The Annals of Applied Statistics
2016, Vol. 10, No. 2, 834–863
DOI: 10.1214/16-AOAS914

MULTILEVEL MODELING OF INSURANCE CLAIMS USING COPULAS

BY PENG SHI[*,1], XIAOPING FENG[*] AND JEAN-PHILIPPE BOUCHER[†]

University of Wisconsin-Madison[*] and Université du Québec à Montréal[†]

In property-casualty insurance, claims management is featured with the modeling of a semi-continuous insurance cost associated with individual risk transfer. This practice is further complicated by the multilevel structure of the insurance claims data, where a contract often contains a group of policyholders, each policyholder is insured under multiple types of coverage, and the contract is repeatedly observed over time. The data hierarchy introduces a complex dependence structure among claims and leads to diversification in the insurer's liability portfolio.

To capture the unique features of policy-level insurance costs, we propose a copula regression for the multivariate longitudinal claims. In the model, the Tweedie double generalized linear model is employed to examine the semi-continuous claim cost of each coverage type, and a Gaussian copula is specified to accommodate the cross-sectional and temporal dependence among the multilevel claims. Estimation and inference is based on the composite likelihood approach and the properties of parameter estimates are investigated through simulation studies. When applied to a portfolio of personal automobile policies from a Canadian insurer, we show that the proposed copula model provides valuable insights to an insurer's claims management process.

1. Introduction and motivation.

General insurance (a.k.a. "nonlife," a.k.a. "property-casualty") protects individuals and organizations from financial losses due to property damage or legal liabilities. It allows policyholders to exchange the risk of a large loss for the certainty of smaller periodic payments of premiums. Insurers allocates the bulk of premium dollars into investment and claims payments. As it is for an insurer to manage its investment portfolio, it is equally important for the insurer to manage its claim portfolio. Claim management is the counterpart of asset management for the claims on the insurer's book.

Claim management is the analytics of insurance costs. It requires applying statistical techniques in the analysis and interpretation of the claims data. In the data-driven industry of general insurance, claim management provides useful insights for insurers to make better business decisions. For instance, analytics helps insurers in identifying risk characteristics for risk screening in underwriting, managing

Received June 2015; revised December 2015.

[1] Supported in part by the Centers of Actuarial Excellence (CAE) Research Grant from the Society of Actuaries and the Fall Research Competition from University of Wisconsin-Madison.

Key words and phrases. Composite likelihood, insurance claims, longitudinal data, multivariate regression, property-casualty insurance, Tweedie distribution.

claim costs and allocating resources for claims handling, refining the classification ratemaking system, as well as understanding excess layers for reinsurance and retention.

The central piece of claim management is claims modeling. In this article, we provide a general framework to look into the process of modeling and estimating insurance cost with a complex structure. It is well known that the insurance cost associated with individual risk transfer presents a unique semi-continuous feature where a significant fraction of zeros is incorporated into an otherwise positive continuous outcome. The portion of zeros corresponds to no claims and the positive component corresponds to the amount of claims. Two strategies are commonly used by practitioners to analyze claim distributions: the two-part approach [see, e.g., Frees (2014)] and the pure premium approach [see, e.g., Jørgensen and Paes de Souza (1994)]. The former decomposes claims cost into a frequency and a severity component, while the latter uses the Tweedie distribution to accommodate the mass probability at zero. Each method has its own strengths and weaknesses. In addition to the statistical considerations, the selection between the two approaches often depends on the types of data available and the preference of the analyst.

Beyond their mixed character, risk- or policy-level general insurance losses are also distinctive in that they can be viewed as the sum of losses from multiple hazard or coverage types. For example, a personal automobile insurance policy could provide both liability and collision coverage. This bundling design complicates the process of claims modeling. Insurers, on the one hand, must analyze claims separately by coverage type both because of the different contract features specific to each coverage type and because predictive dimensions generally relate differently to the various coverage types. On the other hand, insurers want to analyze the multiple types of claims jointly because they are interrelated. The first effort in this line of study is due to Frees and Valdez (2008) and Frees, Shi and Valdez (2009) where the authors extended the frequency-severity model to a three-component framework to incorporate claim type.

Complex design of modern insurance products brings new challenges in modeling insurance costs. One of them is the multilevel structure often encountered in property-casualty insurance, where a contract contains a group of policyholders, each policyholder is insured under multiple types of coverage, and the contract is repeatedly observed over time. For instance, a commercial automobile insurance policy covers both bodily injury and property damage for a fleet of vehicles, a worker's compensation contract provides indemnity cost and medical care payment for all employees of an organization, and an employment-based group health insurance compensates costs of medical care utilization for office-based visits, hospital stays and emergency room usage. The data hierarchy introduces a complex dependence structure among claims and leads to diversification in the insurer's liability portfolio. In our study, the claims data are from personal automobile insurance in Ontario, Canada. An insurance policy provides coverage for the motor vehicles in a household. The number of vehicles per household ranges from one to four and each vehicle is insured under four types of coverage, that is, accident benefit, civil liability, collision and all risk. The portfolio is observed over a 4-year period, from 2003 to 2006. In claims modeling, one expects to capture the cluster effects (household), the cross-sectional dependence among multiple claim types, as well as the serial correlation in the longitudinal context.

Motivated by the above observations, this article further advances the claims modeling in property-casualty insurance. To capture the unique features of policy-level insurance costs, we propose a copula regression for the multivariate longitudinal claims. Specifically, for the claims cost of each type, we consider using the Tweedie distribution to accommodate the massive zeros. In the Tweedie distribution, we perform regression on both mean and dispersion using the double generalized linear model framework [Jørgensen (1987)]. In the insurance claims data, all available predictors are at the risk level, such as primary owner and vehicle characteristics. We allow the set of covariates to vary by claim type.

The multilevel structure of claims are accommodated using dependence models. We use a Gaussian copula to join the mixed outcome of claim costs. Refer to Nelsen (2006) for an introduction and Joe (2015) for recent development on copulas. For our purpose, we specify three sources of dependence: the correlation among claims from multiple vehicles within the same household, the cross-sectional dependence among multiple types of claims, and the temporal association for the longitudinal claim cost of each type. These explicit relations and their implied association are specified in the dispersion matrix of the Gaussian copula, and the dependence parameters are readily interpretable. We show that the proposed dependence model has a direct link with the mixed linear model on transformed data. Another important feature of our data is the lack of balance. The unbalanced claim costs could be due to the difference in the number of vehicles of a household, type of coverage for a vehicle or length of observation period. The Gaussian copula provides flexibility in this sense, assuming that the "missing" observations are ignorable.

Because of the mixed nature of claim costs, estimation of the Gaussian copula model using the full maximum likelihood involves multidimensional integration. For a household with four cars with each being covered by a comprehensive policy (four types of coverage), a four-year period of observation means a $4 \times 4 \times 4 = 64$ dimensional integration. As a solution, we resort to the composite likelihood method for model estimation and comparison [see Varin, Reid and Firth (2011) for an overview]. Before fitting the model to the insurance data, we investigate the finite sample properties of parameter estimates using simulation studies. Using the Gaussian copula and composite likelihood, statistical efficiency is sacrificed to gain the computational advantage and interpretability of the dependence parameters.

In the application of the personal automobile insurance, we examine the claims distribution at both individual and portfolio levels. At the individual level, we demonstrate basic ratemaking and claims triage under a simplified risk classification system. At the portfolio level, we emphasize the importance of dependence modeling and its implications on an insurer's risk management practice. We show that the central limit theorem collapses when aggregating correlated risks in the portfolio.

Section 2 describes the automobile insurance claims dataset and its important characteristics that motivate the multilevel modeling framework. Section 3 proposes the statistical model and discusses the inference based on a composite likelihood method. The specification of the dependence structure in the model is detailed in the Appendix. Section 4 investigates the finite sample properties of parameter estimates using simulated data. In Section 5 we fit the model to the real data and show its implications on the insurer's claim management. Concluding remarks are provided in Section 6.

2. Automobile insurance data.

We examine an insurance claims dataset of personal automobile insurance obtained from a property-casualty insurer in Canada. The data represent the insurer's book of business written in the province of Ontario over the period 2003–2006. Both public and private insurance programs coexist in Canada. Ontario uses a private insurance system. The industry is made up of more than 100 private companies that are overseen by the government agency Financial Services Commission of Ontario. Contrary to the public system, private insurance has more incentives to use advanced actuarial approach and refined risk classification in underwriting and ratemaking. This emphasizes the importance of the statistical analysis in our study.

As in most developed countries, automobile insurance is required for all motorists and is enforced by Ontario law. An insurance contract could provide four types of coverage: (1) "accident benefit" provides the insured with medical care payments and income replacement benefits if injured in an automobile accident, regardless of who caused the accident. (2) "all risk" covers the damage to the insured's vehicle caused by hazards other than collision, such as fire, theft and hail etc. (3) "civil liability" is a combined bodily injury and property damage coverage. It pays claims if the insured is liable for the bodily injury or property damage of a third party. (4) "collision" covers the financial losses when an insured vehicle is involved in a collision with another object, including another vehicle. Coverages (1) and (3) are compulsory and are included in the standard policy. Coverages (2) and (4) are optional and available through the comprehensive policy. Policyholders of standard and comprehensive policies often show distinct driving behavior due to different risk levels and incentives, known as information asymmetry in the economics literature [see, e.g., Shi, Zhang and Valdez (2012)]. To provide focus, we limit our analysis to the comprehensive policy, and our final sample contains 87,670 policies after some screening in the preliminary analysis.

One interesting feature of the data is its multilevel structure. The level-one unit is the insurance policy and the level-two unit is the insured vehicle. In personal

TABLE 1
Distribution of the number of insured vehicles per policy

Number of vehicles	**1**	**2**	**3**	**4**	**Total**
Frequency	77,352	10,058	253	7	87,670
Percentage	88.23	11.47	0.289	0.01	100

automobile insurance, it is common that a single policy is purchased to insure all vehicles within the same household. The distribution of the number of insured vehicles per policy is summarized in Table 1. About 12% of policies in our data insure more than one vehicle, among which the majority insure two vehicles, and it is rare for a policy to insure more than three vehicles. This percentage is lower than the actual number of households owning multiple cars. Consider a household with three cars, two of them are insured under a standard policy and the other one is insured under a comprehensive policy. Only the vehicle in the comprehensive policy is retained in the sample and the two vehicles in the standard policy are removed for our study. Because the insurance database only contains policy ID, we do not even know that these three vehicles are from the same household.

The outcome variable of our interest is the insurance claims cost. The four types of claims indicate the multivariate nature of the data. We examine insurance claim cost by coverage type and look into the vector of claims cost; Figure 1 displays their distributions. The upper panel shows the violin plots using data in 2003 [see Hintze and Nelson (1998) for details on violin plot]. One noticeable feature is the semi-continuity, where the large number of zeros correspond to no claims. In our data, this probability is about 91% regardless of coverage type. Another observation is the long tails in the individual claims cost. This is more pronounced in the liability coverage partly due to the large legal defense cost. Data in all years exhibit consistent properties. The longitudinal nature indicates another hierarchy in the multilevel data. The lower panel shows the average insurance cost over time. The accident benefit coverage shows a higher variation, but in general we observe a relatively stable pattern. In our application, one can think of the average cost as the pure premium for the insurance contract. The premium shows a wide range across coverage type, with civil liability and all risk being the most and least expensive coverages respectively. This relation is also true for different risk levels as shown in the data analysis. The different distributional features shown in Figure 1 motivate the insurer to analyze claims data by coverage type.

The insurance data also contains a set of predictors that could explain the variation in claims cost. It is a common practice for property-casualty insurers using indicators in the risk classification system. Hence, all predictors available are binary. Table 2 summarizes the description of these predictors and their sample averages by year. Three broad categories of covariates are commonly believed to affect

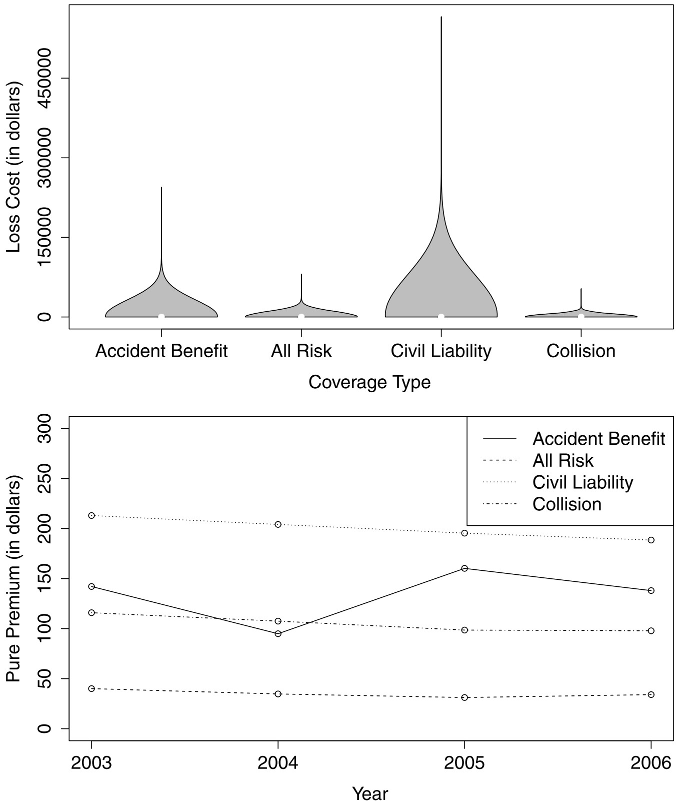

FIG. 1. *Distributions of claims cost by coverage type. The upper panel shows the violin plot and the lower panel shows the average cost over time.*

insurance cost: (1) Policyholder's characteristics. Our data contains indicators on the driver's age, marital status and whether he/she is a homeowner. Because of the nonlinear age effect, we differentiate young drivers and senior citizens. (2) Driving history. Years of experience and conviction history are used in the analysis. (3) Vehicle's characteristics. Vehicle age is an indicator of ownership at purchase. The purpose of the car indicates whether it is a lease vehicle and whether it is used for business. The usage of the vehicle is measured by the mileage driven and the number of drivers. For a vehicle with multiple drivers, the driver's characteristics correspond to the primary driver. As anticipated, the driver's characteristics show larger variation, while the vehicle's characteristics are very consistent over time.

All the covariates obtained from the insurer are dichotomized. First, insurers discretize continuous predictors to simplify the risk classification and to capture potential nonlinear effects. Second, regulations often limit the way predictors can be used in the ratemaking.

TABLE 2
Sample mean of predictors by year

Variable	Description	2003	2004	2005	2006
Young	= 1 if age between 16 and 25	2.77	2.23	1.84	1.60
Senior	= 1 if age more than 60	15.18	16.68	18.27	19.97
Marital	= 1 if married	71.92	72.85	73.62	73.95
Homeowner	= 1 if homeowner	42.15	63.81	76.67	80.42
Experience	= 1 if more than ten years of experience	90.14	91.58	92.97	93.68
Conviction	= 1 if positive number of convictions	8.86	6.24	2.99	1.89
Newcar	= 1 if new car	89.52	89.56	89.59	89.67
Leasecar	= 1 if lease car	15.39	15.31	15.35	15.15
Business	= 1 if business use	3.43	3.68	3.85	3.96
Highmilage	= 1 if drive more than 10,000 miles	72.97	71.52	69.53	67.73
Multidriver	= 1 if more than two drivers	3.42	5.18	7.26	9.38

3. Modeling.

3.1. *Multivariate Tweedie model.* Consider an insurance portfolio consisting of N policies. For the ith $(= 1, \dots, N)$ policy, let K_i denote the number of vehicles, J_i the number of coverage types and T_i the number of observation periods. Let y_{ikjt} denote the insurance cost of coverage type j in the tth period for the kth vehicle in policy i. The quantity of interest is the vector of claims defined as $\mathbf{y}_i = (y_{ikjt})_{k=1,\dots,K_i, j=1,\dots,J_i, t=1,\dots,T_i}$.

Note that y_{ikjt} follows a mixed distribution in that it consists of a discrete mass at zero and a positive continuous component. We consider the Tweedie distribution that has non-negative support and can have a positive probability at zero [Tweedie (1984)]. With appropriate parameterization, the Tweedie distribution can be shown as a member of the exponential dispersion family [Jørgensen (1987)], with the density function given by

$$f(y; \mu, p, \phi) = \exp\left[\frac{1}{\phi}\left(\frac{-y}{(p-1)\mu^{p-1}} - \frac{\mu^{2-p}}{2-p}\right) + S(y; \phi)\right],$$

where

$$(1)\quad S(y; \phi) = \begin{cases} 0, & \text{if } y = 0, \\ \ln \sum_{n\geq 1} \left\{ \dfrac{(1/\phi)^{1/(p-1)} y^{(2-p)/(p-1)}}{(2-p)(p-1)^{(2-p)/(p-1)}} \right\}^n \dfrac{1}{n!\Gamma(n(2-p)/(p-1))y}, & \\ \quad \text{if } y > 0. & \end{cases}$$

With this parameterization, mean and variance of the Tweedie random variable are μ and $\phi\mu^p$, respectively, where ϕ is the dispersion parameter and p is the power parameter that controls the variance of the distribution. This result is rather appealing because it suggests that the theories of generalized linear models are ready to apply [McCullagh and Nelder (1989)].

The Tweedie distribution becomes a Poisson distribution when $p = 1$ and a gamma distribution when $p = 2$. The more interesting range of p for our application is between 1 and 2. In this case, the Tweedie random variable can be generated from a Poisson sum of gamma random variables [Smyth (1996)]. From $p = 1$ to $p = 2$, the Tweedie distribution gradually loses its mass at zero as it shifts from a Poisson distribution to a gamma distribution. The compound Poisson presentation also provides a nature interpretation for insurance claims modeling. One can think of the claims cost per year for a policyholder as sum of a series of independent gamma random variables and the number of claims in a year as a Poisson random variable.

Denote the density and cumulative distribution functions of y_{ikjt} as $f_j(y_{ikjt})$ and $F_j(y_{ikjt})$, respectively. To allow for covariates, we employ the double generalized linear model to perform regression analysis on both mean and dispersion of the Tweedie outcome. When modeling the cost of insurance claims, dispersion modeling is necessary, as it increases the precision of prediction [Smyth and Jørgensen (2002)]. Define $f_j(y_{ikjt}) = f(y_{ikjt}; \mu_{ikjt}, p_j, \phi_{ikjt})$. With log link functions, we specify

$$g_\mu(\mu_{ikjt}) = \log(\mu_{ikjt}) = \mathbf{x}'_{ikjt}\boldsymbol{\beta}_j,$$

$$g_\phi(\phi_{ikjt}) = \log(\phi_{ikjt}) = \mathbf{z}'_{ikjt}\boldsymbol{\gamma}_j.$$

Here $\mathbf{x}_{ikjt}$ and $\mathbf{z}_{ikjt}$ are vectors of covariates in the mean and dispersion regression, respectively, and $\boldsymbol{\beta}_j$ and $\boldsymbol{\gamma}_j$ are the associated regression coefficients. We allow for different sets of covariates for the mean and dispersion, and because of the distributional differences in the coverage types as shown in Section 2, we allow parameters $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and p to depend on the claim type j.

Another commonly used technique to incorporate mass zeros into an otherwise continuous distribution is the censored regression. The classical example is the Tobit model [Tobin (1958)]. Relaxing the normality assumption, recent literature also proposes censored regression with a Student-t distribution [see, e.g., Arellano-Valle et al. (2012)] and scales mixtures of normal distributions [see, e.g., Castro et al. (2014) and Garay et al. (2016)]. We choose the Tweedie model over censored regression because the Tweedie distribution is a Poisson sum of gamma random variables, and thus is more in line with the frequency-severity modeling framework commonly used in practice [Frees (2014)].

The multilevel structure of the insurance data is accommodated using dependence models. We use a parametric copula function to model the complex dependence embedded in the vector of claims cost. To simplify the presentation, we relabel $\mathbf{y}_i = (\tilde{y}_{i1}, \dots, \tilde{y}_{im_i})$, where $m_i = K_i \times J_i \times T_i$, denoting the total number of observations for policy i. Then the cumulative distribution function of $\mathbf{y}_i$ can be expressed in terms of a copula function H_i, that is,

$$G(\mathbf{y}_i) = H_i\bigl(F(\tilde{y}_{i1}), \dots, F(\tilde{y}_{im_i})\bigr), \tag{2}$$

where F is the cumulative distribution function associated with (1). Note that $\mathbf{y}_i$ is a vector of mixed random variables. Without loss of generality, assume that the first q_i components $(\tilde{y}_{i1}, \dots, \tilde{y}_{iq_i})$ are continuous and the rest of the $m_i - q_i$ components $(\tilde{y}_{iq_i+1}, \dots, \tilde{y}_{im_i})$ are discrete. The density function of $\mathbf{y}_i$ is shown as

$$g(\mathbf{y}_i) = \prod_{l=1}^{q_i} f(y_l) h_i^{q_i}\bigl(F(\tilde{y}_{i1}), \dots, F(\tilde{y}_{im_i})\bigr), \tag{3}$$

where

$$h_i^{q_i}(w_1, \dots, w_{m_i}) = \frac{\partial^{q_i}}{\partial w_1 \cdots \partial w_{q_i}} H_i(w_1, \dots, w_{m_i}).$$

Let $m = \max\{m_1, \dots, m_N\}$. We consider the Gaussian copula with the distributional function given by

$$H(w_1, \dots, w_m; \boldsymbol{\Sigma}) = \Phi_m\bigl(\Phi^{-1}(w_1), \dots, \Phi^{-1}(w_m); \boldsymbol{\Sigma}\bigr),$$

where Φ_m and Φ denote the distributional function of an m-variate normal with zero mean and correlation matrix $\boldsymbol{\Sigma}$ and the standard univariate normal respectively. It can be shown that [see, e.g., Song, Li and Yuan (2009) and Shi (2016)]

$$\begin{aligned} h^q&(w_1, \dots, w_m; \boldsymbol{\Sigma}) \\ &= (2\pi)^{-(m-q)/2} |\boldsymbol{\Sigma}|^{-1/2} \\ &\quad \times \int_{-\infty}^{\Phi^{-1}(w_{q+1})} \cdots \int_{-\infty}^{\Phi^{-1}(w_m)} \exp\left\{\frac{1}{2}(\mathbf{s}_1', \mathbf{s}_2')\boldsymbol{\Sigma}^{-1}(\mathbf{s}_1', \mathbf{s}_2')' - \frac{1}{2}\mathbf{s}_1'\mathbf{s}_1\right\} d\mathbf{s}_2. \end{aligned}$$

With the Gaussian copula, the lack of balance can be easily addressed using the subclass of H and h^q. That is, for policy i, we specify $H_i(\cdot) = H(\cdot; \mathbf{A}_i \boldsymbol{\Sigma} \mathbf{A}_i')$ and $h_i^{q_i}(\cdot) = h^q(\cdot; \mathbf{A}_i \boldsymbol{\Sigma} \mathbf{A}_i')$. Here $\mathbf{A}_i = [\iota_1, \dots, \iota_{m_i}, \mathbf{0}, \dots, \mathbf{0}]_{m_i \times m}$ and ι_r is a column vector with the rth element being 1 and 0 otherwise, and $\mathbf{0}$ is a column vector of zeros.

The dependency among the vector of claims cost is captured by the correlation matrix $\boldsymbol{\Sigma}$ in the Gaussian copula. In our context, one wants to accommodate three types of association, the correlation among vehicles insured under the same policy, the dependence among multiple types of claims for a given vehicle, and the temporal relationship for a particular type of coverage. To achieve these purposes, we specify $\boldsymbol{\Sigma} = \mathbf{B}_{K\times K} \otimes \mathbf{P}_{(TJ)\times(TJ)}$, where $\otimes$ denotes the Kronecker product, and

$$\mathbf{B}_{K\times K} = \begin{pmatrix} 1 & \delta & \cdots & \delta \\ \delta & 1 & \cdots & \delta \\ \vdots & \vdots & \ddots & \vdots \\ \delta & \delta & \cdots & 1 \end{pmatrix},$$

$$\mathbf{P}_{(TJ)\times(TJ)} = \begin{pmatrix} \sigma_{11}\mathbf{P}_{11} & \sigma_{12}\mathbf{P}_{12} & \cdots & \sigma_{1J}\mathbf{P}_{1J} \\ \sigma_{21}\mathbf{P}_{21} & \sigma_{22}\mathbf{P}_{22} & \cdots & \sigma_{2J}\mathbf{P}_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{J1}\mathbf{P}_{J1} & \sigma_{J2}\mathbf{P}_{J2} & \cdots & \sigma_{JJ}\mathbf{P}_{JJ} \end{pmatrix}.$$

The cluster effect is captured by an exchangeable correlation $\mathbf{B}_{K\times K}$ that is implied by the household-specific random effect. The dependence due to the multivariate longitudinal observations for a given vehicle is captured by $\mathbf{P}_{(TJ)\times(TJ)}$, where $\sigma_{jj'} = \sigma_{j'j}$ and $\mathbf{P}_{jj'} = \mathbf{P}_{j'j}$. This is a commonly used specification in models of several time series [see, e.g., Greene (2007)]. Here $\sigma_{jj'}$ represents the cross-sectional correlation between coverage type j and j' in the same time period, known as the concurrent or contemporaneous correlation coefficient in time series analysis. $\mathbf{P}_{jj}$ is the serial correlation for the insurance costs of coverage j. $\mathbf{P}_{jj'}$ $(j \neq j')$ is the correlation across coverage types j and j'. Note that this matrix is in general not symmetric. The diagonal elements are ones and the off-diagonal elements indicate the lead-lag relationship between component series. Extending the method in Parks (1967), we specify the concurrent correlation $\sigma_{jj'}$ and serial correlation $\mathbf{P}_{jj}$, and let the lag correlation $\mathbf{P}_{jj'}$ be determined implicitly. With the AR(1) serial correlation, we have

$$\mathbf{P}_{jj'} = \begin{pmatrix} 1 & \rho_{j'} & \cdots & \rho_{j'}^{T-1} \\ \rho_j & 1 & \cdots & \rho_{j'}^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_j^{T-1} & \rho_j^{T-2} & \cdots & 1 \end{pmatrix},$$

$$\sigma_{jj'} = \begin{cases} 1, & \text{if } j = j', \\ \dfrac{\tau_{jj'}\sqrt{1-\rho_j^2}\sqrt{1-\rho_{j'}^2}}{1-\rho_j\rho_{j'}}, & \text{if } j \neq j'. \end{cases}$$

We detail the specification of $\boldsymbol{\Sigma}$ and establish its connection to the linear model on transformed data in the Appendix.

3.2. *Inference*. For inference purposes, we employ the composite likelihood method [Lindsay (1988)]. Because of the mixed nature of the insurance cost, the likelihood function of model (3) involves multidimensional integration. In our application, an insurance contract covering four vehicles would imply a 64-dimensional integration. Thus, full maximum likelihood estimation is computationally challenging and the computational difficulty increases as the number of time periods becomes larger. To minimize the computational burden, we use the pairwise likelihood [Cox and Reid (2004)], which provides relatively high estimation efficiency as shown in the literature. See Varin (2008, 2011) for reviews on the composite likelihood approach.

On another note, the trade-off between the computational challenge and the efficiency loss using the composite likelihood method is due to the estimation of the probability mass function of the Gaussian copula. One alternative strategy could be to explore a more flexible dependence modeling approach such as the pairwise copula construction based on vines [see, e.g., Aas et al. (2009), Smith et al. (2010) and Panagiotelis, Czado and Joe (2012)]. However, we find the Gaussian copula is particularly useful in our application in that it is ready to apply to the unbalanced data and the dependence parameters have intuitive interpretations. Considering the applied nature of this work, we make sacrifices to balance the interpretability, complexity and computation of the model.

The pairwise composite likelihood function for policy i is defined as

$$l_i(\boldsymbol{\theta}; \mathbf{y}_i) = \sum_{k=1}^{K_i}\left(\sum_{j=1}^{J_i}\sum_{t<t'}\ell(\boldsymbol{\theta}; y_{ikjt}, y_{ikjt'}) + \sum_{j<j'}\sum_{t,t'=1}^{T_i}\ell(\boldsymbol{\theta}; y_{ikjt}, y_{ikj't'})\right)$$

$$+\sum_{k<k'}\sum_{j,j'=1}^{J_i}\sum_{t,t'=1}^{T_i}\ell(\boldsymbol{\theta}; y_{ikjt}, y_{ik'j't'}),$$

where $\ell(\boldsymbol{\theta}; y_{ikjt}, y_{ik'j't'}) = \log(L(\boldsymbol{\theta}; y_{ikjt}, y_{ik'j't'}))$ and

$$L(\boldsymbol{\theta}; y_{ikjt}, y_{ik'j't'})$$
$$= \begin{cases} H\bigl(F_j(y_{ikjt}), F_{j'}(y_{ik'j't'}); \tilde{\rho}_{kjtk'j't'}\bigr), \\ \qquad \text{if } y_{ikjt} = 0 \text{ and } y_{ik'j't'} = 0, \\ f_j(y_{ikjt})h_1\bigl(F_j(y_{ikjt}), F_{j'}(y_{ik'j't'}); \tilde{\rho}_{kjtk'j't'}\bigr), \\ \qquad \text{if } y_{ikjt} > 0 \text{ and } y_{ik'j't'} = 0, \\ f_{j'}(y_{ik'j't'})h_2\bigl(F_j(y_{ikjt}), F_{j'}(y_2); \tilde{\rho}_{kjtk'j't'}\bigr), \\ \qquad \text{if } y_{ikjt} = 0 \text{ and } y_{ik'j't'} > 0, \\ f_j(y_{ikjt})f_{j'}(y_{ik'j't'})h\bigl(F_j(y_{ikjt}), F_{j'}(y_{ik'j't'}); \tilde{\rho}_{kjtk'j't'}\bigr), \\ \qquad \text{if } y_{ikjt} > 0 \text{ and } y_{ik'j't'} > 0, \end{cases}$$

with $\tilde{\rho}_{kjtk'j't'} = \delta^{\mathcal{I}(k\neq k')}\sigma_{jj'}^{\mathcal{I}(j\neq j')}\rho_{j'}^{\mathcal{I}(t<t')|t-t'|}\rho_{j}^{\mathcal{I}(t>t')|t-t'|}$. Then the total composite likelihood for the portfolio of policies can be expressed as

$$l(\boldsymbol{\theta}; \mathbf{y}) = \sum_{i=1}^{N}\frac{1}{m_i - 1}l_i(\boldsymbol{\theta}; \mathbf{y}_i), \tag{4}$$

where $1/(m_i - 1)$ is the weight assigned for the ith policy [see, e.g., Zhao and Joe (2005) and Joe and Lee (2009)].

The composite likelihood estimator is defined as $\hat{\boldsymbol{\theta}}_N = \arg\max_{\boldsymbol{\theta}} l(\boldsymbol{\theta}; \mathbf{y})$. Denote the composite score function as $S_N(\boldsymbol{\theta}) = \partial l(\boldsymbol{\theta}; \mathbf{y})/\partial\boldsymbol{\theta}$. To estimate the variance of $\hat{\boldsymbol{\theta}}_N$, we use the Godambe information matrix [Godambe (1960)], defined as

$$\mathbf{G}_N(\boldsymbol{\theta}) = \mathbf{R}_N(\boldsymbol{\theta})\boldsymbol{\Omega}_N^{-1}(\boldsymbol{\theta})\mathbf{R}_N(\boldsymbol{\theta}), \tag{5}$$

where $\mathbf{R}_N(\boldsymbol{\theta}) = -E(\partial S_N(\boldsymbol{\theta})/\partial\boldsymbol{\theta}')$ and $\boldsymbol{\Omega}_N(\boldsymbol{\theta}) = \mathrm{Var}(S_N(\boldsymbol{\theta}))$. Under regularity conditions [Molenberghs and Verbeke (2005), pages 190–191] on the bivariate log-likelihood functions, we can apply the central limit theorem to the composite likelihood score statistic, leading to the result that the composite likelihood estimator, $\hat{\boldsymbol{\theta}}_N$, is asymptotically normally distributed when $N \to \infty$,

$$\sqrt{N}\mathbf{G}_N^{1/2}(\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{I}). \tag{6}$$

The sample estimate of sensitivity matrix $\mathbf{R}_N(\boldsymbol{\theta})$ is given by

$$\hat{\mathbf{R}}_N(\boldsymbol{\theta}) = -\frac{1}{N}\sum_{i=1}^{N}\frac{\partial^2 l_i(\boldsymbol{\theta};\mathbf{y}_i)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'},$$

and the numerical Hessian matrix is used to approximate the second order derivative. The sample estimate of variability matrix $\boldsymbol{\Omega}_N(\boldsymbol{\theta})$ is expressed by the outer product of the composite score functions as

$$\hat{\boldsymbol{\Omega}}_N(\boldsymbol{\theta}) = \frac{1}{N}\sum_{i=1}^{N}\frac{\partial l_i(\boldsymbol{\theta};\mathbf{y}_i)}{\partial\boldsymbol{\theta}}\frac{\partial l_i(\boldsymbol{\theta};\mathbf{y}_i)}{\partial\boldsymbol{\theta}'}.$$

Thus, the asymptotic covariance matrix can be approximated by $\hat{\mathbf{G}}_N^{-1}(\hat{\boldsymbol{\theta}}_N)/N$. Furthermore, model comparison is based on the composite likelihood version of AIC [Varin and Vidoni (2005)] and BIC [Gao and Song (2010)], which are respectively defined as

$$\begin{aligned}\mathrm{CLAIC} &= -2cl(\boldsymbol{\theta};\mathbf{y}) + 2\,\mathrm{tr}\big(\boldsymbol{\Omega}(\boldsymbol{\theta})\mathbf{R}^{-1}(\boldsymbol{\theta})\big),\\ \mathrm{CLBIC} &= -2cl(\boldsymbol{\theta};\mathbf{y}) + \log(N)\,\mathrm{tr}\big(\boldsymbol{\Omega}(\boldsymbol{\theta})\mathbf{R}^{-1}(\boldsymbol{\theta})\big).\end{aligned}$$

4. Numerical experiments. The properties of the composite likelihood estimates are investigated using simulated data. In the simulation, we set $J = 2$, $K = 2$ and $T = 4$, that is, a policy covers two vehicles and provides two types of coverage for each vehicle in a four-year period. Data are generated from the multivariate Tweedie model in Section 3.1. In the marginal distribution, we use Tweedie(μ_j, p_j, ϕ_j) with the following specification for coverage type $j =$ 1 and 2:

$$\mu_j = \exp(\beta_{j0} + \beta_{j1}X_1 + \beta_{j2}X_2), \qquad \phi_j = \exp(\gamma_{j0} + \gamma_{j1}X_1 + \gamma_{j2}X_2),$$

TABLE 3
The resulting association parameter of $\delta\tau\rho$ under different parameter scenarios

	$\boldsymbol{\tau = 0.15}$	$\boldsymbol{\tau = 0.55}$	$\boldsymbol{\tau = 0.95}$
		$\rho = 0.15$	
$\delta = 0.15$	0.003	0.012	0.021
$\delta = 0.55$	0.012	0.045	0.078
$\delta = 0.95$	0.021	0.078	0.135
		$\rho = 0.55$	
$\delta = 0.15$	0.012	0.045	0.078
$\delta = 0.55$	0.045	0.166	0.287
$\delta = 0.95$	0.078	0.287	0.496
		$\rho = 0.95$	
$\delta = 0.15$	0.021	0.078	0.135
$\delta = 0.55$	0.078	0.287	0.496
$\delta = 0.95$	0.135	0.496	0.857

where $X_1 \sim$ Bernoulli(0.5) and $X_2 \sim$ Bernoulli(0.6) independently. In the joint distribution, we use the Gaussian copula with the correlation matrix specified as

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \delta \\ \delta & 1 \end{pmatrix} \otimes \left(\begin{array}{cc} \begin{pmatrix} 1 & \rho_1 & \rho_1^2 & \rho_1^3 \\ \rho_1 & 1 & \rho_1 & \rho_1^2 \\ \rho_1^2 & \rho_1 & 1 & \rho_1 \\ \rho_1^3 & \rho_1^2 & \rho_1 & 1 \end{pmatrix} & \sigma_{12} \begin{pmatrix} 1 & \rho_2 & \rho_2^2 & \rho_2^3 \\ \rho_1 & 1 & \rho_2 & \rho_2^2 \\ \rho_1^2 & \rho_1 & 1 & \rho_2 \\ \rho_1^3 & \rho_1^2 & \rho_1 & 1 \end{pmatrix} \\ \sigma_{12} \begin{pmatrix} 1 & \rho_1 & \rho_1^2 & \rho_1^3 \\ \rho_2 & 1 & \rho_1 & \rho_1^2 \\ \rho_2^2 & \rho_2 & 1 & \rho_1 \\ \rho_2^3 & \rho_2^2 & \rho_2 & 1 \end{pmatrix} & \begin{pmatrix} 1 & \rho_2 & \rho_2^2 & \rho_2^3 \\ \rho_2 & 1 & \rho_2 & \rho_2^2 \\ \rho_2^2 & \rho_2 & 1 & \rho_2 \\ \rho_2^3 & \rho_2^2 & \rho_2 & 1 \end{pmatrix} \end{array} \right).$$

Here, $\sigma_{12} = \tau_{12}\sqrt{1-\rho_1^2}\sqrt{1-\rho_2^2}/(1-\rho_1\rho_2)$.

In the above specification, parameters ρ_1, ρ_2 and τ_{12} can be interpreted as correlation coefficient, and they are bounded between -1 and 1. Therefore, so is parameter σ_{12}. Parameter δ captures the within-cluster dependence and $0 < \delta < 1$ (see Appendix for details). To obtain some intuition about the resulting dependence, we consider the special case $\rho_1 = \rho_2 = \rho$ and $\tau_{12} = \tau$, and we report the resulting association parameter of $\delta \times \sigma_{12} \times \rho = \delta\tau\rho$ under different parameter scenarios in Table 3. The combination of the three values of each parameter allows for a wide range of dependence. The insights obtained from the table apply to the general parameter setting.

In the simulation, we examine three scenarios, weak dependence ($\delta = \tau = \rho = 0.15$), moderate dependence ($\delta = \tau = \rho = 0.55$) and strong dependence ($\delta = \tau = \rho = 0.95$). The true parameters and simulation results are displayed in Tables 4, 5

TABLE 4
Simulation for different sample sizes (number of policies) under weak dependence

	Estimate (mean)		**Relative bias**		**SD**		**SE**		**MSE**	
Parameter	$N = 200$	**500**	**200**	**500**	**200**	**500**	**200**	**500**	**200**	**500**
$\beta_{10} = 1$	0.932	0.978	−0.068	−0.022	0.295	0.220	0.302	0.190	0.092	0.049
$\beta_{11} = 1.5$	1.534	1.501	0.022	0.001	0.217	0.131	0.210	0.134	0.048	0.017
$\beta_{12} = 0.5$	0.529	0.511	0.058	0.022	0.295	0.202	0.297	0.187	0.088	0.041
$\beta_{20} = 1$	0.982	1.004	−0.018	0.004	0.212	0.136	0.220	0.141	0.045	0.019
$\beta_{21} = 0.5$	0.509	0.493	0.017	−0.013	0.221	0.131	0.218	0.138	0.049	0.017
$\beta_{22} = 2$	1.999	1.994	−0.001	−0.003	0.211	0.136	0.218	0.140	0.044	0.018
$p_1 = 1.2$	1.195	1.197	−0.004	−0.002	0.024	0.013	0.021	0.013	0.001	0.000
$p_2 = 1.4$	1.393	1.397	−0.005	−0.002	0.026	0.017	0.026	0.017	0.001	0.000
$\gamma_{10} = 5$	4.994	4.993	−0.001	−0.001	0.115	0.076	0.120	0.077	0.013	0.006
$\gamma_{11} = 1$	0.995	1.012	−0.005	0.012	0.098	0.061	0.096	0.060	0.010	0.004
$\gamma_{12} = -1$	−0.991	−0.994	−0.009	−0.006	0.115	0.079	0.120	0.078	0.013	0.006
$\gamma_{20} = 4$	3.989	4.000	−0.003	0.000	0.123	0.076	0.111	0.071	0.015	0.006
$\gamma_{21} = 0$	0.010	0.004	–	–	0.113	0.071	0.110	0.070	0.013	0.005
$\gamma_{22} = 1$	1.012	1.009	0.012	0.009	0.135	0.087	0.126	0.079	0.018	0.008
$\rho_1 = 0.15$	0.153	0.150	0.022	0.001	0.061	0.048	0.097	0.059	0.004	0.002
$\rho_2 = 0.15$	0.158	0.157	0.053	0.045	0.058	0.043	0.076	0.050	0.003	0.002
$\tau = 0.15$	0.147	0.145	−0.022	−0.034	0.082	0.049	0.079	0.050	0.007	0.002
$\delta = 0.15$	0.137	0.146	−0.088	−0.030	0.068	0.053	0.079	0.050	0.005	0.003

and 6, respectively. We consider different sample sizes (number of policies) N and report the results for $N = 200$ and 500. For each simulated sample, we estimate parameters by maximizing the composite likelihood function. The reported results are based on 100 replications.

Within each table, we first report the mean and standard deviation of the point estimates for each parameter. The average estimates are very close to the corresponding true parameters for both $N = 200$ and 500. We further confirm this relation by calculating the relative bias of the estimates. As expected, increasing sample size reduces the estimation bias, and when $N = 500$ the biases for most parameters are almost zero. Next, we examine the standard error of the estimator. In each replication, the standard error is estimated using the Godambe information matrix described in Section 3.2. Its average (denoted by SE in the table) is comparable with the nominal standard deviation (SD) of point estimates, indicating the accuracy of the uncertainty estimates. Finally, we report the mean squared error (MSE) of the parameter estimates. Consistent results are observed that a larger sample size leads to more accurate estimates and that estimates with less uncertainty can be obtained with a larger number of policies.

When comparing across tables with different degrees of dependence, one notices that, for regression parameters, stronger dependence generally increases both bias and uncertainty of the estimates for a given sample size. This is because the

TABLE 5
Simulation for different sample sizes (number of policies) under moderate dependence

	Estimate (mean)		**Relative bias**		**SD**		**SE**		**MSE**	
Parameter	$N = 200$	**500**	**200**	**500**	**200**	**500**	**200**	**500**	**200**	**500**
$\beta_{10} = 1$	0.905	0.966	−0.095	−0.034	0.335	0.226	0.331	0.209	0.121	0.052
$\beta_{11} = 1.5$	1.534	1.505	0.023	0.003	0.198	0.131	0.222	0.142	0.040	0.017
$\beta_{12} = 0.5$	0.552	0.517	0.105	0.034	0.293	0.178	0.296	0.188	0.089	0.032
$\beta_{20} = 1$	0.983	0.995	−0.017	−0.005	0.255	0.157	0.248	0.162	0.065	0.025
$\beta_{21} = 0.5$	0.504	0.498	0.008	−0.005	0.270	0.146	0.232	0.150	0.073	0.021
$\beta_{22} = 2$	2.002	1.989	0.001	−0.005	0.194	0.138	0.209	0.137	0.038	0.019
$p_1 = 1.2$	1.195	1.199	−0.004	−0.001	0.021	0.015	0.021	0.014	0.000	0.000
$p_2 = 1.4$	1.396	1.400	−0.003	0.000	0.023	0.018	0.027	0.017	0.001	0.000
$\gamma_{10} = 5$	4.994	4.993	−0.001	−0.001	0.138	0.079	0.119	0.078	0.019	0.006
$\gamma_{11} = 1$	1.014	1.014	0.014	0.014	0.089	0.060	0.095	0.061	0.008	0.004
$\gamma_{12} = -1$	−1.006	−1.000	0.006	0.000	0.126	0.078	0.117	0.077	0.016	0.006
$\gamma_{20} = 4$	3.995	4.001	−0.001	0.000	0.130	0.084	0.114	0.074	0.017	0.007
$\gamma_{21} = 0$	−0.009	0.009	–	–	0.120	0.073	0.112	0.072	0.014	0.005
$\gamma_{22} = 1$	1.011	0.994	0.011	−0.006	0.140	0.082	0.123	0.078	0.020	0.007
$\rho_1 = 0.55$	0.539	0.544	−0.020	−0.011	0.073	0.044	0.067	0.045	0.006	0.002
$\rho_2 = 0.55$	0.543	0.545	−0.013	−0.010	0.061	0.042	0.061	0.041	0.004	0.002
$\tau = 0.55$	0.550	0.541	−0.001	−0.017	0.074	0.045	0.067	0.046	0.005	0.002
$\delta = 0.55$	0.537	0.536	−0.024	−0.026	0.075	0.051	0.068	0.045	0.006	0.003

effective sample size is smaller when dependence is stronger. In contrast, for dependence parameters, stronger dependence indicates a higher signal to noise ratio, and thus decreases bias of the point estimate and the associated uncertainty estimate.

Finally, we investigate the robustness of the Gaussian copula model with respect to the copula misspecification. In principal, the proposed modeling framework and inference method applies to any elliptical copula. We only examine the t copula because elliptical copulas other than Gaussian and t are rarely used in applications. Specifically, we generate data from the t copula model and estimate with the Gaussian copula model. In the data-generating process, specifications for marginals and the dispersion matrix are the same as above. We fix the degrees of freedom (df) in the t copula at $\text{df} = 1$ and $\text{df} = 5$. The estimation results are exhibited in Table 7. One finds similar patterns as observed in Table 5 (the Gaussian copula). In addition, the mean squared errors of the estimates in both marginal and dependence parameters are small, indicating that the Gaussian copula is robust with respect to misspecification. As anticipated, higher estimation precision is achieved when the degrees of freedom in the t copula is larger, because the t copula is approaching the Gaussian copula. One notices that the estimation for $\text{df} = 5$ is already close to those in Table 5.

TABLE 6
Simulation for different sample sizes (number of policies) under strong dependence

	Estimate (mean)		**Relative bias**		**SD**		**SE**		**MSE**	
Parameter	**$N = 200$**	**500**	**200**	**500**	**200**	**500**	**200**	**500**	**200**	**500**
$\beta_{10} = 1$	0.800	0.907	−0.200	−0.093	0.619	0.302	0.486	0.309	0.423	0.100
$\beta_{11} = 1.5$	1.516	1.507	0.011	0.004	0.156	0.090	0.153	0.099	0.024	0.008
$\beta_{12} = 0.5$	0.660	0.549	0.320	0.098	0.436	0.203	0.307	0.192	0.215	0.044
$\beta_{20} = 1$	0.948	0.954	−0.052	−0.046	0.369	0.212	0.346	0.224	0.139	0.047
$\beta_{21} = 0.5$	0.515	0.500	0.031	0.000	0.184	0.104	0.164	0.104	0.034	0.011
$\beta_{22} = 2$	2.017	1.996	0.009	−0.002	0.131	0.071	0.115	0.077	0.018	0.005
$p_1 = 1.2$	1.192	1.199	−0.007	−0.001	0.025	0.016	0.023	0.016	0.001	0.000
$p_2 = 1.4$	1.392	1.397	−0.005	−0.002	0.034	0.020	0.028	0.020	0.001	0.000
$\gamma_{10} = 5$	4.976	4.991	−0.005	−0.002	0.178	0.097	0.121	0.081	0.032	0.009
$\gamma_{11} = 1$	1.021	0.999	0.021	−0.001	0.097	0.058	0.087	0.057	0.010	0.003
$\gamma_{12} = -1$	−0.989	−0.989	−0.011	−0.011	0.159	0.079	0.122	0.078	0.025	0.006
$\gamma_{20} = 4$	3.992	3.994	−0.002	−0.001	0.130	0.094	0.135	0.090	0.017	0.009
$\gamma_{21} = 0$	0.013	0.002	–	–	0.104	0.070	0.102	0.068	0.011	0.005
$\gamma_{22} = 1$	1.004	1.017	0.004	0.017	0.120	0.067	0.105	0.071	0.014	0.005
$\rho_1 = 0.95$	0.946	0.948	−0.004	−0.002	0.020	0.013	0.018	0.011	0.000	0.000
$\rho_2 = 0.95$	0.946	0.948	−0.004	−0.003	0.018	0.012	0.018	0.011	0.000	0.000
$\tau = 0.95$	0.956	0.952	0.007	0.002	0.021	0.016	0.022	0.014	0.000	0.000
$\delta = 0.95$	0.943	0.946	−0.007	−0.004	0.025	0.013	0.021	0.013	0.001	0.000

5. Application in risk analysis.

5.1. *Estimation results.* The proposed approach is applied to the portfolio of automobile insurance policies introduced in Section 2. The composite likelihood estimates are summarized in Table 8. In the Tweedie marginals, we fit log-linear models to both the mean and the dispersion for each type of claim. The set of covariates is allowed to vary by coverage type. Due to the relatively small number of predictors, a backward selection procedure is adopted to select the covariates in the marginal model. We start with all available covariates and remove the least important one in each iteration until the AIC statistic deteriorates. The orthogonal parametrization in the Tweedie distribution allows to select variables in the mean and dispersion models in a sequential manner. We report a more parsimonious model by retaining only those important predictors.

The effects of covariates on claim frequency and severity differ either in direction or size. Not surprisingly, we observe their significant effects on the mean as well as the dispersion of the Tweedie model. For instance, the length of driving experience shows a negative effect on the mean but a positive effect on the dispersion of claims cost regardless of the coverage type. There are common factors affecting all types of claims (such as senior and conviction in the mean, and homeowner in the dispersion). Their effects across coverage types are noticeably consistent in

TABLE 7
Robustness of Gaussian copula

	Estimate (mean)		**Relative bias**		**SD**		**SE**		**MSE**	
Parameter	$N = 200$	**500**	**200**	**500**	**200**	**500**	**200**	**500**	**200**	**500**
					df = 1					
$\beta_{10} = 1$	0.818	0.901	−0.182	−0.099	0.560	0.361	0.449	0.285	0.346	0.140
$\beta_{11} = 1.5$	1.514	1.519	0.010	0.013	0.230	0.142	0.197	0.127	0.053	0.020
$\beta_{12} = 0.5$	0.630	0.576	0.260	0.152	0.424	0.265	0.324	0.207	0.196	0.076
$\beta_{20} = 1$	0.903	0.959	−0.097	−0.041	0.342	0.243	0.310	0.200	0.126	0.061
$\beta_{21} = 0.5$	0.534	0.536	0.067	0.073	0.238	0.162	0.201	0.131	0.058	0.028
$\beta_{22} = 2$	2.011	2.009	0.005	0.004	0.180	0.109	0.165	0.107	0.033	0.012
$p_1 = 1.2$	1.187	1.190	−0.011	−0.008	0.028	0.018	0.024	0.016	0.001	0.000
$p_2 = 1.4$	1.380	1.390	−0.014	−0.007	0.033	0.021	0.032	0.022	0.002	0.001
$\gamma_{10} = 5$	4.980	4.987	−0.004	−0.003	0.153	0.103	0.137	0.092	0.024	0.011
$\gamma_{11} = 1$	1.029	1.015	0.029	0.015	0.102	0.066	0.095	0.062	0.011	0.005
$\gamma_{12} = -1$	−0.983	−0.987	−0.017	−0.013	0.157	0.107	0.138	0.090	0.025	0.012
$\gamma_{20} = 4$	3.963	3.970	−0.009	−0.008	0.161	0.098	0.136	0.091	0.027	0.010
$\gamma_{21} = 0$	0.035	0.013	–	–	0.115	0.080	0.115	0.076	0.015	0.007
$\gamma_{22} = 1$	1.041	1.024	0.041	0.024	0.125	0.081	0.125	0.081	0.017	0.007
$\rho_1 = 0.55$	0.858	0.856	0.559	0.557	0.041	0.024	0.031	0.022	0.096	0.094
$\rho_2 = 0.55$	0.853	0.854	0.550	0.552	0.037	0.023	0.030	0.021	0.093	0.093
$\tau = 0.55$	0.847	0.836	0.540	0.520	0.048	0.028	0.040	0.026	0.091	0.083
$\delta = 0.55$	0.834	0.833	0.516	0.514	0.042	0.028	0.036	0.025	0.082	0.081
					df = 5					
$\beta_{10} = 1$	0.851	0.946	−0.149	−0.054	0.415	0.238	0.377	0.240	0.195	0.060
$\beta_{11} = 1.5$	1.500	1.521	0.000	0.014	0.237	0.133	0.223	0.141	0.056	0.018
$\beta_{12} = 0.5$	0.615	0.529	0.230	0.057	0.343	0.192	0.305	0.193	0.131	0.038
$\beta_{20} = 1$	0.953	0.982	−0.047	−0.018	0.298	0.193	0.272	0.179	0.091	0.038
$\beta_{21} = 0.5$	0.501	0.494	0.001	−0.012	0.248	0.146	0.232	0.147	0.061	0.021
$\beta_{22} = 2$	2.036	2.006	0.018	0.003	0.222	0.143	0.196	0.130	0.051	0.020
$p_1 = 1.2$	1.189	1.195	−0.009	−0.005	0.025	0.015	0.022	0.015	0.001	0.000
$p_2 = 1.4$	1.393	1.396	−0.005	−0.003	0.031	0.020	0.027	0.018	0.001	0.000
$\gamma_{10} = 5$	4.982	4.987	−0.004	−0.003	0.147	0.094	0.123	0.079	0.022	0.009
$\gamma_{11} = 1$	1.007	1.008	0.007	0.008	0.101	0.064	0.096	0.061	0.010	0.004
$\gamma_{12} = -1$	−0.981	−0.982	−0.019	−0.018	0.137	0.089	0.119	0.077	0.019	0.008
$\gamma_{20} = 4$	3.986	3.993	−0.004	−0.002	0.111	0.079	0.120	0.078	0.013	0.006
$\gamma_{21} = 0$	0.014	0.003	–	–	0.119	0.073	0.113	0.072	0.014	0.005
$\gamma_{22} = 1$	1.019	1.007	0.019	0.007	0.133	0.081	0.122	0.079	0.018	0.007
$\rho_1 = 0.55$	0.690	0.697	0.254	0.268	0.060	0.044	0.055	0.041	0.023	0.024
$\rho_2 = 0.55$	0.686	0.685	0.248	0.246	0.062	0.040	0.052	0.037	0.022	0.020
$\tau = 0.55$	0.683	0.671	0.241	0.220	0.066	0.038	0.057	0.040	0.022	0.016
$\delta = 0.55$	0.671	0.672	0.220	0.222	0.061	0.043	0.059	0.039	0.018	0.017

TABLE 8

Composite likelihood estimates of the multilevel Tweedie model

	Accident benefit		All risk		Civil liability		Collision		Dependence model		
	Est.	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**	**Parameter**	**Est.**	**S.E.**
					Mean						
Intercept	5.732	0.127	3.485	0.092	5.337	0.063	4.467	0.066	ρ_1	0.163	0.022
Young	−0.911	0.221							ρ_2	0.051	0.013
Senior	−0.466	0.105	−0.467	0.054	−0.183	0.039	−0.142	0.038	ρ_3	0.099	0.015
Marital					−0.155	0.030	−0.102	0.030	ρ_4	0.101	0.011
Homeowner	−0.225	0.074	−0.163	0.038	0.065	0.028			τ_{12}	0.436	0.010
Experience	−0.324	0.115	−0.221	0.068	−0.227	0.042	−0.383	0.045	τ_{13}	0.049	0.018
Conviction	0.807	0.140	0.368	0.078	0.136	0.054	0.209	0.057	τ_{14}	0.641	0.009
Newcar			0.327	0.067	0.136	0.047	0.399	0.050	τ_{23}	0.013	0.012
Leasecar			0.544	0.048	0.170	0.034	0.428	0.034	τ_{24}	0.351	0.007
Business			0.261	0.091	0.342	0.064			τ_{34}	0.021	0.010
Highmilage	−0.578	0.075			0.082	0.029	0.194	0.031	δ	0.082	0.020
Multidriver					0.428	0.050	0.402	0.048	σ_{12}	0.434	–
p	1.703	0.005	1.631	0.003	1.577	0.003	1.440	0.003	σ_{13}	0.048	–
					Dispersion						
Intercept	7.107	0.050	6.505	0.053	6.330	0.031	6.751	0.025	σ_{23}	0.013	–
Young			−0.288	0.068					σ_{24}	0.350	–
Senior	0.151	0.048	0.098	0.026	0.187	0.019	−0.049	0.019	σ_{34}	0.021	–
Marital	0.174	0.036	−0.100	0.021			0.068	0.016			
Homeowner	−0.099	0.035	0.072	0.019	0.043	0.014	0.051	0.015			
Experience	0.489	0.047	−0.169	0.039	0.216	0.021	0.097	0.023			
Conviction					−0.095	0.027					
Newcar			−0.093	0.032	−0.149	0.023					
Leasecar					−0.094	0.017	−0.116	0.017			
Business											
Highmilage			−0.096	0.020							
Multidriver	−0.237	0.060					−0.184	0.025			

TABLE 9
Goodness-of-fit statistics for alternative dependence specification

Model	**Description**	**CLAIC**	**CLBIC**
$M0$	independence	958,513	959,142
$M1$	no temporal correlation	957,747	958,375
$M2$	no cross-sectional dependence	958,501	959,130
$M3$	no cluster effect	957,734	958,363
$M4$	no dispersion	958,491	959,120
$M5$	the proposed model	**957,730**	**958,358**

the direction, but could differ substantially in the size. In the dependence structure, we observe mild serial correlation in claims cost, which is explained by the short sampling period and the sparsity in the mixed outcome. Strong cross-sectional association is found among various types of claims. The cluster effect is statistically significant though relatively weak.

Table 9 compares the proposed model with alternative model specifications. Because dependence modeling is of primary interest for our application, we consider a nested dependence structure to emphasize the effect of various types of associations among claims cost. These nested cases are as follows: $M0$ assumes total independence, ignoring all types of dependence among claims; $M1$ assumes no serial correlation in either type of claim; $M2$ examines the longitudinal claims cost of each type separately by assuming independence among claim types; $M3$ ignores the cluster effect, assuming all vehicles under the same policy are independent. To examine the effect of dispersion, we also look into the Tweedie GLM without dispersion modeling ($M4$). The dependence structure in the mean regression $M4$ is the same as in the proposed copula model $M5$. We report in Table 9 the CLAIC and CLBIC statistics described in Section 3.2, with the preferred model highlighted. Both statistics suggest the favorite fit of the proposed model. The goodness-of-fit statistics of $M0$ and $M4$ are close, suggesting modeling dispersion and dependence are equally important in terms of the reported statistics. Consistent with the size of the dependence parameters reported in Table 8, ignoring the cross-sectional correlation among claim types results in the largest penalty in the model fit. Therefore, we use the proposed model $M5$ in the following applications.

5.2. *Basic ratemaking*. One unique feature that makes insurance differ from other commodities is that the pricing of an insurance contract is based on its future cost. Therefore, claims modeling becomes important in that it provides feedback to the classification and prediction of risks. To demonstrate the prediction, we consider a simplified risk classification system. The core element in the basic ratemaking is the expected cost of a policy. In this application, we consider the total cost of the four types of coverage. Ranking from low to high, there are five risk classes, Superior, Excellent, Good, Fair and Poor as defined in Table 10.

TABLE 10
Risk profile of hypothetical ratemaking classes

	Ratemaking classes				
	Superior	**Excellent**	**Good**	**Fair**	**Poor**
Young	0	0	0	0	0
Senior	1	1	1	1	0
Marital	1	1	1	1	0
Homeowner	1	1	1	1	0
Experience	1	1	0	1	0
Conviction	0	0	1	1	1
Newcar	0	1	1	1	1
Leasecar	0	1	0	1	1
Business	0	1	1	1	1
Highmilage	1	0	1	0	0
Multidriver	0	0	0	1	1

We report the expected cost of each risk class in Figure 2. To incorporate uncertainty in parameter estimates, we show their distributions instead of point estimates. The distributions are derived based on the Monte Carlo simulation from the asymptotic distributions of composite likelihood estimators. The straight line in Figure 2 corresponds to the 95% prediction interval. As anticipated, the expected claims costs of the four types of coverage increase from low risk to high risk class, and the differences among risk classes are statistically significant. As anticipated, the expected claims cost increases from low risk to high risk class, and the differences among risk classes are substantive. For comparison, we also show the actual observed average cost of each risk class as indicated by the solid dot in Figure 2. Because of the large number of observations in each risk class, the observed average cost is consistent with the prediction.

It is worth stressing that the risk classification is not limited to the ratemaking. Insurance is a highly regulated industry. In practice, regulation might not allow insurers to use certain risk factors in pricing, even if evidence supports the predictive power. In this case, risk classification is still useful because insurers could use such risk factors in the underwriting to prescreen risks and thus mitigate the adverse selection.

5.3. *Claims triage.* In response to claims filed by its customers, an insurer needs to investigate the claim, determine the coverage and legal liability, and settle the claim. This process is known as claims adjusting that aims to fulfill the insurer's promises to its policyholders. Claims adjusting is integral to establishing an insurer's relationship to its policyholders. The reputation of the insurer in settling claims directly impacts the marketing and retention of policyholder insurance.

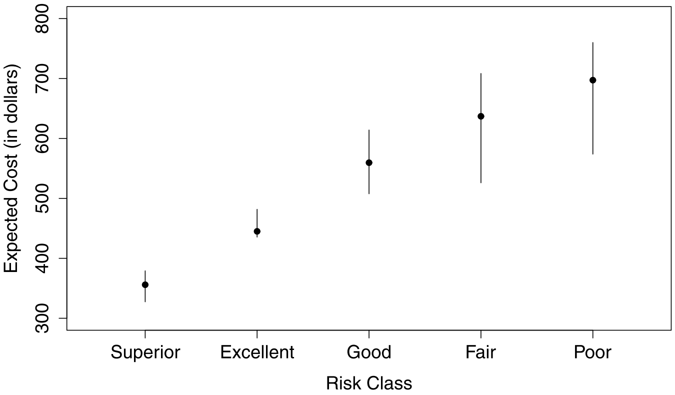

FIG. 2. *Distribution of expected cost by risk class. The straight line corresponds to the* 95% *prediction interval, and the solid dot indicates the observed average cost.*

Insurers often employ claims triage in the claims adjusting process so as to provide effective and efficient services to their customers. This is in the same spirit of nurse triage in emergency rooms, where a trained nurse performs an early assessment of patients to ensure they receive appropriate attention with the requisite degree of urgency. The purpose of claims triage is to decide early on the nature of the claim and assign an adjuster accordingly to ensure prompt and appropriate handling of the claim. For instance, standard stable claims might cost the insurer the same regardless of who handles the claim. However, large volatile claims could have a wide range of possible outcomes, thus, proper claim intervention from experienced adjusters may reduce the ultimate claim payment significantly.

Relevant questions are how many cohorts of adjusters an insurer should have in the claims department and how to allocate claims to each cohort. We provide solutions through cluster analysis. From a risk management perspective, underwriting controls the frequency of claims while claim settlement controls the severity. Hence, the quantity of interest is the cost of a claim given occurrence. Let Y_i denote the cost for policyholder i. If Y_i follows the Tweedie distribution, then one has

$$E(Y_i \mid Y_i > 0) = \frac{E(Y_i)}{1 - F_i(0)} = \frac{\mu_i}{1 - \exp(-(1/\phi_i)(\mu_i^{2-p}/(2-p)))}. \tag{7}$$

Based on the available predictors in Table 2, there are in total 1536 possible risk profiles. We calculate the severity of claims for each risk and perform a model-based clustering using normal mixture models.

Because claims of different types often require unique expertise from the adjusters, we perform the cluster analysis separately for each type of coverage. For illustration purposes, we report in Figure 3 the results for the coverage for property damage of the policyholder, collision and all risk. For each coverage type, we

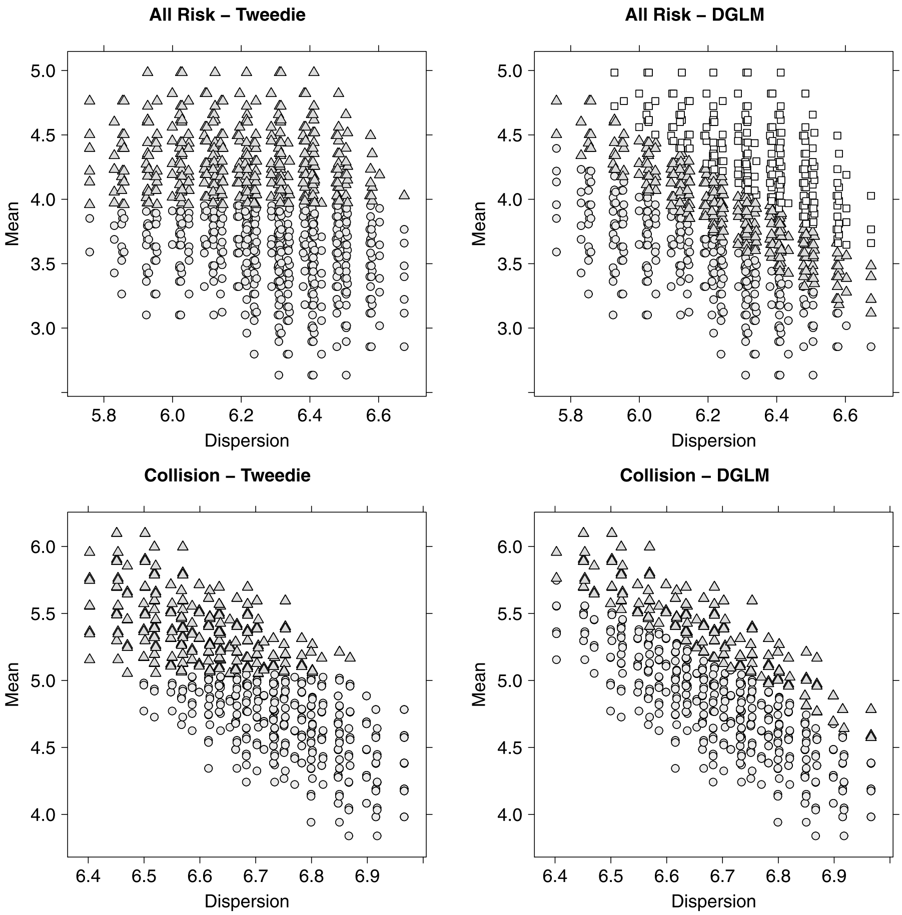

FIG. 3. *Comparison of clustering between the Tweedie and the double GLM models. The upper panel corresponds to the all risk coverage and the lower panel corresponds to the collision coverage.*

compare the clustering results for models $M4$ (Tweedie) and $M5$ (double GLM). In Figure 3, we display the risk clusters on the two-dimensional space determined by the mean (μ) and dispersion (ϕ) from the proposed model $M5$. The difference is as anticipated: with constant dispersion, the Tweedie model classifies risks based solely on the mean because the conditional expected cost in (7) is proportional to the mean. In contrast, the double GLM considers the heterogeneity in dispersion and thus takes it into account in the clustering. The analysis also demonstrates the value added by dispersion modeling.

5.4. *Portfolio management.* As a second application, we are interested in the distribution of insurance costs for a block of business. Consider a hypothetical portfolio consisting of 5000 policies that are evenly distributed in the five risk

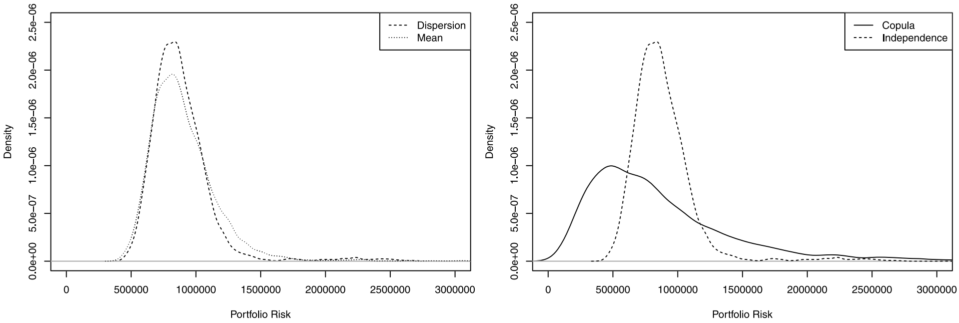

FIG. 4. *Comparison of portfolio risk. The left panel compares the predictive distribution of portfolio claims from the mean and dispersion models. The right panel compares the predictive distribution of portfolio claims from the copula and independence models.*

classes defined in Table 10. The claim distribution of the portfolio is provided in Figure 4. We examine the effect of dispersion modeling and dependence modeling on the portfolio risk. The left panel compares the claim distribution from the Tweedie GLM with and without dispersion modeling. To focus on this effect, we assume independence among claims. As the central limit theorem predicts, the effect of dispersion modeling is less pronounced on the portfolio risk than the individual risk. The right panel compares the claim distribution from the independent and the copula-based Tweedie double GLM. In contrast, the central limit theorem collapses in this case and the dependence modeling plays a critical role in risk aggregation.

The above analysis provides insight for the insurer to make better business decisions, such as to determine the appropriate risk capital and to implement the best reinsurance arrangement. To determine the risk capital for the portfolio, we employ two risk measures that have been widely used in both actuarial and financial studies, the value-at-risk (VaR) and the conditional tail expectation (CTE) [Klugman, Panjer and Willmot (2012), Chapter 3.5]. Both metrics focus on the tail of the distribution, with the VaR being a percentile measure and the CTE providing the expected value conditional on exceeding the VaR. Reinsurance, as a mean of risk management, transfers a portion of an insurer's risks to reinsurers. We examine two types of reinsurance arrangements, quota share reinsurance and excess-of-loss reinsurance. The former is a form of proportional reinsurance under which a fixed percentage of claims is ceded to the reinsurer. The latter is a nonproportional reinsurance where the reinsurer assumes all the losses above a specified dollar amount, the retention limit.

We report in Tables 11 and 12 for the required risk capital determined by VaR and CTE, respectively. Standard deviations are based on 1000 replications. In each case, we consider a variety of reinsurance arrangements and compare the implications from the independence and copula models. For quota share reinsurance,

TABLE 11
Estimated risk capital by VaR

	95% VaR				**99% VaR**			
	Independence		**Copula**		**Independence**		**Copula**	
	Estimate	**SD**	**Estimate**	**SD**	**Estimate**	**SD**	**Estimate**	**SD**
Full coverage	1,232,195	28,967	2,126,942	96,500	2,225,927	118,051	3,298,770	244,766
Quota = 0.25	924,147	21,725	1,595,207	72,375	1,669,445	88,538	2,474,078	183,575
Quota = 0.5	616,098	14,484	1,063,471	48,250	1,112,964	59,025	1,649,385	122,383
Quota = 0.75	308,049	7,242	531,736	24,125	556,482	29,513	824,693	61,192
Retention = 1 m	1,000,000	0	1,000,000	0	1,000,000	0	1,000,000	0
Retention = 1.5 m	1,232,195	28,967	1,500,000	0	1,500,000	0	1,500,000	0
Retention = 2 m	1,232,195	28,967	1,996,368	15,307	1,993,295	37,087	2,000,000	0

risk measures are calculated based on a rescaled distribution, and for excess-of-loss reinsurance, risk measures are calculated based on a truncated distribution. Dependence among risks plays a critical role in the capital determination. Positive (negative) association implies higher (lower) variability in the aggregate losses. The effects are linear and nonlinear for quota share and excess-of-loss reinsurance, respectively.

6. Conclusion.

In this work, we advanced modeling of insurance claims with a complex data structure that often exhibits in property-casualty insurance. The data are complex in that they are multivariate and multilevel. The multivariate nature is because a vehicle is insured by multiple types of coverage. The multilevel structure is because a policy covers more than one vehicle and they are observed

TABLE 12
Estimated risk capital by CTE

	95% CTE				**99% CTE**			
	Independence		**Copula**		**Independence**		**Copula**	
	Estimate	**SD**	**Estimate**	**SD**	**Estimate**	**SD**	**Estimate**	**SD**
Full coverage	1,757,925	152,819	2,907,977	188,560	2,601,139	556,144	4,268,984	644,687
Quota = 0.25	1,142,047	48,367	2,311,163	140,351	2,281,018	275,428	3,302,446	373,194
Quota = 0.5	910,513	14,185	1,705,420	93,570	1,478,016	165,732	2,381,024	213,539
Quota = 0.75	891,409	10,571	1,144,965	45,633	898,713	12,691	1,442,155	91,992
Retention = 1 m	1,245,359	37,423	1,632,082	43,660	1,245,359	37,423	1,632,082	43,660
Retention = 1.5 m	1,757,925	152,819	2,204,311	77,457	2,206,040	233,886	2,204,311	77,457
Retention = 2 m	1,757,925	152,819	2,760,684	137,491	2,422,570	317,148	2,764,584	137,344

over time. The proposed multivariate regression model is sufficiently flexible to handle our complex insurance data.

A main contribution of this article is the introduction of the regression framework for the multivariate semi-continuous claims in the multilevel context. We used the Tweedie distribution to accommodate the semi-continuous nature of claims cost while, at the same time, allowing for covariates in both mean and dispersion. One innovation in our approach is the employment of dependence modeling to accommodate the complex relationship among insurance claims. We used the Gaussian copula because of its flexibility in handling unbalanced data and the interpretability of the dependence parameters. We focused on the Gaussian copula and investigated misspecification with respect to the t copula.

The proposed method can potentially be extended in two areas. First, other copulas in the elliptical family possess similar flexibility in dependence modeling as the Gaussian copula, and thus are sensible candidates for our data. However, the properties of the composite likelihood estimator for elliptical copulas are worth further investigating in future research. Second, regression can be performed on the copula in addition to the marginal model. Adding covariates to the correlation structure could be difficult because of the constraints in the association matrix. One possible solution for future exploration is through the modified Cholesky decomposition [see, e.g., Pourahmadi (1999, 2000, 2007), among others].

The modeling approach developed in this article was motivated by the claims data in personal automobile insurance. However, it finds applications in a much broader context. As pointed out already, the multilevel structure exhibited by our claims data is very common in property-casualty insurance, including major personal lines (personal auto and homeowner) and commercial lines (worker's compensation, commercial multi-peril, commercial auto). The property-casualty insurance represents an important sector in the developed economy. The size of the market provides sufficient motivation for our work. Beyond the insurance market, the proposed model has potential application in the modeling of health care utilization, where a household in a private health plan or an employer in a group health plan forms the cluster, and the consumption of various types of care services is the outcome of interest. This provides additional motivation for the proposed method.

There are other possible approaches to modeling this type of data. One strategy is to use techniques from multivariate longitudinal data [Fahrmeir and Tutz (2001), Chapter 7.2]. Because the Tweedie density is not analytically tractable, the likelihood-based method for the Tweedie linear mixed model is difficult [see, e.g., Dunn and Smyth (2005, 2008) and Zhang (2013)]. The dispersion model in this context is another challenge. Another possibility is the two-part model for the semi-continuous longitudinal data [see, e.g., Olsen and Schafer (2001)]. However, the two-part framework is not ready to apply due to the multivariate and multilevel nature of our data. Since both strategies involve inference on the prediction of random quantities, we feel that the proposed approach is more flexible and easier to implement, especially when focus of the application is the predictive distribution of the outcome variables.

APPENDIX

This section provides a foundation for the dependence structure used in the Gaussian copula model. Consider the linear model for the transformed data:

$$\varepsilon_{ikjt} = \Phi^{-1}\big(F(y_{ikjt}; \mu_{ikjt}, p_j, \phi_{ikjt})\big) = \rho_j \varepsilon_{ikjt-1} + u_{ikjt} + v_{ijt},$$

where the three components of ε_{ikjt} are assumed to be independent and $\mathrm{Var}(\varepsilon_{ikjt}) = 1$. Let $\mathbf{u}_{ikt} = (u_{ik1t}, \dots, u_{ikJt})'$ and $\mathbf{v}_{it} = (v_{i1t}, \dots, v_{iJt})'$, and

$$\mathrm{Var}(\mathbf{u}_{ikt}) = \begin{pmatrix} \sigma_u^{(11)} & \cdots & \sigma_u^{(1J)} \\ \vdots & \ddots & \vdots \\ \sigma_u^{(J1)} & \cdots & \sigma_u^{(JJ)} \end{pmatrix}, \qquad \mathrm{Var}(\mathbf{v}_{it}) = \begin{pmatrix} \sigma_v^{(11)} & \cdots & \sigma_v^{(1J)} \\ \vdots & \ddots & \vdots \\ \sigma_v^{(J1)} & \cdots & \sigma_v^{(JJ)} \end{pmatrix}.$$

We show that the above model implies the dependence structure $\mathbf{\Sigma} = \mathbf{B}_{K\times K} \otimes \mathbf{P}_{(TJ)\times(TJ)}$ specified in Section 3.1 if and only if $\mathrm{Var}(\mathbf{v}_{it}) = \lambda \,\mathrm{Var}(\mathbf{u}_{ikt})$. As shown below, this assumption implies the exchangeable correlation in $\mathbf{B}_{K\times K}$ and the separable dependence in $\mathbf{\Sigma}$. In our context, the exchangeable structure means that correlation between vehicles under the same policy is introduced by a latent household specific heterogeneity. The separable dependence refers to the identical coverage-temporal relation for all vehicles within the same household. Both are reasonable assumptions if one thinks that individuals from the same household might possess similar risk characteristics and risk appetite and they might drive their vehicles interchangeably.

Let $\boldsymbol{\varepsilon}_{ikj} = (\varepsilon_{ikj1}, \dots, \varepsilon_{ikjT})'$. We consider four scenarios in dependence analysis. The first is regarding the serial correlation among insurance costs for each coverage type. It can be shown that

$$\mathrm{Var}(\boldsymbol{\varepsilon}_{ikj}) = \begin{pmatrix} 1 & \rho_j & \cdots & \rho_j^{T-1} \\ & 1 & \cdots & \rho_j^{T-2} \\ & & \ddots & \vdots \\ & & & 1 \end{pmatrix} = \mathbf{P}_{jj}.$$

The second is regarding dependence among different types of claims cost for a given vehicle. For the same time period $t = t'$,

$$\mathrm{Cov}(\varepsilon_{ikjt}, \varepsilon_{ikj't}) = \frac{\sigma_u^{(jj')} + \sigma_v^{(jj')}}{1 - \rho_j \rho_j'} = \frac{1+\lambda}{\lambda} \frac{\sigma_v^{(jj')}}{1 - \rho_j \rho_j'} := \sigma_{jj'}.$$

For the different time periods $t \neq t'$,

$$\mathrm{Cov}(\varepsilon_{ikjt}, \varepsilon_{ikj't'}) = \begin{cases} \rho_j^{t-t'} \mathrm{Cov}(\varepsilon_{ikjt'}, \varepsilon_{ikj't'}) = \rho_j^{t-t'} \sigma_{jj'}, & \text{if } t > t', \\ \rho_{j'}^{t'-t} \mathrm{Cov}(\varepsilon_{ikjt}, \varepsilon_{ikj't}) = \rho_{j'}^{t'-t} \sigma_{jj'}, & \text{if } t < t'. \end{cases}$$

Thus, we have

$$\mathrm{Cov}(\boldsymbol{\varepsilon}_{ikj},\boldsymbol{\varepsilon}_{ikj'}) = \sigma_{jj'}\begin{pmatrix} 1 & \rho_{j'} & \cdots & \rho_{j'}^{T-1} \\ \rho_j & 1 & \cdots & \rho_{j'}^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_j^{T-1} & \rho_j^{T-2} & \cdots & 1 \end{pmatrix} = \sigma_{jj'}\mathbf{P}_{jj'}.$$

The third is the dependence between losses of a particular type of coverage but from different vehicles under the same policy. For the same period $t = t'$,

$$\mathrm{Cov}(\varepsilon_{ikjt},\varepsilon_{ik'jt}) = \frac{\sigma_v^{(jj)}}{1-\rho_j^2} = \frac{\lambda}{1+\lambda} := \delta,$$

and $0 < \delta < 1$. Note that the cluster-specific effect only introduces positive correlation. For the different time periods $t \neq t'$,

$$\mathrm{Cov}(\varepsilon_{ikjt},\varepsilon_{ikj't'}) = \begin{cases} \rho_j^{t-t'}\,\mathrm{Cov}(\varepsilon_{ikjt'},\varepsilon_{ik'jt'}) = \delta\rho_j^{t-t'}, & \text{if } t > t', \\ \rho_{j'}^{t'-t}\,\mathrm{Cov}(\varepsilon_{ikjt},\varepsilon_{ik'jt}) = \delta\rho_{j'}^{t'-t}, & \text{if } t < t'. \end{cases}$$

Hence, one obtains

$$\mathrm{Cov}(\boldsymbol{\varepsilon}_{ikj},\boldsymbol{\varepsilon}_{ik'j}) = \delta\begin{pmatrix} 1 & \rho_j & \cdots & \rho_j^{T-1} \\ & 1 & \cdots & \rho_j^{T-2} \\ & & \ddots & \vdots \\ & & & 1 \end{pmatrix} = \delta\mathbf{P}_{jj}.$$

The fourth is the dependence between losses of different coverage types and from different vehicles insured by the same contract. For the same period $t = t'$,

$$\mathrm{Cov}(\varepsilon_{ikjt},\varepsilon_{ik'j't}) = \frac{\sigma_v^{(jj')}}{1-\rho_j\rho_{j'}} = \frac{\lambda}{1+\lambda}\sigma_{jj'} = \delta\sigma_{jj'}.$$

For the different time periods $t \neq t'$,

$$\mathrm{Cov}(\varepsilon_{ikjt},\varepsilon_{ik'j't'}) = \begin{cases} \rho_j^{t-t'}\,\mathrm{Cov}(\varepsilon_{ikjt'},\varepsilon_{ik'j't'}) = \rho_j^{t-t'}\delta\sigma_{jj'}, & \text{if } t > t', \\ \rho_{j'}^{t'-t}\,\mathrm{Cov}(\varepsilon_{ikjt},\varepsilon_{ik'j't}) = \rho_{j'}^{t'-t}\delta\sigma_{jj'}, & \text{if } t < t'. \end{cases}$$

In addition,

$$\mathrm{Cov}(\varepsilon_{ikjt},\varepsilon_{ik'j't}) = \frac{\sigma_v^{(jj')}}{1-\rho_j\rho_{j'}} = \frac{\tau_{jj'}\sqrt{\sigma_v^{(jj)}}\sqrt{\sigma_v^{(j'j')}}}{1-\rho_j\rho_{j'}} = \delta\tau_{jj'}\frac{\sqrt{1-\rho_j^2}\sqrt{1-\rho_{j'}^2}}{1-\rho_j\rho_{j'}}.$$

This justifies the reparameterization $\sigma_{jj'} = \tau_{jj'}\sqrt{1-\rho_j^2}\sqrt{1-\rho_{j'}^2}/(1-\rho_j\rho_{j'})$ and provides a natural interpretation of parameter $\tau_{jj'}$. Therefore,

$$\operatorname{Cov}(\boldsymbol{\varepsilon}_{ikj}, \boldsymbol{\varepsilon}_{ik'j'}) = \delta\sigma_{jj'} \begin{pmatrix} 1 & \rho_{j'} & \cdots & \rho_{j'}^{T-1} \\ \rho_j & 1 & \cdots & \rho_{j'}^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_j^{T-1} & \rho_j^{T-2} & \cdots & 1 \end{pmatrix} = \delta\sigma_{jj'}\mathbf{P}_{jj'}.$$

Acknowledgments. We are grateful to Editor Thomas Brendan Murphy, the Associate Editor and two anonymous reviewers for their constructive comments and suggestions that helped improved the quality and presentation of the paper.

REFERENCES

Aas, K., Czado, C., Frigessi, A. and Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance Math. Econom.* **44** 182–198. MR2517884

Arellano-Valle, R. B., Castro, L. M., González-Farías, G. and Muñoz-Gajardo, K. A. (2012). Student-*t* censored regression model: Properties and inference. *Stat. Methods Appl.* **21** 453–473. MR2992913

Castro, L. M., Lachos, V. H., Ferreira, G. P. and Arellano-Valle, R. B. (2014). Partially linear censored regression models using heavy-tailed distributions: A Bayesian approach. *Stat. Methodol.* **18** 14–31. MR3151861

Cox, D. R. and Reid, N. (2004). A note on pseudolikelihood constructed from marginal densities. *Biometrika* **91** 729–737. MR2090633

Dunn, P. K. and Smyth, G. K. (2005). Series evaluation of Tweedie exponential dispersion model densities. *Stat. Comput.* **15** 267–280. MR2205390

Dunn, P. K. and Smyth, G. K. (2008). Evaluation of Tweedie exponential dispersion model densities by Fourier inversion. *Stat. Comput.* **18** 73–86. MR2416440

Fahrmeir, L. and Tutz, G. (2001). *Multivariate Statistical Modelling Based on Generalized Linear Models*, 2nd ed. Springer, New York. MR1832899

Frees, E. (2014). Frequency and severity models. In *Predictive Modeling Applications in Actuarial Sciences* (E. Frees, G. Meyers and R. Derrig, eds.) 138–166. Cambridge Univ. Press, Cambridge.

Frees, E. W., Shi, P. and Valdez, E. A. (2009). Actuarial applications of a hierarchical insurance claims model. *Astin Bull.* **39** 165–197. MR2749883

Frees, E. W. and Valdez, E. A. (2008). Hierarchical insurance claims modeling. *J. Amer. Statist. Assoc.* **103** 1457–1469. MR2655723

Gao, X. and Song, P. X.-K. (2010). Composite likelihood Bayesian information criteria for model selection in high-dimensional data. *J. Amer. Statist. Assoc.* **105** 1531–1540. MR2796569

Garay, A. M., Lachos, V. H., Bolfarine, H. and Cabral, C. R. (2016). Linear censored regression models with scale mixtures of normal distributions. *Statistical Papers*. To appear.

Godambe, V. P. (1960). An optimum property of regular maximum likelihood estimation. *Ann. Math. Stat.* **31** 1208–1211. MR0123385

Greene, W. (2007). *Econometric Analysis*. Prentice Hall, New Jersey.

Hintze, J. L. and Nelson, R. D. (1998). Violin plots: A box plot-density trace synergism. *Amer. Statist.* **52** 181–184.

Joe, H. (2015). *Dependence Modeling with Copulas*. *Monographs on Statistics and Applied Probability* **134**. CRC Press, Boca Raton, FL. MR3328438

JOE, H. and LEE, Y. (2009). On weighting of bivariate margins in pairwise likelihood. *J. Multivariate Anal.* **100** 670–685. MR2478190

JØRGENSEN, B. (1987). Exponential dispersion models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **49** 127–162. MR0905186

JØRGENSEN, B. and PAES DE SOUZA, M. C. (1994). Fitting Tweedie's compound Poisson model to insurance claims data. *Scand. Actuar. J.* **1** 69–93. MR1286486

KLUGMAN, S., PANJER, H. and WILLMOT, G. (2012). *Loss Models*: *From Data to Decisions*, 4th ed. Wiley, New York.

LINDSAY, B. G. (1988). Composite likelihood methods. In *Statistical Inference from Stochastic Processes* (*Ithaca*, *NY*, 1987). *Contemp. Math.* **80** 221–239. Amer. Math. Soc., Providence, RI. MR0999014

MCCULLAGH, P. and NELDER, J. A. (1989). *Generalized Linear Models*, 2nd ed. Chapman & Hall, London. MR3223057

MOLENBERGHS, G. and VERBEKE, G. (2005). *Models for Discrete Longitudinal Data*. Springer, New York. MR2171048

NELSEN, R. B. (2006). *An Introduction to Copulas*, 2nd ed. Springer, New York. MR2197664

OLSEN, M. K. and SCHAFER, J. L. (2001). A two-part random-effects model for semicontinuous longitudinal data. *J. Amer. Statist. Assoc.* **96** 730–745. MR1946438

PANAGIOTELIS, A., CZADO, C. and JOE, H. (2012). Pair copula constructions for multivariate discrete data. *J. Amer. Statist. Assoc.* **107** 1063–1072. MR3010894

PARKS, R. W. (1967). Efficient estimation of a system of regression equations when disturbances are both serially and contemporaneously correlated. *J. Amer. Statist. Assoc.* **62** 500–509. MR0216633

POURAHMADI, M. (1999). Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika* **86** 677–690. MR1723786

POURAHMADI, M. (2000). Maximum likelihood estimation of generalised linear models for multivariate normal covariance matrix. *Biometrika* **87** 425–435. MR1782488

POURAHMADI, M. (2007). Cholesky decompositions and estimation of a covariance matrix: Orthogonality of variance-correlation parameters. *Biometrika* **94** 1006–1013. MR2376812

SHI, P. (2016). Insurance ratemaking using a copula-based multivariate Tweedie model. *Scand. Actuar. J.* **3** 198–215. MR3435180

SHI, P., ZHANG, W. and VALDEZ, E. A. (2012). Testing adverse selection with two-dimensional information: Evidence from the Singapore auto insurance market. *Journal of Risk and Insurance* **79** 1077–1114.

SMITH, M., MIN, A., ALMEIDA, C. and CZADO, C. (2010). Modeling longitudinal data using a pair-copula decomposition of serial dependence. *J. Amer. Statist. Assoc.* **105** 1467–1479. MR2796564

SMYTH, G. K. (1996). Regression analysis of quantity data with exact zeros. In *Proceedings of the Second Australia-Japan Workshop on Stochastic Models in Engineering*, *Technology and Management* (R. Wilson, S. Osaki and D. Murthy, eds.) 17–19. Gold Coast, Australia.

SMYTH, G. K. and JØRGENSEN, B. (2002). Fitting Tweedie's compound Poisson model to insurance claims data: Dispersion modelling. *Astin Bull.* **32** 143–157. MR1930491

SONG, P. X.-K., LI, M. and YUAN, Y. (2009). Joint regression analysis of correlated data using Gaussian copulas. *Biometrics* **65** 60–68. MR2665846

TOBIN, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica* **26** 24–36. MR0090462

TWEEDIE, M. C. K. (1984). An index which distinguishes between some important exponential families. In *Statistics*: *Applications and New Directions* (*Calcutta*, 1981) 579–604. Indian Statist. Inst., Calcutta. MR0786162

VARIN, C. (2008). On composite marginal likelihoods. *AStA nnv. Stat. Anal.* **92** 1–28. MR2414624

VARIN, C., REID, N. and FIRTH, D. (2011). An overview of composite likelihood methods. *Statist. Sinica* **21** 5–42. MR2796852

VARIN, C. and VIDONI, P. (2005). A note on composite likelihood inference and model selection. *Biometrika* **92** 519–528. MR2202643

ZHANG, Y. (2013). Likelihood-based and Bayesian methods for Tweedie compound Poisson linear mixed models. *Stat. Comput.* **23** 743–757. MR3247830

ZHAO, Y. and JOE, H. (2005). Composite likelihood estimation in multivariate data analysis. *Canad. J. Statist.* **33** 335–356. MR2193979

P. SHI
WISCONSIN SCHOOL OF BUSINESS
UNIVERSITY OF WISCONSIN-MADISON
5281 GRAINGER HALL
975 UNIVERSITY AVE
MADISON, WISCONSIN 53706
USA
E-MAIL: pshi@bus.wisc.edu

X. FENG
DEPARTMENT OF STATISTICS
UNIVERSITY OF WISCONSIN-MADISON
1300 UNIVERSITY AVENUE
MADISON, WISCONSIN 53706
USA
E-MAIL: xfeng25@wisc.edu

J.-P. BOUCHER
DÉPARTEMENT DE MATHÉMATIQUES
UNIVERSITÉ DU QUÉBEC À MONTRÉAL
BUREAU PK-5720
201, AVENUE DU PRÉSIDENT-KENNEDY
MONTRÉAL (QUÉBEC), H2X 3Y7
CANADA
E-MAIL: boucher.jean-philippe@uqam.ca