P. Multivariate dispersion models generated from Gaussian copula. *Scand. J. Stat.* **2000**, *27*, 305–320.
19. Frees, E.W.; Wang, P. Credibility using copulas. *N. Am. Actuar. J.* **2005**, *9*, 31–48.
20. Song, X.K. *Correlated Data Analysis: Modeling, Analytics, and Applications*; Springer Science & Business Media: New York, NY, USA, 2007.
21. Kolev, N.; Paiva, D. Copula-based regression models: A survey. *J. Stat. Plan. Inference* **2009**, *139*, 3847–3856.
22. Sklar, M. *Fonctions de Répartition À N Dimensions et Leurs Marges*; Université Paris 8: Paris, France, 1959.
23. Nelsen, R.B. *An Introduction to Copulas*; Springer Science & Business Media: New York, NY, USA, 1999; Volume 139.
24. Frees, E.W.; Valdez, E.A. Understanding relationships using copulas. *N. Am. Actuar. J.* **1998**, *2*, 1–25.
25. Joe, H. *Dependence Modelling with Copulas*; CRC Press: Boca Raton, FL, USA, 2014.
26. Patton, A.J. Modelling asymmetric exchange rate dependence\*. *Int. Econ. Rev.* **2006**, *47*, 527–556.
27. Acar, E.F.; Craiu, R.V.; Yao, F. Dependence calibration in conditional copulas: A nonparametric approach. *Biometrics* **2011**, *67*, 445–453.
28. Joe, H. *Multivariate Models and Multivariate Dependence Concepts*; CRC Press: Boca Raton, FL, USA, 1997.
29. Song, P.X.K.; Li, M.; Zhang, P. Vector generalized linear models: A Gaussian copula approach. In *Copulae in Mathematical and Quantitative Finance*; Springer: New York, NY, USA, 2013; pp. 251–276.
30. Nikoloulopoulos, A.K. Copula-based models for multivariate discrete response data. In *Copulae in Mathematical and Quantitative Finance*; Springer: New York, NY, USA, 2013; pp. 231–249.
31. Genest, C.; Nešlehová, J. A primer on copulas for count data. *Astin Bull.* **2007**, *37*, 475–515.
32. Nikoloulopoulos, A.K.; Karlis, D. Regression in a copula model for bivariate count data. *J. Appl. Stat.* **2010**, *37*, 1555–1568.
33. Nikoloulopoulos, A.K. On the estimation of normal copula discrete regression models using the continuous extension and simulated likelihood. *J. Stat. Plan. Inference* **2013**, *143*, 1923–1937.
34. Genest, C.; Nikoloulopoulos, A.K.; Rivest, L.P.; Fortin, M. Predicting dependent binary outcomes through logistic regressions and meta-elliptical copulas. *Braz. J. Probab. Stat.* **2013**, *27*, 265–284.
35. Panagiotelis, A.; Czado, C.; Joe, H. Pair copula constructions for multivariate discrete data. *J. Am. Stat. Assoc.* **2012**, *107*, 1063–1072.
36. Shi, P. Insurance ratemaking using a copula-based multivariate Tweedie model. *Scand. Actuar. J.* **2016**, *2016*, 198–215.

37. SAS (Statistical Analysis System) Institute Incorporated. *SAS/STAT 9.2 User's Guide, Second Edition*; SAS Institute Inc.: 2010. Available online: <http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#titlepage.htm> (accessed on 30 September 2015).
38. Frees, E.W.; Wang, P. Copula credibility for aggregate loss models. *Insur. Math. Econ.* **2006**, *38*, 360–373.
39. Shi, P. Multivariate longitudinal modeling of insurance company expenses. *Insur. Math. Econ.* **2012**, *51*, 204–215.
40. Genest, C.; Boies, J.C. Detecting dependence with kendall plots. *J. Am. Stat. Assoc.* **2003**, *57*, 275–284.
41. Genest, C.; Rivest, L.P. Statistical inference procedures for bivariate archimedean copulas. *J. Am. Stat. Assoc.* **1993**, *88*, 1034–1043.
42. Kojadinovic, I.; Segers, J.; Yan, J. Large-sample tests of extreme value dependence for multivariate copulas. *Can. J. Stat.* **2011**, *39*, 703–720.
43. Genest, C.; Kojadinovic, I.; Neslehova, J.; Yan, J. A goodness-of-fit test for bivariate extreme value copulas. *Bernoulli* **2011**, *17*, 253–275.
44. Bahraoui, Z.; Bolance, C.; Perez-Marin, A.M. Testing extreme value copulas to estimate the quantile. *SORT* **2014**, *38*, 89–102.
45. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed.; Springer Science & Business Media: New York, NY, USA, 2009.
46. Czado, C.; Kastenmeier, R.; Brechmann, E.C.; Min, A. A mixed copula model for insurance claims and claim sizes. *Scand. Actuar. J.* **2012**, *2012*, 278–305.
47. Cox, D.R.; Snell, E.J. A general definition of residuals. *J. R. Stat. Soc. Ser. B* **1968**, *30*, 248–275.
48. Rüschenendorf, L. On the distributional transform, Sklar's theorem, and the empirical copula process. *J. Stat. Plan. Inference* **2009**, *139*, 3921–3927.
49. Vuong, Q.H. Likelihood ratio tests for model selection and non-nested hypotheses. *Econom. J. Econom. Soc.* **1989**, *57*, 307–333.
50. Frees, E.W.; Meyers, G.; Cummings, A.D. Summarizing insurance scores using a Gini index. *J. Am. Stat. Assoc.* **2011**, *106*, doi:10.1198/jasa.2011.tm10506.
51. Frees, E.W.; Gao, J.; Rosenberg, M.A. Predicting the frequency and amount of health care expenditures. *N. Am. Actuar. J.* **2011**, *15*, 377–392.
52. Gschlößl, S.; Czado, C. Spatial modelling of claim frequency and claim size in non-life insurance. *Scand. Actuar. J.* **2007**, *2007*, 202–225.
53. Song, P.X.K.; Fan, Y.; Kalbfleisch, J.D. Maximization by parts in likelihood inference. *J. Am. Stat. Assoc.* **2005**, *100*, 1145–1158.
54. Krämer, N.; Brechmann, E.C.; Silvestrini, D.; Czado, C. Total loss estimation using copula-based regression models. *Insur. Math. Econ.* **2013**, *53*, 829–839.
55. Shi, P.; Feng, X.; Ivantsova, A. Dependent frequency-severity modeling of insurance claims. *Insur. Math. Econ.* **2015**, *64*, 417–428.
56. Hua, L. Tail negative dependence and its applications for aggregate loss modeling. *Insur. Math. Econ.* **2015**, *61*, 135–145.
57. Brechmann, E.; Czado, C.; Paterlini, S. Flexible dependence modeling of operational risk losses and its impact on total capital requirements. *J. Bank. Financ.* **2014**, *40*, 271–285.
58. Li, J.; Zhu, X.; Chen, J.; Gao, L.; Feng, J.; Wu, D.; Sun, X. Operational risk aggregation across business lines based on frequency dependence and loss dependence. *Math. Probl. Eng.* **2014**, *2014*. Available online: <http://www.hindawi.com/journals/mpe/2014/404208/> (accessed on 20 February 2016).
59. Frees, E.W.; Lee, G. Rating endorsements using generalized linear models. *Variance* **2015**. Available online: <http://www.variancejournal.org/issues/> (accessed on 20 February 2016).
60. Local Government Property Insurance Fund. *Board of Regents of the University of Wisconsin System* **2011**. Website: <https://sites.google.com/a/wisc.edu/local-government-property-insurance-fund> (accessed on 20 February 2016).
61. Prentice, R.L. Discrimination among some parametric models. *Biometrika* **1975**, *62*, 607–614.

62. Yang, X. Multivariate Long-Tailed Regression with New Copulas. Ph.D. Thesis, University of Wisconsin-Madison, Madison, WI, USA, 2011.
63. Prentice, R.L. A log gamma model and its maximum likelihood estimation. *Biometrika* **1974**, *61*, 539–544.



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).