Pair Copula Constructions for Insurance Experience Rating

Peng Shi
Wisconsin School of Business
University of Wisconsin - Madison
Email: pshi@bus.wisc.edu

Lu Yang
Department of Statistics
University of Wisconsin - Madison
Email: luyang@stat.wisc.edu

October 26, 2016

Abstract

In non-life insurance, insurers use experience rating to adjust premiums to reflect policyholders' previous claim experience. Performing prospective experience rating can be challenging when the claim distribution is complex. For instance, insurance claims are semicontinuous in that a fraction of zeros is often associated with an otherwise positive continuous outcome from a right-skewed and long-tailed distribution. Practitioners use credibility premium that is a special form of the shrinkage estimator in the longitudinal data framework. However, the linear predictor is not informative especially when the outcome follows a mixed distribution.

In this article, we introduce a mixed vine pair copula construction framework for modeling semicontinuous longitudinal claims. In the proposed framework, a two-component mixture regression is employed to accommodate the zero inflation and thick tails in the claim distribution. The temporal dependence among repeated observations is modeled using a sequence of bivariate conditional copulas based on a mixed D-vine. We emphasize that the resulting predictive distribution allows insurers to incorporate past experience into future premiums in a nonlinear fashion and the classic linear predictor can be viewed as a nested case.

In the application, we examine a unique claims dataset of government property insurance from the state of Wisconsin. Due to the discrepancies between the claim and premium distributions, we employ an ordered Lorenz curve to evaluate the predictive performance. We show that the proposed approach offers substantial opportunities for separating risks and identifying profitable business when compared with alternative experience rating methods.

Keywords: Government insurance, Mixed D-vine, Mixture regression, Predictive distribution, Zero inflation

1 Introduction

In non-life (property, liability, and health) insurance, insurers use experience rating to adjust premiums to reflect policyholders' past loss experience. Premiums decrease (increase) if the experience of a policyholder is better (worse) than that assumed in the manual rate - a premium rate developed from the experience of a large number of homogeneous policies defined by the insurer's risk classification system. Experience rating can be prospective or retrospective. We restrict our consideration to the former that points to a predictive modeling application.

Experience rating improves insurance market efficiency as a dynamic contract mechanism under information asymmetry, therefore providing a competitive advantage for insurers deft at its employment over the rival firms. First, an insurer's risk classification system might not be perfect. Unobserved heterogeneity remains after all underwriting criteria are accounted for. Experience rating allows insurers to further separate good risks from bad risks, and thus helps mitigate adverse selection. Second, adjusting premium based on past experience gives policyholders incentives for loss prevention, which is known as moral hazard in the economics literature.

The statistical component of experience rating is to model longitudinal insurance claims and to infer the predictive distribution of future claims given previous loss experience. This can be a difficult task when the risk distribution is complex. For instance, the distribution of claims is well known to be a mixture of zeros and a right-skewed and long-tailed distribution. The degenerate distribution at zero corresponds to no claims and the positive thick-tailed distribution describes the amount of claims given occurrence. In the case of the property insurance provided by the Wisconsin local government property fund (Section 2), for the single coverage on buildings and contents, on average about 70% of policyholders have zero claim per year, and the coefficients of skewness and kurtosis of the (conditional) claim amount are 21.18 and 570.63, respectively.

1.1 Credibility Theory and Longitudinal Data

Insurers use credibility ratemaking to perform prospective experience rating (adjust future premium based on past experience) on a risk or a group of risks. Credibility theory has a long history in actuarial science, with fundamental contributions dating back to the 1900s (Mowbray (1914) and Whitney (1918)). The intuitive concept of credibility premium is to express the expected future claim of a given risk class as a weighted sum of the average claim from the risk class and the average claim over all other risk classes, which begs the question that how much of the experience of a given policyholder is due to the random variation in the underlying risk and how much is due to the policyholder being better or worse than average. The classic work of Bühlmann (1967) provided a systematic solution using what is known as random-effects framework, thereby modern theory of credibility has developed and flourished. Frees et al. (1999) established the link between the credibility theory in actuarial science and the longitudinal data models in statistics, and noted that the credibility predictor is a special form of the shrinkage estimator in the longitudinal data framework.

The longitudinal data interpretation of credibility theory suggests additional models and techniques that actuaries can use in experience rating. In the current literature, there are two popular modeling frameworks for analyzing longitudinal, and in general, clustered data. One approach is mixed models where a mean model is specified conditional on cluster-specific random effects (Diggle et al. (2002)). The other approach is marginal models using generalized estimating equations (Liang and Zeger (1986)). The former is more relevant to the prediction application in our context.

Due to the zero inflation and long tails exhibited in the insurance claims, standard longitudinal data models are not ready to apply to insurers' experience rating. One attractive approach to characterizing the complex structure of semicontinuous longitudinal data is the two-part mixed models with correlated random effects (see, for example, Olsen and Schafer (2001)). The two-part model, decomposing the semicontinuous outcome into a zero component and a continuous component, has long found its use in modeling insurance claims (Frees (2014)). An alternative available strategy is based on a mixed distribution with mass probability at zero that is constructed from some structured process. One example that is often used in insurance claims modeling is the Tweedie compound Poisson model (Jørgensen and de Souza (1994) and Smyth and Jørgensen (2002)). Inference for the Tweedie distribution and Tweedie mixed model was investigated by Dunn and Smyth (2005; 2008) and Zhang (2013), respectively.

However, both approaches are subject to several difficulties in the current application of predictive modeling. First, likelihood-based estimation is computationally expensive, especially with big data such as in insurance. Second, prediction of random effects in nonlinear models is never an easy task, which further hinders the derivation of the predictive distribution. Third, the structural assumption of the subject-specific heterogeneity implies a symmetric relation among repeated observations, and thus limits the way past experience is incorporated into the prediction.

1.2 Copula Regression for Repeated Measurements

Reminiscent of the marginal models for longitudinal data, recent literature has observed rapid growth of copula regression models for repeated measurements (see Joe (2014) for the recent advancement on copulas). Marginal models specify the mean model and covariance separately. In the covariance model, dependence is formulated using working correlation matrix and is treated as nuisance parameter. Therefore, marginal models are suitable when the mean model is of central interest and are not appropriate for prediction. In contrast, in copula regression, univariate margins are specified by (semi-)parametric regressions, while the cluster or serial dependence is modeled through a multivariate copula. Masarotto et al. (2012) provided a comprehensive methodological review on marginal regression models using Gaussian copulas.

The first effort of using copula regression for experience rating in insurance is due to Frees and Wang (2005), where a t-copula was proposed to accommodate the serial correlation in the severity of automobile claims from a cross section of risk classes. However, predictive applications of copula model for semicontinuous longitudinal insurance claims are rarely found in the literature. Two examples are Frees and Wang (2006) and Shi et al. (2016). The former adopted the two-part

approach by decoupling the claims cost into a frequency and a (conditional) severity component, and specified elliptical copulas for each longitudinal component. The latter considered the Tweedie model for the marginal and employed the Gaussian copula to analyze the semi-continuous claims in a multilevel context.

In this article, we introduce a mixed vine pair copula construction framework for modeling semi-continuous longitudinal claims. In the proposed framework, a two-component mixture regression is employed to accommodate the zero inflation and thick tails in the claim distribution. The temporal dependence among repeated observations is modeled using a sequence of bivariate conditional copulas based on a mixed D-vine. The proposed approach enjoys several advantages compared with the methods available in the existing literature. First, the mixture regression combines the merits of both the two-part and the Tweedie models. Unlike the Tweedie, it allows the analyst to use different sets of predictors for the frequency and severity of claims. In the meanwhile, it does not require separate copulas for the frequency and severity components, and thus avoids the unbalanced data issue in the conditional severity model. It is worth stressing that modeling marginal is of no less importance than modeling dependence. Inference of dependence will be biased if marginals are not correctly specified. In the data analysis, we carefully examine the marginal distribution of claims prior to the specification of the dependence among them. Second, compared with the elliptical copulas in the current literature of experience rating, the mixed vine pair copula construction allows for more flexible dependence structure by using asymmetric bivariate copulas as building blocks. In addition, the computational burden is much lower than the case of elliptical copulas when there are discrete components in the response variable. This feature has significant practical value where the true data generating process is unknown. Third, for the purposes of experience rating, we are interested in one particular type of statistical inference - prediction. Under the pair copula framework, it is straightforward to derive the predictive distribution of future claim given past experience without referring to the Bayesian approach. We also point out that many existing credibility predictors can be viewed in the proposed approach.

The main contribution of this article to the literature is the introduction of the vine pair copula constructions for mixed data and the novel application in insurance experience rating. Vine copulas have been studied for both continuous and discrete data. Following the seminal work of Bedford and Cooke (2001, 2002) on this new class of graphical models, Kurowicka and Cooke (2006) and Aas et al. (2009) are among the first to exploit the idea of building a multivariate model through a series of bivariate copulas for continuous data. Smith et al. (2010) employed a Bayesian approach and investigated copula selection in the D-vine for longitudinal data. More recently, Panagiotelis et al. (2012) introduced the discrete analogue to the vine pair copula construction. Stöber (2013) and Stöber et al. (2015) studied the theory and applications of pair copula constructions for mixed data. Our work fills the blank in the literature on vine copulas for a special type of mixed outcome - hybrid data. Specifically, in Stöber's work, "mixed data" corresponds to the case where a copula is used to join a continuous distribution and a discrete distribution. In contrast, we use "mixed data" to refer to the case of a hybrid distribution, i.e. a random variable with both discrete and

continuous components, and a copula is used to join two mixed or hybrid distributions. Note that we motivate the mixed D-vine structure using the predictive nature of the application in insurance experience rating. However, the notion of mixed vine pair copula construction easily extends to regular vines.

The rest of the article are organized as follows: Section 2 describes the local government property insurance fund in the state of Wisconsin and the characteristics of the claim data for the property coverage on buildings and contents. Section 3 introduces the mixed vine pair copula construction for semicontinuous clustered data, and discusses model inference. Section 4 provides results on the application in experience rating in the property insurance in Wisconsin. Technical details are summarized in Appendix, along with a simulation study where we investigate the estimation and copula selection for the mixed D-vine model.

2 Wisconsin Local Government Property Insurance Fund

2.1 Background

The Local Government Property Insurance Fund (LGPIF) was established by the Chapter 605 of the Wisconsin Statutes and is administered by the Wisconsin Office of the Commissioner of Insurance. The purpose of the LGPIF is to make property insurance available for local government units, such as counties, cities, towns, villages, school districts, and library boards, etc. The LGPIF is designed to moderate the budget effects of uncertain insurable events for local government entities with separate budgetary responsibilities and does not provide coverage for state government buildings.

The LGPIF offers three major types of coverage for local government properties: building and contents, inland marine (construction equipment), and motor vehicles. It covers all causes of property losses with certain exclusions. Such exclusions include those resulting from flood, earthquake, wear and tear, extremes in temperature, mold, war, nuclear reactions, and embezzlement or theft by an employee.

The fund operates, to some extent, as a stand-alone property insurer in that it charges premiums and pays claims to its policyholders, i.e. local government units. In terms of size, the fund is currently insuring over a thousand entities. On average, it writes approximately $25 million in premiums and $75 billion in coverage each year. However, the fund differs from proprietary insurance companies in its operations. First, the fund has only one state employee who supervises the day-to-day operations by contracting for specialized services, such as claim management and policy administration. Second, the LGPIF is not allowed to deny coverage, although local government units can secure insurance in the open market.

2.2 Data Characteristics

In experience rating, we examine the insurance coverage for building and contents. Data are collected for 1,019 local government entities over six years from 2006 to 2011. Due to the role of "residual market" of the LGPIF, attrition is a rare event at least during our sampling period. For

the same reason, the policyholders' experience becomes particularly important for pricing insurance contracts because other sources of market data may not be relevant. We use data in years 2006-2010 to develop the model and reserve the data of 2011 for validation.

The quantity of interest is the entity-level cost of claims that serves as the basis for determining the pure premium. Similar to private commercial insurers, the government insurance fund keeps track of claims for its pool of policyholders, from which we derive the total claim cost of each entity for each year. In addition, the fund further breaks down the total cost of claims by the cause of losses, known as peril in property insurance. In this application, we examine the total cost as well as the cost by peril. Statistically speaking, one might prefer to analyze claims by peril presuming that more information is revealed at granular level observations. On the other hand, one might argue for the simplicity of aggregating data across perils in the sense of sufficient statistics. In practice, the choice often depends on the preference of the analyst and the type of data collected by the insurer. We view this as an empirical question and compare the predictive performance of both common practices.

Table 1 summarizes the distribution of claim cost by year. The first panel corresponds to the total claims and the other three correspond to losses caused by water, fire, and other perils, respectively. Water and fire (including smoke) damages are among those of highest frequency of occurrence. Examples of other perils include lightning strikes, windstorms and hail, explosion. All four outcomes are semicontinuous in that a significant portion of zeros is associated with an otherwise positive continuous outcome. The zeros imply no claims and the positive component indicates the size of claims. In the table, we report the probability of zero claim denoted by p_0. For instance, regardless of the cause of loss, about 72.3% entities did not report any claim during year 2006. As expected, the percentage of zeros is larger when decomposing claims by peril, and water and fire damages are more common than other perils.

Table 1: Distribution of the claim cost by year and by peril

Total				Water			
Year	p_0	Mean	SD	Year	p_0	Mean	SD
2006	0.723	71,338	499,690	2006	0.847	67,082	536,722
2007	0.677	51,225	178,513	2007	0.821	13,791	27,256
2008	0.716	40,439	146,224	2008	0.850	15,260	41,462
2009	0.722	36,932	143,783	2009	0.851	20,995	52,598
2010	0.626	94,784	728,353	2010	0.821	119,892	986,053
Fire				Other			
Year	p_0	Mean	SD	Year	p_0	Mean	SD
2006	0.855	19,893	58,387	2006	0.920	81,808	515,961
2007	0.863	43,441	139,471	2007	0.863	59,010	201,994
2008	0.873	40,087	167,739	2008	0.887	36,357	118,778
2009	0.879	32,975	87,335	2009	0.912	35,606	167,402
2010	0.840	33,384	111,541	2010	0.816	47,330	243,250

Conditioning on at least one claims, we also present in Table 1 the mean and standard deviation

of the amount of claims. The large standard deviation is as anticipated and is indicative of the skewness and thick tails in the claim size distribution. Another noticeable feature in severity is the substantial variation across years, especially for water damages. This is in contrast with claim frequency where temporal variation is less pronounced. We attribute the temporal variation in the claim size to the heavy tails of the underlying distribution and we accommodate such data feature by using a flexible parametric regression. To visualize the size distribution, Figure 1 displays the violin plot of the amount of claims by year and by peril. One can think of violin plot as a marriage of box plot and density trace (see Hintze and Nelson (1998) for more details). The plots suggest that the occurrence of extremely large losses is not unusual and the claims related to water damages are more volatile than fire and other perils. Overall, zero inflation and heavy tails in the claim cost distribution, as shown in Table 1 and Figure 1, motivate the two-component mixture regression in Section 3.1.

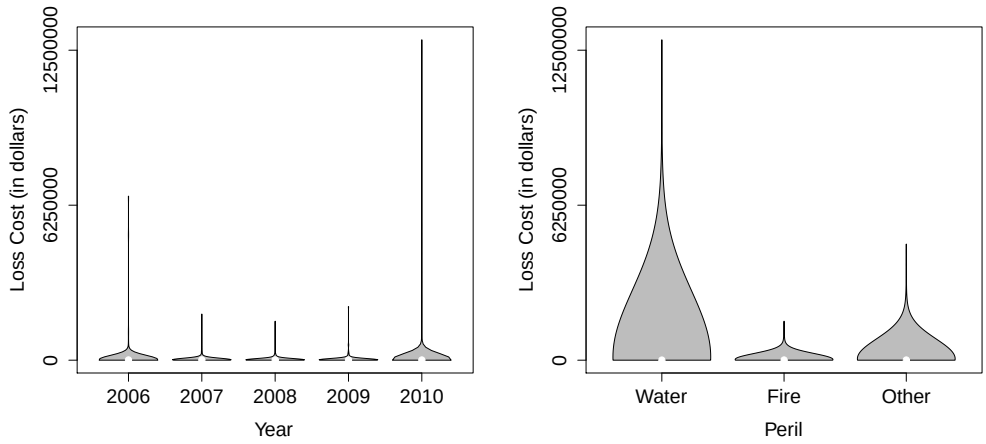

Figure 1: Violin plots of the amount of claims by year and by peril.

In a risk classification system, an insurer uses observed policyholder and contract characteristics to explain the variability in the insurance claims and then reflects such heterogeneity in the ratemaking. For example, the large claim amount could, to certain extent, relate to the size of the coverage. Table 2 presents the rating variables, their descriptions, and the associated descriptive statistics. Unlike personal lines of business (such as automobile and homeowner insurance), we have a very limited number of predictors used in the rating system, which is not unusual in commercial insurance ratemaking. One rating variable is the entity type that indicates whether the covered buildings belong to a city, county, school, town, village, or a miscellaneous entity such as fire stations. Apparently the entity type does not change over the years. For example, about 15% policyholders are city entities and 30% are school districts. We set miscellaneous entity (TypeMisc) as reference level in the analysis. As an incentive to prevent and mitigate loss, the fund offers credits for the different types of fire alarms. In our case, the policyholder receives a 5% discount in premium if automatic smoke alarms are installed in some of the main rooms within a building, a

10% discount if alarms are installed in all of the main rooms, and a 15% discount if the alarms are 24/7 monitored. No alarm credit (AC00) is omitted as reference level in the regression analysis. The alarm credit is often subject to the underwriter's discretion. The increasing temporal pattern in the alarm credit might be due to the fact that policyholders are responsive to the incentives and the advanced alarm system becomes accessible at a lower cost. Because of the skewness in the amount of coverage (in million dollars), we report its mean and standard deviation (in parenthesis) of the coverage amount in the log scale. The statistics indicate a relatively small variation in coverage overtime.

Table 2: Description and summary statistics of covariates †

Variable	Description	Year =				
		2006	2007	2008	2009	2010
TypeCity	=1 if entity type is city	0.146				
TypeCounty	=1 if entity type is county	0.061				
TypeSchool	=1 if entity type is school	0.291				
TypeTown	=1 if entity type is town	0.164				
TypeVillage	=1 if entity type is village	0.231				
AC05	=1 indicate 5% alarm credit	0.025	0.026	0.034	0.054	0.074
AC10	=1 indicate 10% alarm credit	0.045	0.051	0.050	0.067	0.084
AC15	=1 indicate 15% alarm credit	0.381	0.399	0.434	0.486	0.544
Coverage	Amount of coverage in log scale	2.065	2.153	2.227	2.285	2.292
		(2.021)	(2.000)	(1.981)	(1.990)	(1.987)

† Standard deviation for continuous covariates is reported in parenthesis.

Through repeated contracting, an insurer expects to gain private information regarding the risk level of its policyholders, and thus competitive advantages over its rivals. In particular, insurers hope to leverage the policyholders' past claim experience into the prediction of future claims. To this end, we explore the serial association of the claim cost over time. To motivate the specification of the mixed D-vine in Section 3.2, we report in Table 3 the partial rank correlations for the total claim cost as well as the claim cost for each peril. Specifically, the partial correlations are calculated recursively using relation:

$$\rho_{jk;l\cup V} = \frac{\rho_{jk;V} - \rho_{jl;V}\rho_{kl;V}}{\sqrt{(1-\rho^2_{jl;V})(1-\rho^2_{kl;V})}},$$

where j, k, l are distinct, V is a subset of $\{1, ..., m\}\backslash\{j, k, l\}$, and $\rho_{jk;V}$ denotes the partial correlation between the jth and kth variables controlling for variables with indexes in V. The starting values in the recursive calculation are sample pairwise correlations.

In each correlation matrix, the upper triangle exhibits the Kendall's *tau* and the lower triangle the Spearman's *rho*. Using the upper triangle of the total claim cost as an example, the Kendall's *tau* between the claim cost in 2006 and 2007 is 0.284, between 2006 and 2008 conditioning on 2007 claims is 0.202, between 2006 and 2009 conditioning on 2007 and 2008 claims is 0.188, and so on. Two general patterns are noted from the table: First, the correlation decreases as one moves from the primary diagonal of the matrix toward its opposite corner in either upper or lower triangles,

indicating that the conditioning set is more informative as two observations become further apart in time; Second, the correlations along the same diagonal are of comparable size with some exceptions for the claims of other perils. These data characteristics support the D-vine specification with the stationarity assumption employed in the application in Section 4.

Table 3: Serial partial correlation for the total claim cost and the claim cost by peril

	Total						Water				
	2006	2007	2008	2009	2010		2006	2007	2008	2009	2010
2006	1	0.284	0.202	0.188	0.133	2006	1	0.243	0.188	0.157	0.086
2007	0.327	1	0.345	0.204	0.126	2007	0.261	1	0.313	0.194	0.164
2008	0.216	0.395	1	0.298	0.285	2008	0.195	0.339	1	0.284	0.211
2009	0.202	0.222	0.338	1	0.301	2009	0.163	0.205	0.303	1	0.245
2010	0.145	0.133	0.323	0.350	1	2010	0.087	0.172	0.226	0.264	1
	Fire						Other				
	2006	2007	2008	2009	2010		2006	2007	2008	2009	2010
2006	1	0.231	0.221	0.188	0.169	2006	1	0.206	0.227	0.172	0.152
2007	0.247	1	0.288	0.208	0.132	2007	0.216	1	0.186	0.152	0.049
2008	0.234	0.306	1	0.188	0.266	2008	0.236	0.197	1	0.192	0.160
2009	0.198	0.219	0.200	1	0.239	2009	0.177	0.158	0.201	1	0.200
2010	0.178	0.137	0.283	0.255	1	2010	0.160	0.051	0.170	0.222	1

3 Modeling Semicontinuous Longitudinal Data

3.1 Marginal Model

Let Y_{it} denote the cost (total or by peril) of claims for policyholder i $(= 1, \cdots, n)$ in year t $(= 1, \cdots, T)$. We consider a two-component mixture model to accommodate the mass probability at zero, the skewness, and the long tails of the distribution. Specifically, Y_{it} is assumed as being generated from a degenerate distribution at zero with probability p_{it} and being generated from a skewed and heavy tailed distribution $G_{it}(\cdot)$ defined on $(0, +\infty)$ with probability $1 - p_{it}$. Assuming independence between the degenerate distribution and the skewed heavy-tailed distribution, the resulting variable follows a mixed distribution. Let $F_{it}(\cdot)$ and $f_{it}(\cdot)$ denote its distribution function and density function, respectively. It is shown:

$$\begin{aligned} F_{it}(y) &= p_{it} + (1 - p_{it})G_{it}(y), \\ f_{it}(y) &= p_{it}I(y = 0) + (1 - p_{it})g_{it}(y). \end{aligned} \tag{1}$$

Here $I(\cdot)$ is the indicator function and g_{it} is the density function associated with G_{it}.

In the above formulation, the zero component models the probability of incurring claims, and the continuous component models the amount of claims given occurrence. Separating the frequency and severity allows for different sets of predictors as well as different effects of the same predictor on each component. This is a common practice in pricing non-life insurance contracts. Using property

insurance as an example, one can think that the probability of having claims is more related to the risk profile of the property, while the amount of payment is, to a great extent, determined at the adjuster's discretion.

For the claim frequency, we consider a logit specification due to the straightforward interpretability of model parameters:

$$\log\left(\frac{p_{it}}{1-p_{it}}\right) = \boldsymbol{x}_{1it}'\boldsymbol{\beta}_1$$

where $\boldsymbol{x}_{1it}$ represents the vector of explanatory variables and $\boldsymbol{\beta}_1$ denotes the corresponding regression coefficients to be estimated. For the claim severity, we employ the generalized beta of the second kind (GB2) distribution. See Shi (2014) for discussions of alternative strategies for handling skewness and heavy tails in insurance claims. The GB2 distribution was introduced by McDonald (1984) and has found extensive applications in the economics literature (McDonald and Xu (1995)). More recently, Frees and Valdez (2008) and Shi and Zhang (2015) considered an alternative parameterization and demonstrated its flexibility in fitting insurance claims. Following this line of studies, we consider the formulation:

$$g_{it}(y) = \frac{\exp(\kappa_1\omega_{it})}{y|\sigma|B(\kappa_1,\kappa_2)[1+\exp(\omega_{it})]^{\kappa_1+\kappa_2}} \tag{2}$$

where $\omega_{it} = (\ln y - \mu_{it})/\sigma$ and $B(\kappa_1,\kappa_2)$ is the Euler beta function. The GB2 is a member of the log location-scale family with location parameter μ_{it}, scale parameter σ, and shape parameters κ_1 and κ_2. With four parameters, the GB2 distribution is very flexible to model skewed and heavy-tailed data. For instance, $\kappa_1 > \kappa_2$ indicates right skewness and $\kappa_1 < \kappa_2$ left skewness. The rth moment is $\mathrm{E}(Y^r) = \exp(\mu_{it}r)B(\kappa_1 + r\sigma, \kappa_2 - r\sigma)/B(\kappa_1,\kappa_2)$ where $-\kappa_2 < r\sigma < \kappa_2$. The location parameter is further modeled as a linear combination of covariates to control for the observed heterogeneity $\mu_{it} = \boldsymbol{x}_{2it}'\boldsymbol{\beta}_2$, with $\boldsymbol{x}_{2it}$ and $\boldsymbol{\beta}_2$ being the vector of predictors and regression coefficients, respectively.

3.2 Dependence Model

3.2.1 General Framework

Consider a vector of random variables $\boldsymbol{Z} = (Z_1, \cdots, Z_m)'$ with each component following a mixed distribution. In this application, we focus on the zero inflated data that mimic the claim cost in non-life insurance. The idea is easily extended to the general mixed case. Let $\boldsymbol{z} = (z_1, \cdots, z_m)'$ denote a realization of $\boldsymbol{Z}$. Below we lay out a general framework to construct a high dimensional mixed distribution $f(z_1, \cdots, z_m)$ by using bivariate pair copulas as building blocks.

Let $\mathcal{V}_m$ denote a vine on m elements. A regular vine consists of $m-1$ trees $\mathcal{T}_l,\ l = 1, \cdots, m-1$, and $\mathcal{T}_l$ is connected by nodes $\mathcal{N}_l$ and edges $\mathcal{E}_l$. Edges in a tree become nodes in the next tree, i.e. $\mathcal{N}_l = \mathcal{E}_{l-1}$ $(l = 2, \cdots, m-1)$. If two nodes in tree $\mathcal{T}_l$ are joined by an edge, the corresponding edges in tree $\mathcal{T}_{l-1}$ share a node. Define edge set of $\mathcal{V}_m$ as $\mathcal{E}(\mathcal{V}_m) = \mathcal{E}_1 \cup \cdots \cup \mathcal{E}_{m-1}$. To develop the mixed vine, we adopt similar notations used in Panagiotelis et al. (2012). Let Z be a scale element of $\boldsymbol{Z}$ and $\boldsymbol{V}$ be a subset of $\boldsymbol{Z}$ satisfying $Z \notin \boldsymbol{V}$. Let V_h be any scalar element of $\boldsymbol{V}$ and $\boldsymbol{V}_{\backslash h}$ its

complement. Specify the conditional bivariate mixed distributions using copula:

$$
\begin{aligned}
&f_{Z,V_h|\boldsymbol{V}_{\setminus h}}(z, v_h|\boldsymbol{v}_{\setminus h}) \\
&= \begin{cases}
C_{Z,V_h;\boldsymbol{V}_{\setminus h}}\left(F_{Z|\boldsymbol{V}_{\setminus h}}(0|\boldsymbol{v}_{\setminus h}), F_{V_h|\boldsymbol{V}_{\setminus h}}(0|\boldsymbol{v}_{\setminus h})\right) & z=0, v_h=0 \\
f_{Z|\boldsymbol{V}_{\setminus h}}(z|\boldsymbol{v}_{\setminus h}) c_{1,Z,V_h;\boldsymbol{V}_{\setminus h}}\left(F_{Z|\boldsymbol{V}_{\setminus h}}(z|\boldsymbol{v}_{\setminus h}), F_{V_h|\boldsymbol{V}_{\setminus h}}(0|\boldsymbol{v}_{\setminus h})\right) & z>0, v_h=0 \\
f_{V_h|\boldsymbol{V}_{\setminus h}}(v_h|\boldsymbol{v}_{\setminus h}) c_{2,Z,V_h;\boldsymbol{V}_{\setminus h}}\left(F_{Z|\boldsymbol{V}_{\setminus h}}(0|\boldsymbol{v}_{\setminus h}), F_{V_h|\boldsymbol{V}_{\setminus h}}(v_h|\boldsymbol{v}_{\setminus h})\right) & z=0, v_h>0 \\
f_{Z|\boldsymbol{V}_{\setminus h}}(z|\boldsymbol{v}_{\setminus h}) f_{V_h|\boldsymbol{V}_{\setminus h}}(v_h|\boldsymbol{v}_{\setminus h}) c_{Z,V_h;\boldsymbol{V}_{\setminus h}}\left(F_{Z|\boldsymbol{V}_{\setminus h}}(z|\boldsymbol{v}_{\setminus h}), F_{V_h|\boldsymbol{V}_{\setminus h}}(v_h|\boldsymbol{v}_{\setminus h})\right) & z>0, v_h>0
\end{cases}
\end{aligned} \tag{3}
$$

where $C_{Z,V_h;\boldsymbol{V}_{\setminus h}}(u_1,u_2)$ and $c_{Z,V_h;\boldsymbol{V}_{\setminus h}}(u_1,u_2)$ are the bivariate copula and density function associated with conditional distributions $F_{Z|\boldsymbol{V}_{\setminus h}}$ and $F_{V_h|\boldsymbol{V}_{\setminus h}}$, respectively. And $c_{k,Z,V_h;\boldsymbol{V}_{\setminus h}}(u_1,u_2) = \partial C_{Z,V_h;\boldsymbol{V}_{\setminus h}}(u_1,u_2)/\partial u_k$, for $k=1,2$. For inference, we require the simplifying assumption that the copula does not directly rely on the conditioning set (see, for example, Haff et al. (2010) and Stoeber et al. (2013)).

To evaluate (3), we further derive the following generic conditional quantities:

$$
\begin{aligned}
f_{Z|\boldsymbol{V}}(z|\boldsymbol{v}) &= f_{Z|V_h,\boldsymbol{V}_{\setminus h}}(z|v_h,\boldsymbol{v}_{\setminus h}) \\
&= \begin{cases}
\dfrac{C_{Z,V_h;\boldsymbol{V}_{\setminus h}}\left(F_{Z|\boldsymbol{V}_{\setminus h}}(0|\boldsymbol{v}_{\setminus h}), F_{V_h|\boldsymbol{V}_{\setminus h}}(0|\boldsymbol{v}_{\setminus h})\right)}{F_{V_h|\boldsymbol{V}_{\setminus h}}(0|\boldsymbol{v}_{\setminus h})} & z=0, v_h=0 \\
\dfrac{f_{Z|\boldsymbol{V}_{\setminus h}}(z|\boldsymbol{v}_{\setminus h}) c_{1,Z,V_h;\boldsymbol{V}_{\setminus h}}\left(F_{Z|\boldsymbol{V}_{\setminus h}}(z|\boldsymbol{v}_{\setminus h}), F_{V_h|\boldsymbol{V}_{\setminus h}}(0|\boldsymbol{v}_{\setminus h})\right)}{F_{V_h|\boldsymbol{V}_{\setminus h}}(0|\boldsymbol{v}_{\setminus h})} & z>0, v_h=0 \\
c_{2,Z,V_h;\boldsymbol{V}_{\setminus h}}\left(F_{Z|\boldsymbol{V}_{\setminus h}}(0|\boldsymbol{v}_{\setminus h}), F_{V_h|\boldsymbol{V}_{\setminus h}}(v_h|\boldsymbol{v}_{\setminus h})\right) & z=0, v_h>0 \\
f_{Z|\boldsymbol{V}_{\setminus h}}(z|\boldsymbol{v}_{\setminus h}) c_{Z,V_h;\boldsymbol{V}_{\setminus h}}\left(F_{Z|\boldsymbol{V}_{\setminus h}}(z|\boldsymbol{v}_{\setminus h}), F_{V_h|\boldsymbol{V}_{\setminus h}}(v_h|\boldsymbol{v}_{\setminus h})\right) & z>0, v_h>0
\end{cases}
\end{aligned} \tag{4}
$$

and

$$
\begin{aligned}
F_{Z|\boldsymbol{V}}(z|\boldsymbol{v}) &= F_{Z|V_h,\boldsymbol{V}_{\setminus h}}(z|v_h,\boldsymbol{v}_{\setminus h}) \\
&= \begin{cases}
\dfrac{C_{Z,V_h;\boldsymbol{V}_{\setminus h}}\left(F_{Z|\boldsymbol{V}_{\setminus h}}(0|\boldsymbol{v}_{\setminus h}), F_{V_h|\boldsymbol{V}_{\setminus h}}(0|\boldsymbol{v}_{\setminus h})\right)}{F_{V_h|\boldsymbol{V}_{\setminus h}}(0|\boldsymbol{v}_{\setminus h})} & z=0, v_h=0 \\
\dfrac{C_{Z,V_h;\boldsymbol{V}_{\setminus h}}\left(F_{Z|\boldsymbol{V}_{\setminus h}}(z|\boldsymbol{v}_{\setminus h}), F_{V_h|\boldsymbol{V}_{\setminus h}}(0|\boldsymbol{v}_{\setminus h})\right)}{F_{V_h|\boldsymbol{V}_{\setminus h}}(0|\boldsymbol{v}_{\setminus h})} & z>0, v_h=0 \\
c_{2,Z,V_h;\boldsymbol{V}_{\setminus h}}\left(F_{Z|\boldsymbol{V}_{\setminus h}}(0|\boldsymbol{v}_{\setminus h}), F_{V_h|\boldsymbol{V}_{\setminus h}}(v_h|\boldsymbol{v}_{\setminus h})\right) & z=0, v_h>0 \\
c_{2,Z,V_h;\boldsymbol{V}_{\setminus h}}\left(F_{Z|\boldsymbol{V}_{\setminus h}}(z|\boldsymbol{v}_{\setminus h}), F_{V_h|\boldsymbol{V}_{\setminus h}}(v_h|\boldsymbol{v}_{\setminus h})\right) & z>0, v_h>0
\end{cases} \\
&= \begin{cases}
\dfrac{C_{Z,V_h;\boldsymbol{V}_{\setminus h}}\left(F_{Z|\boldsymbol{V}_{\setminus h}}(z|\boldsymbol{v}_{\setminus h}), F_{V_h|\boldsymbol{V}_{\setminus h}}(0|\boldsymbol{v}_{\setminus h})\right)}{F_{V_h|\boldsymbol{V}_{\setminus h}}(0|\boldsymbol{v}_{\setminus h})} & v_h=0 \\
c_{2,Z,V_h;\boldsymbol{V}_{\setminus h}}\left(F_{Z|\boldsymbol{V}_{\setminus h}}(z|\boldsymbol{v}_{\setminus h}), F_{V_h|\boldsymbol{V}_{\setminus h}}(v_h|\boldsymbol{v}_{\setminus h})\right) & v_h>0
\end{cases}
\end{aligned} \tag{5}
$$

Define

$$\tilde{f}_{Z,V_h|\boldsymbol{V}_{\setminus h}}(z, v_h|\boldsymbol{v}_{\setminus h}) := \frac{f_{Z,V_h|\boldsymbol{V}_{\setminus h}}(z, v_h|\boldsymbol{v}_{\setminus h})}{f_{Z|\boldsymbol{V}_{\setminus h}}(z|\boldsymbol{v}_{\setminus h}) f_{V_h|\boldsymbol{V}_{\setminus h}}(v_h|\boldsymbol{v}_{\setminus h})}, \tag{6}$$

one can express the joint distribution of $\boldsymbol{Z}$ using the bivariate building blocks as:

$$f_{\boldsymbol{Z}}(z_1, \cdots, z_m) = \prod_{j=1}^{m} f_{Z_j}(z_j) \prod_{[Z,V_h|\boldsymbol{V}_{\setminus h}] \in \mathcal{E}(\mathcal{V}_m)} \tilde{f}_{Z,V_h|\boldsymbol{V}_{\setminus h}}(z, v_h|\boldsymbol{v}_{\setminus h}). \tag{7}$$

Definition (6) is the ratio of the bivariate distribution to the product of marginals given the conditioning set. Thus, one can interpret (6) as the (conditional) "dependence ratio" with a ratio of one indicating conditional independence. Each ratio corresponds to one building block in the pair copula construction. Equation (7) shows that the joint distribution can be expressed as the product of marginals and the bivariate building blocks. Detailed discussion is provided in Appendix A.3. Formulation (7) provides a general framework for the pair copula construction in that both continuous and discrete vines can be viewed in the same framework as well. We articulate this point in more detail using the example of D-vine in Section 3.2.2.

3.2.2 Mixed D-Vine

For this application, we focus on a specific vine - D-vine. We use the term "mixed D-Vine" to refer to a D-Vine with a distribution that is a combination of a discrete and continuous components. Due to its simplicity, the D-vine is one of the most popular vine structures used in applied studies. An example of a D-vine on five elements is exhibited in Figure 2. The key feature of the D-vine is that the nodes of each tree only connect adjacent nodes. For instance, the nodes in the first tree represent ordered marginals, and the edges in each tree becomes the nodes in the next tree. Each edge corresponds to a (conditional) bivariate distribution that we construct using a parametric copula. The edges of the entire vine indicate the bivariate building blocks that contribute to the pair copula constructions.

In longitudinal data, a cross-sectional subjects are repeatedly observed over time. The temporal order makes D-vine a natural choice. Consider a mixed variables for T periods. The joint distribution of $(Z_1, \cdots, Z_T)$ can be expressed based on a D-vine as:

$$\begin{aligned} f_{\boldsymbol{Z}}(z_1, \cdots, z_T) &= f(z_T|z_{T-1}, \cdots, z_1) \times \cdots \times f(z_2|z_1) f(z_1) \\ &= \prod_{t=1}^{T} f_t(z_t) \prod_{t=2}^{T} \prod_{s=1}^{t-1} \tilde{f}_{s,t|(s+1):(t-1)}(z_s, z_t|z_{s+1}, \cdots, z_{t-1}). \end{aligned} \tag{8}$$

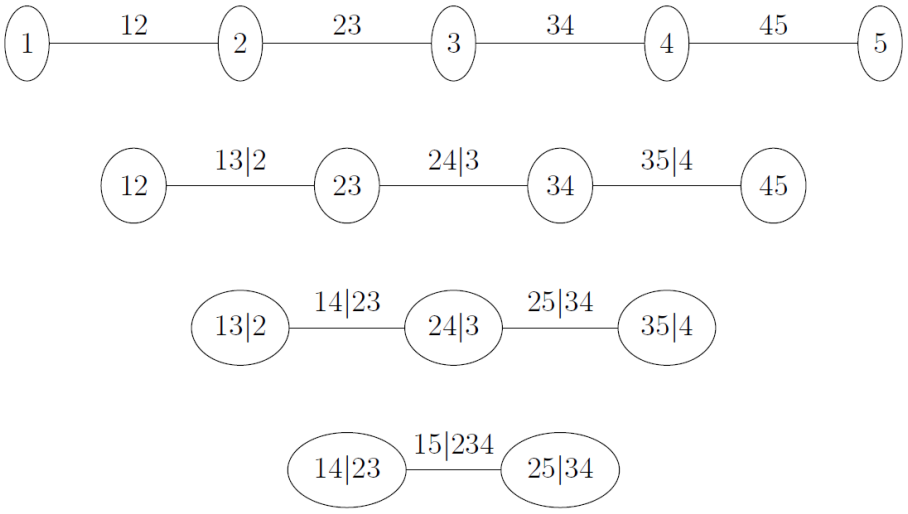

Figure 2: A 5-dimension D-vine

where using (6), we show:

$$\tilde{f}_{s,t|(s+1):(t-1)}(z_s, z_t|z_{s+1}, \cdots, z_{t-1})$$

$$= \begin{cases} \frac{C_{s,t;(s+1):(t-1)}\left(F_{s|(s+1):(t-1)}(0|z_{s+1}, \cdots, z_{t-1}), F_{t|(s+1):(t-1)}(0|z_{s+1}, \cdots, z_{t-1})\right)}{F_{s|(s+1):(t-1)}(0|z_{s+1}, \cdots, z_{t-1})F_{t|(s+1):(t-1)}(0|z_{s+1}, \cdots, z_{t-1})} & z_s = 0, z_t = 0 \\ \frac{c_{1,s,t;(s+1):(t-1)}\left(F_{s|(s+1):(t-1)}(z_s|z_{s+1}, \cdots, z_{t-1}), F_{t|(s+1):(t-1)}(0|z_{s+1}, \cdots, z_{t-1})\right)}{F_{t|(s+1):(t-1)}(0|z_{s+1}, \cdots, z_{t-1})} & z_s > 0, z_t = 0 \\ \frac{c_{2,s,t;(s+1):(t-1)}\left(F_{s|(s+1):(t-1)}(0|z_{s+1}, \cdots, z_{t-1}), F_{t|(s+1):(t-1)}(z_t|z_{s+1}, \cdots, z_{t-1})\right)}{F_{s|(s+1):(t-1)}(0|z_{s+1}, \cdots, z_{t-1})} & z_s = 0, z_t > 0 \\ c_{s,t;(s+1):(t-1)}\left(F_{s|(s+1):(t-1)}(z_s|z_{s+1}, \cdots, z_{t-1}), F_{t|(s+1):(t-1)}(z_t|z_{s+1}, \cdots, z_{t-1})\right) & z_s > 0, z_t > 0 \end{cases} \tag{9}$$

There are two points worth stressing. First, the decomposition in (8) is not unique. The order of these random variables determines pair copula building blocks and each decomposition corresponds to a graphical model with a specific vine structure. For a T dimensional vector, there are $\frac{T!}{2} \times 2^{\binom{T-2}{2}}$ possible vine trees, which points to a vine selection problem (see, for example Dißmann et al. (2013), Gruber et al. (2015), Panagiotelis et al. (2015)). We choose the D-vine due to the longitudinal nature of the application. Hence vine selection is not concern for this study. However, the other aspect of model selection - copula selection - is of more importance and we will discuss this issue in Section 3.3. Second, both continuous and discrete pair copula constructions can be viewed in this general framework. To be more specific, we recognize the following two cases that can be derived using (6):

(1) Continuous vine (Aas et al. (2009))

$$\begin{aligned}&\tilde{f}_{s,t|(s+1):(t-1)}(z_s, z_t|z_{s+1}, \cdots, z_{t-1})\\ =&c_{s,t;(s+1):(t-1)}\left(F_{s|(s+1):(t-1)}(z_s|z_{s+1}, \cdots, z_{t-1}), F_{t|(s+1):(t-1)}(z_t|z_{s+1}, \cdots, z_{t-1})\right)\end{aligned}$$

(2) Discrete vine (Panagiotelis et al. (2012))

$$\begin{aligned}&\tilde{f}_{s,t|(s+1):(t-1)}(z_s, z_t|z_{s+1}, \cdots, z_{t-1})\\ =&\frac{\sum_{i_1=0,1}\sum_{i_2=0,1}(-1)^{i_1+i_2}C_{s,t;(s+1):(t-1)}\left(F_{s|(s+1):(t-1)}(z_s-i_1|z_{s+1}, \cdots, z_{t-1}), F_{t|(s+1):(t-1)}(z_t-i_2|z_{s+1}, \cdots, z_{t-1})\right)}{f_{s|(s+1):(t-1)}(z_s|z_{s+1}, \cdots, z_{t-1})f_{t|(s+1):(t-1)}(z_t|z_{s+1}, \cdots, z_{t-1})}\end{aligned}$$

3.3 Inference

Due to the parametric nature of the proposed model, we employ likelihood-based method for estimation. Consider a portfolio of n policyholders, the total log likelihood function is

$$ll(\boldsymbol{\theta}, \boldsymbol{\zeta}) = \sum_{i=1}^{n}\sum_{t=1}^{T}\log f_{it}(y_{it}) + \sum_{i=1}^{n}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\log \tilde{f}_{i,s,t|(s+1):(t-1)}(y_{is}, y_{it}|y_{i,s+1}, \cdots, y_{i,t-1}) \tag{10}$$

where $f_{it}(\cdot)$ and $F_{it}(\cdot)$ are specified by (1), $\tilde{f}_{i,s,t|(s+1):(t-1)}(\cdot|\cdot)$ is specified by (9), and $\boldsymbol{\theta}$ and $\boldsymbol{\zeta}$ summarize the parameters in marginals and the mixed D-vine, respectively. Note that the model allows for unbalanced data provided that there are no intermittent missing values. The model can be estimated by two methods: joint maximum likelihood estimation (MLE) and inference function for margins (IFM) (see Joe (2005)). The joint MLE is a full likelihood approach and estimates all model parameters simultaneously. In a two-stage IFM, one estimates the marginal parameters ($\boldsymbol{\theta}$) from a separate univariate likelihood and then estimates the dependence parameters ($\boldsymbol{\zeta}$) from the multivariate likelihood with the marginal parameters given from the first stage. Compared with the joint MLE, the IFM is more computationally efficient by sacrificing the statistical efficiency. Therefore, the IFM is more practical for predictive applications where the statistical efficiency is of secondary concern. We examine both methods in Section A.4.

To implement the likelihood (10), one needs to evaluate the marginal densities and the bivariate building blocks (9) corresponding to each edge in the D-vine. We first calculate the marginal densities according to (1). Then we calculate (9) on a tree-by-tree basis from lower to higher orders. In the calculation of (9) for each tree, we use the copula of the current tree and the conditional cdf derived from the previous tree. An algorithm for evaluating the likelihood function for the mixed D-vine is provided in Appendix A.1.

In addition, we explore a sequential method that estimates and selects the bivariate copulas on a tree-by-tree basis. We start with the first tree, estimating the parameters and selecting the appropriate copulas from a given set of candidates. Fixing the parameters in the first tree, we then estimate the dependence parameters in the second tree for the candidate copulas and select the

optimal. We continue estimating parameters and selecting copulas for the next tree of a higher order while holding the parameters fixed in all previous trees. If an independence copula is selected for a certain tree, we then truncate the vine, i.e. assume conditional independence in all higher order trees (see, for example, Brechmann et al. (2012)). We use a heuristic procedure based on a commonly used model selection method Akaike information criterion (AIC) to select the copula. The sequential approach reduces the number of models to compare extensively and thus helps to fast select an appropriate model for applied studies. The benefit could be substantial in the case of big data or high dimensional dependence. In this application with a short panel of five-year observations, under the stationary assumption with nine candidate copulas, the sequential approach compares 9×4 different models in contrast to 9^4 models in an exhaustive search. The performance of the sequential method is investigated using simulation studies.

4 Application in Experience Rating

4.1 Model Fitting

We apply the proposed approach to the LGPIF claim data for the property coverage of building and contents. Separate models are fit for the total claim cost as well as the claim cost by peril. To summarize, the two-component mixture regression is used to accommodate the semicontinuous claim cost. The mass probability at zero is modeled using a logit regression and the amount of claims is modeled using a GB2 regression. Due to the limited number of predictors, we did not perform variable selection but instead included all available covariates in the two components. In the preliminary analysis, we explored the potential nonlinear effect of the coverage amount using scatter plot smoothing techniques (see, for instance, Ruppert et al. (2003)) and we found the linear term of coverage in log scale quite satisfactory. The estimation results are summarized in Table 4. Parameters are estimated by the IFM and standard errors are calculated using the Godambe information matrix.

Table 4: Estimates of the two-component mixture regression

Total Claim			Water			Fire			Other		
Logit	Est.	Std.	Logit	Est.	Std.	Logit	Est.	Std.	Logit	Est.	Std.
(Intercept)	2.776	0.158	(Intercept)	3.688	0.216	(Intercept)	3.985	0.247	(Intercept)	4.272	0.272
TypeCity	-1.139	0.164	TypeCity	-1.105	0.209	TypeCity	-1.025	0.237	TypeCity	-0.988	0.257
TypeCounty	-1.813	0.203	TypeCounty	-1.244	0.230	TypeCounty	-1.890	0.253	TypeCounty	-1.345	0.274
TypeSchool	-0.162	0.160	TypeSchool	0.095	0.210	TypeSchool	-0.137	0.237	TypeSchool	-0.589	0.253
TypeTown	-0.194	0.204	TypeTown	-0.680	0.264	TypeTown	0.054	0.354	TypeTown	-0.306	0.365
TypeVillage	-0.887	0.156	TypeVillage	-0.849	0.209	TypeVillage	-1.054	0.237	TypeVillage	-0.677	0.267
AC05	-0.327	0.170	AC05	-0.093	0.236	AC05	-0.262	0.236	AC05	-0.108	0.247
AC10	-0.266	0.150	AC10	-0.204	0.195	AC10	-0.182	0.206	AC10	-0.146	0.209
AC15	-0.273	0.087	AC15	-0.291	0.108	AC15	-0.098	0.116	AC15	-0.012	0.123
log(Coverage)	-0.453	0.034	log(Coverage)	-0.468	0.041	log(Coverage)	-0.481	0.044	log(Coverage)	-0.528	0.046
GB2	Est.	Std.	GB2	Est.	Std.	GB2	Est.	Std.	GB2		
(Intercept)	7.569	0.210	(Intercept)	7.234	0.257	(Intercept)	7.339	0.377	(Intercept)	7.306	0.424
TypeCity	-0.483	0.184	TypeCity	-0.263	0.237	TypeCity	-0.858	0.286	TypeCity	-0.538	0.337
TypeCounty	-0.412	0.199	TypeCounty	-0.607	0.257	TypeCounty	-0.685	0.297	TypeCounty	-0.724	0.358
TypeSchool	-0.438	0.186	TypeSchool	-0.613	0.248	TypeSchool	-0.720	0.292	TypeSchool	0.015	0.335
TypeTown	-0.034	0.239	TypeTown	-0.107	0.299	TypeTown	-0.342	0.398	TypeTown	-0.241	0.454
TypeVillage	-0.285	0.177	TypeVillage	-0.246	0.235	TypeVillage	-0.360	0.278	TypeVillage	-0.224	0.338
AC05	0.121	0.184	AC05	0.478	0.294	AC05	0.077	0.264	AC05	0.218	0.318
AC10	-0.239	0.158	AC10	-0.144	0.210	AC10	0.467	0.223	AC10	-0.555	0.267
AC15	-0.127	0.091	AC15	-0.009	0.123	AC15	0.078	0.125	AC15	-0.138	0.162
log(Coverage)	0.546	0.035	log(Coverage)	0.377	0.049	log(Coverage)	0.433	0.050	log(Coverage)	0.428	0.058
σ	0.868	0.127	σ	0.593	0.116	σ	0.791	0.190	σ	1.100	0.259
κ_1	1.352	0.310	κ_1	0.941	0.268	κ_1	1.670	0.691	κ_1	2.023	0.801
κ_2	1.039	0.224	κ_2	0.569	0.145	κ_2	0.893	0.297	κ_2	1.304	0.484

The results suggest that the entity type explains some heterogeneity in both claim frequency and severity. The effect of the alarm credit is a little counterintuitive which is to some extent explained by the estimation uncertainty. This counterintuitive effect could also imply some moral hazard issue. As anticipated, the odds of claim occurrence is higher for larger contract (due to higher exposure to risk), and so is the expected aggregate amount of claims. In the severity model, $\phi_1 > \phi_2$ in all the fitted GB2 distribution implies the positive skewness in the amount of claims. As indicated by the relation between ϕ_1 ϕ_2 and σ, their (theoretical) second moments do not even exist, which is consistent with the long tails in the distributions shown in Figure 1.

To demonstrate the goodness-of-fit of the GB2 distribution, we present in Figure 3 the *qq*-plots of the Cox-Snell residuals that is defined as $r_{it} = \Phi^{-1}(G_{it}(y_{it}))$. The match between the theoretical and empirical quantiles suggests the favorable fit of the GB2 distribution. The plots also indicate a slight lack of fit in the left tails except for the claims related to fire damages. However, for the purposes of ratemaking, we are more interested in the large claims that correspond to the right tails of the distribution. The left tails represent small claims and are less of a concern in this application.

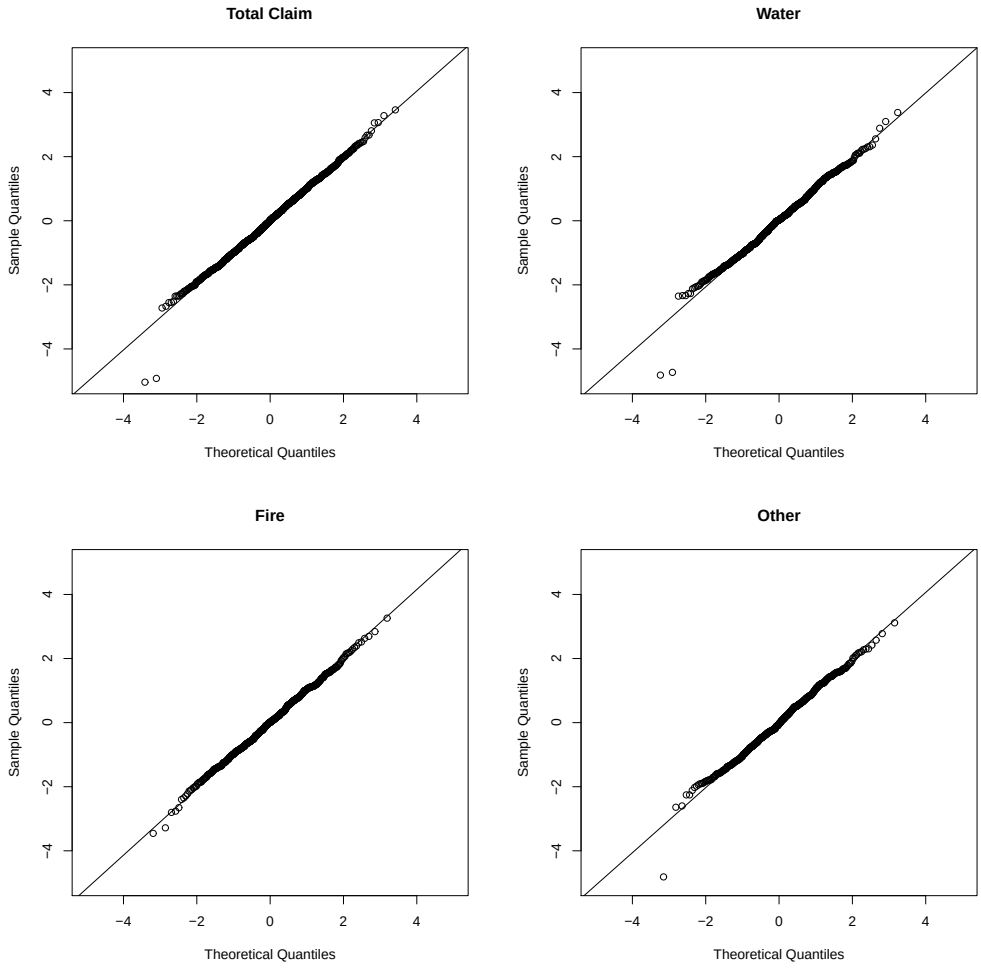

Figure 3: QQ plots of the GB2 distribution for total claims and claims by peril.

Pair copula constructions based on a mixed D-vine are used to accommodate the serial dependence among the longitudinal semicontinuous claim costs. We consider a candidate set of nine bivariate copulas as building block, including the Gaussian, Student's t, Gumbel, Clayton, Frank, Joe, survival Gumbel, survival Clayton, and survival Joe copulas. For the purpose of prediction, we further impose a stationarity assumption that all conditional pairs in a given tree share the same dependence. One should not view this assumption as a limitation of the proposed approach. In traditional longitudinal models, one would need a structured serial correlation such as autoregressive or exchangeable so as to borrow strength from past experience for future prediction. In the same spirit, the stationarity assumption is only required for prediction but not necessary for other types of statistical inference for the proposed longitudinal model.

Table 5 summarizes the selected bivariate copulas for the mixed D-vine, the estimated association parameters, and the corresponding Kendall's *tau*s for the total claim cost as well as the claim cost by peril. We followed the procedure in Section 3.3 to select the copula and to decide the optimal truncation. With five years of data, we have at most four trees in each of the mixed D-vines. For example, the mixed D-vine for the losses due to other perils is truncated at the third tree. The model is calibrated using the IFM. In general, the Kendall's *tau* decreases when moving from lower order trees to higher order trees. The decreasing pattern suggests that the conditioning set in higher order trees explains more of the association between the two nodes. This is consistent with the first principal of building vine trees that the (conditional) pairs with stronger association should receive higher priority. The reported Kendall's *tau* represents a partial relation in the same sense of the partial correlation. However, because of the discrete component in the marginal distributions, the inferred associations from the estimated copulas do not necessarily match the partial correlations from the data as reported in Table 3 (see Genest and Nešlehová (2007)), although they present a similar decreasing pattern.

Table 6 compares the goodness-of-fit statistics of the selected D-vine with two special cases, a fully truncated model and a fully simplified model. The former uses independence copula for all (conditional) pairs, and the latter uses Gaussian copula for all (conditional) pairs. It is not surprising that the mixed D-vine is superior to the independence copula, confirming the significant temporal association in the zero-inflated longitudinal data. When compared with the Gaussian copula, the favorable fit of the mixed D-vine (smaller AIC and BIC statistics) emphasizes the value added by the flexible dependence structure (such as asymmetric and non-linear association) embraced by the pair copula constructions. Such flexibility plays a crucial role in the dependence modeling for nonnormal outcomes such as heavy-tailed and discrete data.

4.2 Prediction

Experience rating is to incorporate policyholders' past claim experience into the future premiums. The mixed D-vine provides a natural structure to derive the predictive distribution, not just a point prediction, of the future claim cost. We stress that this is another advantage of using pair copula constructions for predictive applications. With elliptical copulas, it is not straightforward to derive

Table 5: Selected copulas for the mixed D-vine with estimated dependence

Total Claim	Copula	Est.	Std.	Kendall's *tau*
$\mathcal{T}_1$	Rotated Joe	1.440	0.062	0.199
$\mathcal{T}_2$	Rotated Joe	1.382	0.063	0.178
$\mathcal{T}_3$	Rotated Joe	1.274	0.066	0.135
$\mathcal{T}_4$	Clayton	0.214	0.097	0.097
Water	Copula	Est.	Std.	Kendall's *tau*
$\mathcal{T}_1$	Rotated Joe	1.962	0.143	0.347
$\mathcal{T}_2$	Rotated Joe	1.685	0.126	0.276
$\mathcal{T}_3$	Rotated Joe	1.535	0.135	0.231
$\mathcal{T}_4$	Rotated Joe	1.302	0.166	0.146
Fire	Copula	Est.	Std.	Kendall's *tau*
$\mathcal{T}_1$	Frank	1.376	0.268	0.150
$\mathcal{T}_2$	Rotated Joe	1.668	0.132	0.271
$\mathcal{T}_3$	Frank	1.229	0.324	0.135
$\mathcal{T}_4$	Rotated Joe	1.500	0.194	0.219
Other	Copula	Est.	Std.	Kendall's *tau*
$\mathcal{T}_1$	Rotated Joe	1.622	0.156	0.258
$\mathcal{T}_2$	Rotated Joe	1.614	0.159	0.255
$\mathcal{T}_3$	Gaussian	0.098	0.054	0.062

Table 6: Goodness-of-fit statistics of the mixed D-vine and its nested models

	Total		Water		Fire		Other	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
Truncated Model	39,708	39,859	21,191	21,341	18,584	18,735	16,701	16,851
Simplied Model	39,624	39,801	21,098	21,275	18,523	18,700	16,667	16,844
Mixed D-vine	39,561	39,737	21,036	21,213	18,496	18,673	16,658	16,834

the predictive distribution when there are discrete components in the marginals. For policyholder i, denoting $\boldsymbol{Y}_i = (Y_{i1}, \cdots, Y_{iT})'$, the conditional distribution of $Y_{i,T+1}$ given $\boldsymbol{Y}_i$ is shown as:

$$f_{Y_{i,T+1}|\boldsymbol{Y}_i}(y) = f_{i,T+1}(y) \prod_{t=2}^{T} \tilde{f}_{i,t,T+1|(t+1):T}(y_{it}, y|y_{i,t+1}, \cdots, y_{i,T}).$$

Here, $f_{i,T+1}(\cdot)$ and $\tilde{f}_{i,t,T+1|(t+1):T}(\cdot|\cdot)$ are defined by (1) and (9), respectively. The derivation of the predictive distribution relies on the conditional independence assumption between Y_{i1} and $Y_{i,T+1}$ given $Y_{i2}, \cdots, Y_{i,T}$. This is sensible given the pattern of the dependence in the mixed D-vine reported in Table 5. Detailed derivation of the predictive density is found in Appendix A.3. Insurers set pure premium as expected cost of the contract, thus the experience adjusted pure premium is $\mathrm{E}(Y_{i,T+1}|\boldsymbol{Y}_i = \boldsymbol{y}_i)$. The predictive mean can be estimated using the Monte Carlo simulation or the numerical integration.

The predictive performance is investigated using the hold-out sample of year 2011. It is well known that the usual loss functions are ill-suited for capturing the differences between the predicted values and the corresponding outcomes in the hold-out sample, due to the high proportion of zeros and the skewness and heavy tails in the distribution of the positive losses. Therefore we turn to alternative statistical measures - the ordered Lorenz curve and the associated Gini index - that have been developed in the recent literature (see Frees et al. (2012) and Frees et al. (2014)). The essential idea of the ordered Lorenz curve is to measure the discrepancy between the premium and loss distributions. Let $B(\boldsymbol{x})$ be the base premium and $P(\boldsymbol{x})$ be the competing premium, both depending on a set of exogenous variables $\boldsymbol{x}$. The ordered premium and loss distributions are defined based on the relativity $R(\boldsymbol{x}) = P(\boldsymbol{x})/B(\boldsymbol{x})$ as:

$$\hat{H}_P(s) = \frac{\sum_{i=1}^{n} B(\boldsymbol{x}_i) I(R(\boldsymbol{x}_i) \leq s)}{\sum_{i=1}^{n} B(\boldsymbol{x}_i)} \quad \text{and} \quad \hat{L}_P(s) = \frac{\sum_{i=1}^{n} y_i I(R(\boldsymbol{x}_i) \leq s)}{\sum_{i=1}^{n} y_i}.$$

The ordered Lorenz curve is the plot of $\left(\hat{H}_P(s), \hat{L}_P(s)\right)$. The 45-degree line, known as the line of equality, indicates the percentage of losses equals the percentage of premiums. The associated Gini index is defined as twice the area between the ordered Lorenz curve and the line of equality, and it may range over $(-1, 1)$. A curve below the line of equality suggests that the insurer could look to the competing premium to identify more profitable contracts.

We make two sets of validations. The first is to compare the proposed experience rated premium with some alternative bases. Table 7 reports the Gini indices associated with the ordered Lorenz curves under three scenarios. The upper panel uses a constant premium base, the middle panel uses the contract premium in year 2011 as the base, and the lower panel uses the non-experience adjusted premium base. Within each panel, we consider the prediction for the total claim cost as well as the claim cost by peril. Two methods are used to derive the prediction for the total claim cost. One directly predicts from the model for the total claim cost, the other predicts the claim cost for each peril and then aggregates them. The constant premium base means that the insurer does

not differentiate good risks and bad risks, and charges all policyholders the average cost. Hence it is not surprising to observe the large and significant Gini indices for all predictions in the upper panel. With both informative predictors and claim experience, insurers will achieve better risk segmentation. In practice, insurers use a finer-grained rating algorithm to classify and price the risk. Fortunately, the LGPIF data contain the actual contract premium for building and contents coverage as well as the premium for each peril. When compared with the contract premiums, the Gini indices become smaller as shown in the middle panel. However, the statistical significance suggests that the insurer can still identify profitable business when looking to the proposed experience adjusted rates. The lower panel demonstrates the importance of experience rating. The non-experience adjusted premium is calculated based on the independence assumption among the repeated observations over time. Indeed, the results are in line with our expectations. The significant positive Gini indices implies that the claim experience provides the insurer opportunities to cream skim (or cherry-pick the low-risk policyholders).

Table 7: Gini indices (percentage) with the constant, contract, and independence premium bases

	Total Claim		Claim by Peril		
Base	Total	By Peril	Water	Fire	Other
Constant	76.16	76.57	69.63	76.58	78.84
	(4.65)	(4.72)	(6.88)	(5.31)	(6.93)
Contract	31.34	33.69	29.29	16.89	38.92
	(8.74)	(10.04)	(8.99)	(19.02)	(9.30)
Independence	36.93	34.19	32.14	26.52	29.80
	(11.2)	(12.14)	(12.63)	(11.06)	(11.78)

Figure 4 displays the ordered Lorenz curves corresponding to Gini indices of the total claim cost prediction in Table 7. The left panel shows the case of the direct prediction and the right panel shows the case of the prediction by peril. For instance, the areas between the curves and the 45-degree line in the left panel are 0.38, 0.15, and 0.18 when using the constant premium, contract premium, and independence premium as bases, respectively.

The second set of validation is to compare the proposed rating algorithm with some off-the-shelf strategies for experience rating. The standard approach to incorporating past claims into future prediction is to use a random effect framework. We examine both the linear and generalized linear models. The former leads to the classic Bülhmann credibility premium. In the latter, we consider the industry benchmark - the Tweedie's compound Poisson model. We evaluate the performance of alternative approaches based on the prediction for the total claim cost for building and contents coverage. The total claim cost is either predicted directly or indirectly by aggregating the claims of different perils. Table 8 briefly summarizes the alternative predicting methods.

For model comparison, we calculate the Gini index matrix as in Table 9. The matrix summarizes the pair-wise Gini indices of the predictions from all candidate models when each of them is successively used as the base premium and the remaining as competing premiums. For example, the first row corresponds to the Gini indices using the experience-adjusted premium from the model

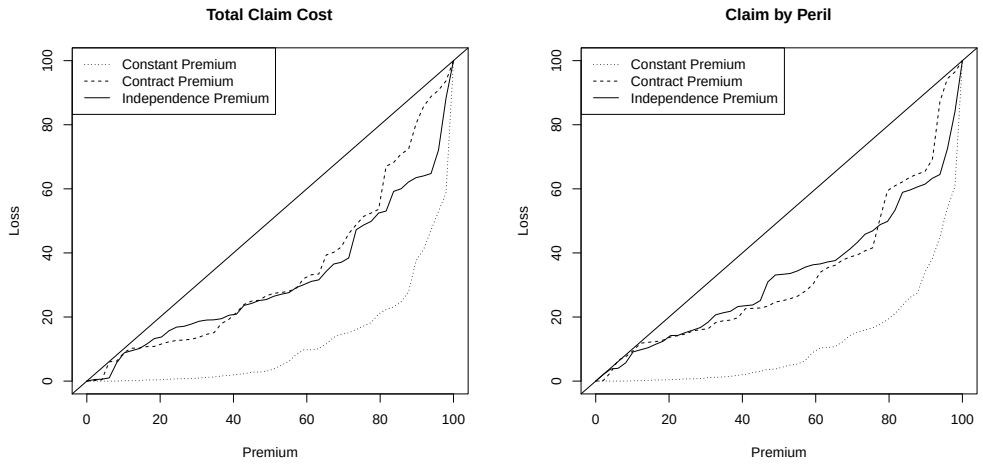

Figure 4: Ordered Lorenz curves using constant, contract, and independence premium bases. The left panel corresponds to the prediction of total claim cost and the right panel corresponds to the prediction of claim by peril.

Table 8: Description of alternative experience rating models

Model	Description
LM.t	Linear mixed model directly predicts the total claim cost
GLM.t	Tweedie mixed effects model directly predicts the total claim cost
COPULA.t	Mixed D-vine approach directly predicts the total claim cost
LM.p	Linear mixed model predicts the claim cost by peril
GLM.p	Tweedie mixed effects model predicts the claim cost by peril
COPULA.p	Mixed D-vine approach predicts the claim cost by peril

LM.t as the base. The sub matrices in the upper left and lower right corners are of our primary interest. One strategy for model selection is to use the proposed premium from the mixed D-vine model to challenge alternative premiums from the off-the-shelf rating algorithms. The upper left matrix compares models that directly predict the total claim cost. The Gini indices are 32.41 and 50.52 for the Bülhmann premium base and the Tweedie premium base, respectively. The lower right matrix compares models that predict the claim cost by peril and then aggregate them. The Gini indices are 29.90 and 56.34 for the Bülhmann premium base and the Tweedie premium base, respectively. The statistical significance confirms that the proposed mixed D-vine model leads to a greater separation among the observations.

Alternatively, to pick the "best" model, one could use a "minimax" strategy to select the base premium model that is the least vulnerable to the competing premium models. That is, we select the model that provides the smallest of the maximal Gini indices among the challenging premiums. For the direct prediction, the maximal Gini indices are 32.41, 50.52, and 14.89 when the base premium corresponds to the linear model, Tweedie model, and copula model, respectively. The copula approach has the smallest maximal Gini index, and hence is the least vulnerable to the alternative predictions. In a similar manner, when predicting by peril, the copula approach also has the smallest maximal Gini index of 15.48. When selecting from all six models, the "minmax" approach picks the predictions by peril from the mixed D-vine model. As a summary, we attribute the superior performance of the proposed method in experience rating to: 1) The two-component mixture model provides substantial flexibility in capturing the unique features in the insurance claim data, such as zero inflation, skewness, and thick tails; 2) The pair copula constructions based on the mixed D-vine allow us to accommodate a wide range of dependence including nonlinear and asymmetric relationship, while the traditional random effects framework limits the way past claims are incorporated into the future prediction.

Table 9: Gini index matrix for six alternative predictions

	LM.t	GLM.t	COPULA.t	LM.p	GLM.p	COPULA.p
LM.t	–	7.50	32.41	1.03	6.80	32.09
		(9.61)	(11.84)	(11.44)	(9.61)	(11.77)
GLM.t	29.22	–	50.52	33.34	-39.34	51.47
	(8.92)		(8.45)	(10.85)	(9.48)	(8.78)
COPULA.t	-4.45	14.89	–	-6.20	15.87	18.22
	(11.68)	(9.05)		(13.76)	(8.98)	(11.61)
LM.p	27.33	23.77	30.53	–	23.65	29.90
	(10.68)	(9.18)	(12.86)		(9.10)	(12.48)
GLM.p	35.50	47.02	55.32	40.19	–	56.34
	(9.30)	(9.24)	(8.44)	(10.69)		(8.66)
COPULA.p	-3.99	14.72	-8.05	-5.00	15.48	–
	(12.25)	(8.95)	(12.20)	(13.34)	(8.64)	

5 Discussion

Motivated by the experience rating practice in non-life insurance, this article introduced pair copula constructions based on the mixed D-vine for modeling the zero-inflated longitudinal insurance claim cost. The proposed approach is shown to achieve better risk segmentation and thus improve market efficiency. The data analysis emphasized the benefits of the mixed vine approach in both fitting the observations in the training sample and predicting the observations in the hold-out sample. The size of the insurance market itself justifies the contribution of the new method. Furthermore, pair copula constructions for mixed outcomes are expected to find applications in many other disciplines. To name a few, in marketing research, retailers are interested in consumers' purchasing behavior; in health care, care providers are interested in the patients' consumption of medical services; in climate studies, scientists are interested in the amount of precipitation. The variables of interest from all these examples are mixed measurements.

A natural extension is to generalize the current mixed D-vine framework for the semi-continuous longitudinal data to the multivariate context. In this application, the total claim cost for building and contents coverage is decomposed into claims by water, fire, and other perils. The losses caused by these perils were treated as three independent longitudinal outcomes separately. However, if the losses from different perils are correlated, it is arguable that one can further improve prediction and experience rating scheme by borrowing strength among perils. To provide intuition, we display in Table 10 the contemporaneous cross-sectional correlation among the claim costs from various perils. The upper and lower triangles report the Kendall's *tau* and Spearman's *rho*, respectively, and the correlations are calculated after controlling for the observed heterogeneity.

Table 10: Rank correlations among claims of different perils

	Water	Fire	Other
Water	1	0.233	0.228
Fire	0.250	1	0.193
Other	0.244	0.205	1

The strong relation suggests some joint modeling strategy. For the prediction purposes, the association that matters most is the lead-lag correlation across perils rather than contemporaneous correlation between perils. In experience rating, one hopes that the past claims in other perils could provide prediction lift for related perils. Although not reported, the strong correlations in Table 10 are indeed associated with significant lead-lag correlations across perils. Recently, Brechmann and Czado (2015) and Smith (2015) discussed possible strategies of constructing vine trees for multivariate time series, which shed some lights on the modeling of multivariate longitudinal data. This topic is being investigated in a separate work.

Appendix

A.1 Algorithm for Computing Likelihood

For a T-vector of zero-inflated outcomes $\boldsymbol{Z} = (Z_1, \cdots, Z_T)$, the likelihood of the mixed D-vine can be calculated using the following algorithm. Note that the algorithm can be extended to the general type of mixed outcomes. Without loss of generality, we focus on the outcome with a mass probability at zero.

1. For $t = 1, \cdots, T$, evaluate F_t and f_t with marginal model (1).
2. For $t = 1, \cdots, T-1$, evaluate $\tilde{f}_{t,t+1}$ using (9) with F_t and F_{t+1}.
3. For $t = 1, \cdots, T-2$:
   (a) Evaluate $F_{t|t+1}$ using (5) with $C_{t,t+1}$;
   (b) Evaluate $F_{t+2|t+1}$ using (4) with $C_{t+1,t+2}$;
   (c) Evaluate $\tilde{f}_{t,t+2;t+1}$ with $F_{t|t+1}$ and $F_{t+2|t+1}$ using (9).
4. For $t = 3, \cdots, T-1$ and $s = 1, \cdots, T-t$:
   (a) Using previous $C_{s,s+t-1;s+1,\cdots,s+t-2}$, calculate $F_{s|s+1,\cdots,s+t-1}$;
   (b) Using previous $C_{s+1,s+t;s+1,\cdots,s+t-1}$, calculate $F_{s+t|s+1,\cdots,s+t-1}$;
   (c) Evaluate $\tilde{f}_{s,s+t;s+1,\cdots,s+t-1}$ with $F_{s|s+1,\cdots,s+t-1}$ and $F_{s+t|s+1,\cdots,s+t-1}$ using (9).
5. The likelihood is calculated as $f_{1:T} = \prod_{t=1}^{T} f_t \prod_{t=2}^{T} \prod_{s=1}^{t-1} \tilde{f}_{s,t|(s+1):(t-1)}$.

A.2 Algorithm for Simulating from Mixed Vine

To simulate from the mixed D-vine, we define function g such that

$$F_{Z|\boldsymbol{V}_{\setminus h}}(z|\boldsymbol{v}_{\setminus h}) = g\left(F_{Z|\boldsymbol{V}}(z|\boldsymbol{v}), F_{V_h|\boldsymbol{V}_{\setminus h}}(v_h|\boldsymbol{v}_{\setminus h})\right).$$

To simplify the notation, let C denote the bivariate copula joining $F_{V_h|\boldsymbol{V}_{\setminus h}}$ and $F_{Z|\boldsymbol{V}_{\setminus h}}$, and c_2 denote its partial derivative with respect to the second argument. Then $g(a, b)$ is solution x to

$$\begin{cases} C(x,b)/b = a & v_h = 0 \\ c_2(x,b) = a & v_h > 0 \end{cases}$$

Here are some examples related to the simulation studies in Section 5. For Archimedean copulas, $C(u_1, u_2) = \psi^{-1}(\psi(u_1) + \psi(u_2))$, where ψ is the generator. We have

$$\begin{cases} C(x,b)/b = \dfrac{\psi^{-1}(\psi(x) + \psi(b))}{b} = a & v_h = 0 \\ c_2(x,b) = \dfrac{\psi'(b)}{\psi'(\psi^{-1}(\psi(x) + \psi(b)))} = a & v_h > 0 \end{cases}$$

Solving for x, one obtains

$$g(a,b) = \begin{cases} \psi^{-1}(\psi(ab) - \psi(b)) & v_h = 0 \\ \psi^{-1}[\psi((\psi')^{-1}(\psi'(b)/(a))) - \psi(b)] & v_h > 0 \end{cases}$$

As another example, for the survival Archimedean copulas $C(u_1, u_2) = u_1 + u_2 - 1 + \psi^{-1}(\psi(1 - u_1) + \psi(1 - u_2))$, we have

$$D_2(x,b) = 1 - \frac{\psi'(1-b)}{\psi'(\psi^{-1}(\psi(1-x) + \psi(1-b)))}.$$

The solution to $D_2(x,b) = a$ is of the form

$$1 - \psi^{-1}[\psi((\psi')^{-1}(\psi'(1-b)/(1-a))) - \psi(1-b)].$$

In the simulation, we use the survival Joe copula, where $\psi(t) = -\ln[1-(1-t)^\theta]$, $\psi^{-1}(t) = 1-(1-e^{-t})^{1/\theta}$, $\psi'(t) = -\frac{\theta(1-t)^{\theta-1}}{1-(1-t)^\theta}$. And $(\psi'(t))^{-1}$ is calculated with numerical root solve.

Below are simulation steps for the model where the marginal is a logit-GB2 mixture regression defined by (1):

1. For $t = 1, \cdots, T$, calculate zero probabilities $F_t(0)$ using the logit model, and calculate location parameter μ_t in the GB2 distribution G_t.

2. Generate $u_1, \cdots, u_T$ following T-dimensional independent $Uniform\ (0,1)$.

3. If $u_1 < F_1(0)$, set $y_1 = 0$ and $F_1(y_1) = F_1(0)$;

   Otherwise, set $F_1(y_1) = u_1$. Then y_1 is solution to $F_1(0) + (1 - F_1(0))G_1(y_1) = u_1$, which has closed form solution

   $$y_1 = \left(\text{qf}\left(\frac{F_1(y_1) - F_1(0)}{1 - F_1(0)}, \text{df1} = 2\kappa_1, \text{df2} = 2\kappa_2\right)\right)^\sigma \exp(\mu_1)\left(\frac{\kappa_1}{\kappa_2}\right)^\sigma$$

   where qf is quantile of F-distribution with degrees of freedom df1 and df2.

4. (a) Calculate $F_{2|1}(0)$ using copula C_{12} with $F_2(0)$ and $F_1(y_1)$.

   If $u_2 < F_{2|1}(0)$, set $y_2 = 0$ and $F_2(y_2) = F_2(0)$.

   Otherwise, set $F_{2|1}(y_2) = u_2$, and calculate $F_2(y_2) = g(F_{2|1}(y_2), F_1(y_1))$. Then

   $$y_2 = \left(\text{qf}\left(\frac{F_2(y_2) - F_2(0)}{1 - F_2(0)}, \text{df1} = 2\kappa_1, \text{df2} = 2\kappa_2\right)\right)^\sigma \exp(\mu_2)\left(\frac{\kappa_1}{\kappa_2}\right)^\sigma.$$

   (b) Calculate $F_{1|2}(y_1)$ using C_{12} with $F_1(y_1)$ and $F_2(y_2)$.

5. For $t = 3, \cdots, T$

(a) Calculate $F_{t|t-1}(0)$ using $F_t(0)$ and $F_{t-1}(y_{t-1})$ with copula $C_{t-1,t}$.

(b) For $s = 2, \cdots t-1$, calculate conditional probability at zero $F_{t|t-1,\cdots,t-s}(0)$ using copula $C_{t,t-s|t-1,\cdots,t-s+1}$ with $F_{t|t-1,\cdots,t-s+1}(0)$ and $F_{t-s|t-1,\cdots,t-s+1}(y_{t-s})$.

(c) If $u_t < F_{t|1,\cdots,t-1}(0)$, set $y_t = 0$ and $F_t(y_t) = F_t(0)$; Otherwise

i. $F_{t|1,\cdots,t-1}(y_t) = u_t$;

ii. If $t > 3$, for $s = 2, \cdots t-1$, calculate $F_{t|t-1,\cdots,s}(y_t) = g(F_{t|t-1,\cdots s-1}(y_t), F_{s-1|t-1,\cdots,s}(y_{s-1}))$ with copula $C_{t,s-1|s,\cdots,t-1}$;

iii. Calculate $F_t(y_t) = g(F_{t|t-1}(y_t), F_{t-1}(y_{t-1})$ with copula $C_{t-1,t}$, then

$$y_t = \left(\text{qf}\left(\frac{F_t(y_t) - F_t(0)}{1 - F_t(0)}, \text{df1} = 2\kappa_1, \text{df2} = 2\kappa_2\right)\right)^{\sigma} \exp(\mu_t) \left(\frac{\kappa_1}{\kappa_2}\right)^{\sigma}.$$

(d) If $t < T$, calculate

i. $F_{t-1|t}(y_{t-1})$ using $F_t(y_t)$ and $F_{t-1}(y_{t-1})$ with copula $C_{t-1,t}$;

ii. For $s = t-2, \cdots, 1$, calculate $F_{s|s+1,\cdots,t}(y_s)$ with $F_{s|s+1,\cdots,t-1}(y_s)$ and $F_{t|s+1,\cdots,t-1}(y_t)$ using copula $C_{s,t|s+1,\cdots,t-1}$.

A.3 Joint and Predictive Distributions Using D-vine

To derive (7), one can decompose the joint density of $\boldsymbol{Z} = (Z_1, \cdots, Z_m)$ into conditional densities. The decomposition is not unique depending on the vine structure. Without loss of generality, we write:

$$f_{\boldsymbol{Z}}(z_1, \cdots, z_m) = f(z_m|z_1, \cdots, z_{m-1}) f(z_{m-1}|z_1, \cdots, z_{m-2}) \cdots f(z_2|z_1) f(z_1).$$

From (6), we calculate each conditional density using

$$f_{Z|V_h, \boldsymbol{V}_{\setminus h}}(z|v_h, \boldsymbol{v}_{\setminus h}) = \tilde{f}_{Z,V_h|\boldsymbol{V}_{\setminus h}} f_{Z|\boldsymbol{V}_{\setminus h}}(z|\boldsymbol{v}_{\setminus h}). \tag{11}$$

By choosing conditioning variable V_h according to the regular vine structure recursively, the likelihood can be expressed into (7).

Consider a D-vine for longitudinal data $\boldsymbol{Z} = (Z_1, \cdots, Z_T)$. Using relation (11) repeatedly, we show

$$\begin{aligned}
f(z_t|z_{t-1}, \cdots, z_1) &= \tilde{f}_{1,t|2:(t-1)}(z_1, z_t|z_2, \cdots, z_{t-1}) f(z_t|z_2, \cdots, z_{t-1}) \\
&= \tilde{f}_{1,t|2:(t-1)}(z_1, z_t|z_2, \cdots, z_{t-1}) \tilde{f}_{2,t|3:(t-1)}(z_2, z_t|z_3, \cdots, z_{t-1}) f(z_t|z_3, \cdots, z_{t-1}) \\
&\vdots \\
&= f(z_t) \prod_{s=1}^{t-1} \tilde{f}_{s,t|(s+1):(t-1)}(z_s, z_t|z_{s+1}, \cdots, z_{t-1}).
\end{aligned}$$

Therefore,

$$
\begin{aligned}
f_{\boldsymbol{Z}}(z_1, \cdots, z_T) &= f(z_1) \prod_{t=2}^{T} f(z_t | z_{t-1}, \cdots, z_1) \\
&= \prod_{t=1}^{T} f_t(z_t) \prod_{t=2}^{T} \prod_{s=1}^{t-1} \tilde{f}_{s,t|(s+1):(t-1)}(z_s, z_t | z_{s+1}, \cdots, z_{t-1}).
\end{aligned}
$$

The predictive density of the outcome in period $T+1$, Z_{T+1}, given $\boldsymbol{Z}$ is:

$$
\begin{aligned}
f_{Z_{T+1}|\boldsymbol{Z}}(z_{t+1}) &= f_{T+1|1:T}(z_{t+1} | z_1, \cdots, z_t) \\
&= \frac{\prod_{t=1}^{T+1} f_t(z_t) \prod_{t=2}^{T+1} \prod_{s=1}^{t-1} \tilde{f}_{s,t|(s+1):(t-1)}(z_s, z_t | z_{s+1}, \cdots, z_{t-1})}{\prod_{t=1}^{T} f_t(z_t) \prod_{t=2}^{T} \prod_{s=1}^{t-1} \tilde{f}_{s,t|(s+1):(t-1)}(z_s, z_t | z_{s+1}, \cdots, z_{t-1})} \\
&= f_{T+1}(z_{T+1}) \prod_{s=1}^{T} \tilde{f}_{s,T+1|(s+1):T}(z_s, z_{T+1} | z_{s+1}, \cdots, z_T) \\
&= f_{T+1}(z_{T+1}) \prod_{s=2}^{T} \tilde{f}_{s,T+1|(s+1):T}(z_s, z_{T+1} | z_{s+1}, \cdots, z_T).
\end{aligned}
$$

The last equality is based on the assumption that Z_1 and Z_{T+1} are conditionally independent given $(Z_2, \cdots, Z_T)$.

A.4 Simulation

This section investigates two issues using simulated data. The first experiment is to compare estimations from the IFM and the joint MLE. The second experiment is to explore the performance of the sequential copula selection algorithm. In the simulation, we set $T = 4$, that is, each policyholder is repeatedly observed for four years. Data are generated from the multivariate model based on the mixed D-vine with marginals from the two-component mixture regression. Specifically, the joint distribution of $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, Y_{i3}, Y_{i4})$ is specified by (8) and (9) where all the bivariate copulas $C_{s,t;(s+1):(t-1)}$ are assumed to be the survival Joe copula. We further assume that the pair copulas in the same tree are identical. There are three trees in total and the associated dependence parameters in the copulas are denoted by $\boldsymbol{\zeta} = (\zeta_1, \zeta_2, \zeta_3)$. The marginal distribution of Y_{it} is specified by (1) with

$$
\begin{aligned}
\text{logit}(p_{it}) &= \beta_0 + \beta_1 X_{1,it} + \beta_2 X_{2,i}, \\
\mu_{it} &= \gamma_0 + \gamma_1 X_{1,it} + \gamma_2 X_{2,i}.
\end{aligned}
$$

Here, X_{1t} is time-varying and X_2 is time constant. The time-varying variable is generated independently over time, and both covariates are generated independently across subjects. To allow for both continuous and discrete predictors, we further assume $X_{1t} \sim N(0, 1)$ and $X_2 \sim Bernoulli(0.4)$. The scale and shape parameters in the GB2 distribution mimic those from the total claim model

and are selected to generate a right skewed and heavy tailed distribution. In particular, the second moment of GB2 does not even exist with the specified parameters.

We consider three scenarios corresponding to different levels of dependence that we quantify using Kendall's *tau*. In the case of weak dependence, Kendall's *tau*s are set to be 0.3, 0.2, and 0.1 for the first, second, and third trees, respectively. They are 0.6, 0.4, and 0.2 for the case of moderate dependence, and 0.9, 0.6, and 0.3 for the case of strong dependence. It is sensible to specify a decreasing pattern of dependence because the conditioning set contains more information when one moves from lower to higher trees. The parameters in the bivariate copulas are specified to match the Kendall's *tau*. Furthermore, the parameters in the marginal regressions are specified to mimic the percentage of zeros and the heavy tails in the insurance claim data. The detailed algorithm for generating data from the proposed model is provided in Appendix A.2.

A.4.1 Parameter Estimation

With the simulated data, we estimate the true model using both the IFM and the joint MLE. Results are summarized in Tables 11, 12, and 13 for scenarios of low, medium, and high dependence, respectively. For each case, we consider different sample sizes (number of policyholders) and present the results for $n = 500$ and $n = 1000$. The reported results are based on 500 replications.

Within each table, we examine the point estimate and the associated uncertainty. We first report the mean estimate and the relative bias for all model parameters. The average estimates are close to the true parameters, which is consistent with the associated small relative bias. Not surprisingly, increasing sample size reduces the estimation bias that is expected to shrink to zero when the sample size is large enough. Next, we examine the estimation uncertainty. Relevant quantities reported in the tables are the average standard error and the nominal standard deviation for each parameter. In each replication, the standard error is calculated based on the Godambe matrix and the nominal standard deviation is derived from the replicated point estimates. The two statistics are comparable, indicating the accuracy of the uncertainty estimates. An examination of the relation between the standard errors under different sample sizes confirms the $\sqrt{n}$ convergence rate. One notices the relatively large bias and uncertainty in the shape parameters of the GB2 distribution. This is due to the long tails of the distribution. Note that the second moment does not even exist for the GB2 distribution used in the simulation.

The effect of dependence on parameter estimation is demonstrated through comparisons across tables. As anticipated, higher dependence results in larger bias and larger standard error. This is because the effective sample size is smaller when the dependence is stronger. It is also noted that such effect is not linear in that the difference in estimates under small and large sample sizes is more pronounced in the scenario of strong dependence than the weak dependence.

Finally, when comparing the estimation of the IFM with the joint MLE, one notes that the two estimating methods produce very consistent results in terms of bias and uncertainty for low and medium dependence cases. The difference between the two methods becomes more noticeable in the high dependence scenario, especially in the dependence parameter. However, such differences

Table 11: Simulation under low dependence ($\tau = 0.3, 0.2, 0.1$)

	IFM				Joint MLE			
	Est.	Rel.Bias	S.E.	Nom.SD	Est.	Rel.Bias	S.E.	Nom.SD
				$n = 500$				
$\beta_0 = 2$	1.998	-0.001	0.093	0.100	2.000	0.000	0.101	0.100
$\beta_1 = -1$	-1.007	0.007	0.067	0.069	-1.009	0.009	0.065	0.065
$\beta_2 = -2$	-2.010	0.005	0.121	0.149	-2.012	0.006	0.142	0.147
$\gamma_0 = 5$	4.952	-0.010	0.316	0.423	4.949	-0.010	0.323	0.445
$\gamma_1 = 1$	1.002	0.002	0.059	0.061	1.002	0.002	0.058	0.058
$\gamma_2 = -1$	-0.993	-0.007	0.117	0.124	-0.993	-0.007	0.119	0.122
$\sigma = 0.8$	0.861	0.076	0.297	0.324	0.864	0.080	0.308	0.335
$\kappa_1 = 1.2$	1.597	0.331	1.255	2.232	1.634	0.362	1.389	2.559
$\kappa_2 = 1$	1.206	0.206	0.687	0.794	1.218	0.218	0.771	0.880
$\zeta_1 = 1.77$	1.763	-0.005	0.132	0.137	1.767	-0.003	0.139	0.137
$\zeta_2 = 1.44$	1.435	-0.006	0.121	0.124	1.437	-0.005	0.123	0.125
$\zeta_3 = 1.19$	1.202	0.006	0.144	0.142	1.202	0.006	0.144	0.143
				$n = 1000$				
$\beta_0 = 2$	2.007	0.004	0.066	0.068	2.008	0.004	0.071	0.068
$\beta_1 = -1$	-1.003	0.003	0.048	0.049	-1.005	0.005	0.046	0.047
$\beta_2 = -2$	-2.009	0.005	0.086	0.101	-2.011	0.005	0.101	0.099
$\gamma_0 = 5$	4.968	-0.006	0.178	0.178	4.968	-0.006	0.177	0.177
$\gamma_1 = 1$	1.001	0.001	0.042	0.043	1.001	0.001	0.041	0.041
$\gamma_2 = -1$	-0.998	-0.002	0.083	0.087	-0.998	-0.002	0.085	0.086
$\sigma = 0.8$	0.839	0.049	0.181	0.191	0.839	0.049	0.181	0.191
$\kappa_1 = 1.2$	1.355	0.129	0.480	0.532	1.355	0.129	0.479	0.532
$\kappa_2 = 1$	1.097	0.097	0.351	0.376	1.097	0.097	0.351	0.376
$\zeta_1 = 1.77$	1.770	-0.001	0.094	0.102	1.771	-0.001	0.098	0.101
$\zeta_2 = 1.44$	1.442	-0.001	0.086	0.090	1.442	-0.001	0.087	0.090
$\zeta_3 = 1.19$	1.199	0.004	0.100	0.099	1.199	0.004	0.101	0.099

Table 12: Simulation under medium dependence ($\tau = 0.6, 0.4, 0.2$)

	IFM				Joint MLE			
	Est.	Rel.Bias	S.E.	Nom.SD	Est.	Rel.Bias	S.E.	Nom.SD
				$n = 500$				
$\beta_0 = 2$	1.997	-0.002	0.093	0.116	1.999	-0.001	0.112	0.110
$\beta_1 = -1$	-1.007	0.007	0.067	0.073	-1.008	0.008	0.059	0.058
$\beta_2 = -2$	-2.007	0.003	0.121	0.174	-2.011	0.005	0.154	0.158
$\gamma_0 = 5$	4.932	-0.014	0.319	0.405	4.938	-0.012	0.304	0.353
$\gamma_1 = 1$	1.004	0.004	0.059	0.062	1.003	0.003	0.049	0.049
$\gamma_2 = -1$	-0.983	-0.017	0.117	0.136	-0.985	-0.015	0.120	0.126
$\sigma = 0.8$	0.866	0.083	0.302	0.339	0.865	0.082	0.299	0.325
$\kappa_1 = 1.2$	1.607	0.339	1.324	2.234	1.564	0.304	1.167	1.710
$\kappa_2 = 1$	1.221	0.221	0.712	0.866	1.217	0.217	0.713	0.860
$\zeta_1 = 3.83$	3.787	-0.010	0.260	0.315	3.815	-0.003	0.313	0.306
$\zeta_2 = 2.22$	2.209	-0.004	0.176	0.187	2.222	0.001	0.182	0.188
$\zeta_3 = 1.44$	1.458	0.010	0.168	0.174	1.462	0.013	0.170	0.179
				$n = 1000$				
$\beta_0 = 2$	2.005	0.003	0.066	0.079	2.006	0.003	0.079	0.075
$\beta_1 = -1$	-1.001	0.001	0.048	0.051	-1.003	0.003	0.042	0.042
$\beta_2 = -2$	-2.007	0.004	0.086	0.117	-2.005	0.003	0.109	0.105
$\gamma_0 = 5$	4.978	-0.004	0.172	0.177	4.981	-0.004	0.169	0.166
$\gamma_1 = 1$	1.003	0.003	0.042	0.043	1.003	0.003	0.035	0.036
$\gamma_2 = -1$	-0.997	-0.003	0.083	0.092	-0.997	-0.003	0.085	0.086
$\sigma = 0.8$	0.821	0.026	0.175	0.169	0.820	0.025	0.173	0.165
$\kappa_1 = 1.2$	1.297	0.081	0.447	0.463	1.291	0.076	0.436	0.437
$\kappa_2 = 1$	1.061	0.061	0.332	0.327	1.059	0.059	0.327	0.319
$\zeta_1 = 3.83$	3.820	-0.002	0.186	0.243	3.834	0.002	0.223	0.229
$\zeta_2 = 2.22$	2.222	0.001	0.125	0.135	2.228	0.004	0.129	0.135
$\zeta_3 = 1.44$	1.454	0.007	0.118	0.123	1.456	0.009	0.119	0.124

Table 13: Simulation under high dependence ($\tau = 0.9, 0.6, 0.3$)

	IFM				Joint MLE			
	Est.	Rel.Bias	S.E.	Nom.SD	Est.	Rel.Bias	S.E.	Nom.SD
				$n = 500$				
$\beta_0 = 2$	2.007	0.003	0.093	0.149	2.002	0.001	0.108	0.108
$\beta_1 = -1$	-1.012	0.012	0.068	0.082	-1.008	0.008	0.043	0.044
$\beta_2 = -2$	-2.016	0.008	0.122	0.205	-2.011	0.005	0.128	0.133
$\gamma_0 = 5$	4.942	-0.012	0.347	0.395	4.962	-0.008	0.216	0.222
$\gamma_1 = 1$	1.008	0.008	0.059	0.064	1.003	0.003	0.025	0.025
$\gamma_2 = -1$	-0.976	-0.024	0.118	0.161	-0.988	-0.012	0.092	0.098
$\sigma = 0.8$	0.906	0.132	0.331	0.372	0.871	0.088	0.220	0.224
$\kappa_1 = 1.2$	1.671	0.392	1.294	1.513	1.468	0.223	0.671	0.740
$\kappa_2 = 1$	1.346	0.346	0.906	1.147	1.196	0.196	0.488	0.524
$\zeta_1 = 18.74$	17.643	-0.058	1.018	1.802	18.884	0.008	1.552	1.482
$\zeta_2 = 3.83$	3.599	-0.059	0.277	0.335	3.837	0.003	0.294	0.307
$\zeta_3 = 1.77$	1.684	-0.050	0.188	0.206	1.800	0.016	0.205	0.210
				$n = 1000$				
$\beta_0 = 2$	2.010	0.005	0.066	0.097	2.005	0.002	0.076	0.073
$\beta_1 = -1$	-1.003	0.003	0.048	0.054	-1.001	0.001	0.030	0.030
$\beta_2 = -2$	-2.013	0.006	0.086	0.133	-2.004	0.002	0.090	0.084
$\gamma_0 = 5$	4.972	-0.006	0.178	0.188	4.986	-0.003	0.130	0.127
$\gamma_1 = 1$	1.001	0.001	0.042	0.049	1.001	0.001	0.018	0.018
$\gamma_2 = -1$	-0.995	-0.005	0.083	0.109	-0.996	-0.004	0.065	0.062
$\sigma = 0.8$	0.836	0.045	0.181	0.206	0.818	0.022	0.134	0.139
$\kappa_1 = 1.2$	1.348	0.123	0.480	0.562	1.276	0.064	0.348	0.364
$\kappa_2 = 1$	1.100	0.100	0.354	0.418	1.052	0.052	0.264	0.282
$\zeta_1 = 18.74$	18.121	-0.033	0.733	1.262	18.822	0.004	1.091	1.045
$\zeta_2 = 3.83$	3.691	-0.035	0.200	0.245	3.831	0.001	0.208	0.215
$\zeta_3 = 1.77$	1.724	-0.027	0.137	0.148	1.801	0.016	0.144	0.148

can eventually be washed out by increasing sample size. The practical implication is that the IFM offers substantial computational efficiency without sacrificing the statistical efficiency, while one should be cautious if the dependence is extremely high.

A.4.2 Copula Specification

The above simulation shows the advantage of using the IFM for estimation provided that the model is correctly specified. This section provides guidance on how to correctly specify the bivariate copulas in the D-vine. Clearly it is not practical to perform an exhaustive search especially when the dimension is high or the candidate pool is large. Even for the current application where all copulas in the same tree are assumed to be identical, an exhaustive search still requires a fair amount of computation.

We explore the sequential approach that is proposed in Section 3.3. Specifically, we consider in this experiment a pool of four candidate copulas, i.e. the Clayton, Frank, Gumbel, and the survival Joe. This set of copulas is chosen so that one has the potential to capture different types of dependence structure, such as tail dependence and asymmetric dependence. For instance, the Frank copula implies a symmetric dependence, the Gumbel copula allows for upper tail dependence, and the other two capture lower tail dependence. To visualize the differences in the dependence structure implied by these candidates, we exhibit the elliptical contours of their densities in Figure 5. The low, medium, and high associations in the figure are defined by Kendall's *tau* being 0.2, 0.5, and 0.8 respectively. One noticeable observation from Figure 5 is that besides both the Clayton and survival Joe copulas accommodate lower tail dependence, the entire dependence structures implied by the two copulas are nearly indistinguishable. To recap, the data are generated from the mixed D-vine where the survival Joe copula is used for all trees and identical copulas are used for each tree. The marginals are from a two-component mixture model featured with a logit regression and a GB2 regression.

In each replication, we use the sequential algorithm to specify the copula on a tree-by-tree basis. For comparison, we also use an exhaustive search to specify the copula for all trees simultaneously. Note that for this experiment, the computational complexity is 4×3 for the former case and 4^3 for the latter case. As a robustness check, we consider the three scenarios of different levels of dependence that are examined in Section A.4.1. The corresponding coverage probabilities based on 500 replications are displayed in Tables 14, 15, and 16. The experiment is also implemented under different sample sizes and the results are consistent.

First, the specification based on the sequential approach is comparable to the exhaustive search, indicating that the cumulative bias from lower to higher trees are ignorable, at least in the low dimension context. We stress that the sole purpose of this experiment is to justify the practical value of the sequential specification algorithm for the mixed D-vines compared to the benchmark of the joint specification. Model selection for regular vines is out of the scope of this study. Second, both methods effectively distinguish the true copula from the Frank and Gumbel copulas. This is attributed to the unique dependence structure of each copula as showed in Figure 5. As expected,

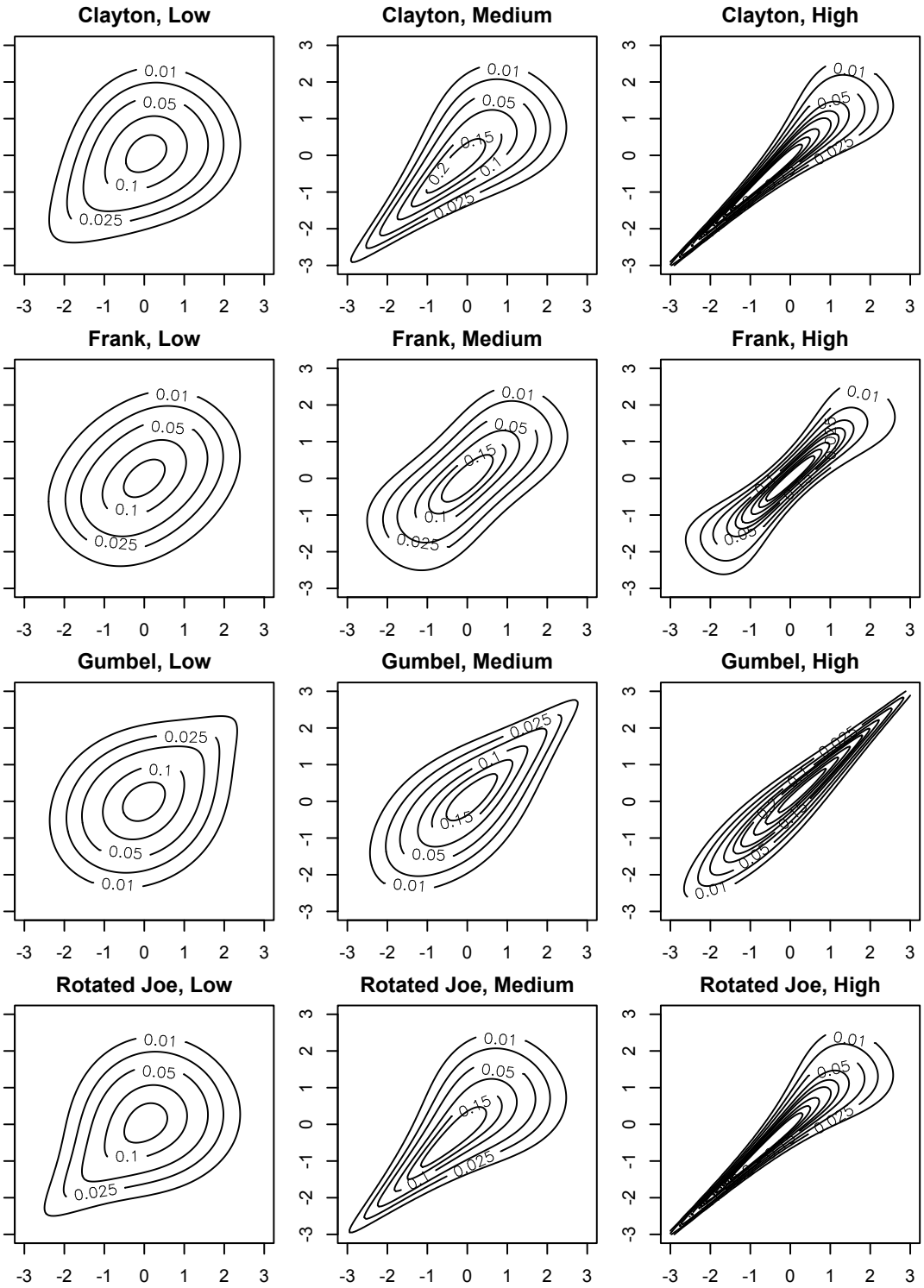

Figure 5: Elliptical contour density plots of candidate copulas at low, medium, and high associations.

Table 14: Coverage probability - low dependence

	Clayton	Frank	Gumbel	Survival Joe
Joint				
$\mathcal{T}_1$ ($\tau = 0.3$)	0.376	0.014	0.000	0.610
$\mathcal{T}_2$ ($\tau = 0.2$)	0.328	0.046	0.002	0.624
$\mathcal{T}_3$ ($\tau = 0.1$)	0.222	0.194	0.084	0.500
Sequential				
$\mathcal{T}_1$ ($\tau = 0.3$)	0.388	0.014	0.000	0.598
$\mathcal{T}_2$ ($\tau = 0.2$)	0.338	0.044	0.002	0.616
$\mathcal{T}_3$ ($\tau = 0.1$)	0.220	0.196	0.084	0.500

Table 15: Coverage probability - medium dependence

	Clayton	Frank	Gumbel	Survival Joe
Joint				
$\mathcal{T}_1$ ($\tau = 0.6$)	0.390	0.000	0.000	0.610
$\mathcal{T}_2$ ($\tau = 0.4$)	0.338	0.000	0.000	0.662
$\mathcal{T}_3$ ($\tau = 0.2$)	0.328	0.048	0.002	0.622
Sequential				
$\mathcal{T}_1$ ($\tau = 0.6$)	0.410	0.004	0.000	0.586
$\mathcal{T}_2$ ($\tau = 0.4$)	0.328	0.000	0.000	0.672
$\mathcal{T}_3$ ($\tau = 0.2$)	0.322	0.048	0.002	0.628

Table 16: Coverage probability - high dependence

	Clayton	Frank	Gumbel	Survival Joe
Joint				
$\mathcal{T}_1$ ($\tau = 0.9$)	0.538	0.060	0.000	0.402
$\mathcal{T}_2$ ($\tau = 0.6$)	0.602	0.000	0.000	0.398
$\mathcal{T}_3$ ($\tau = 0.3$)	0.502	0.024	0.000	0.474
Sequential				
$\mathcal{T}_1$ ($\tau = 0.9$)	0.548	0.040	0.000	0.412
$\mathcal{T}_2$ ($\tau = 0.6$)	0.610	0.000	0.000	0.390
$\mathcal{T}_3$ ($\tau = 0.3$)	0.498	0.028	0.000	0.474

identification at extremely low and high dependence becomes more difficult. Specifically, in the tree with $\tau = 0.1$, both methods misspecify the Frank copula and the Gumbel copula at 20% and 8% chance, respectively. The misspecification rate becomes ignorable when there is moderate dependence, but signals a slight rebound at high dependence $\tau = 0.9$. Third, neither approach successfully distinguishes the true copula from the Clayton copula. We do not find this result surprising considering the similarity in the contour plots of the Clayton and survival Joe copulas on one hand, and the non-identifiability issue due to the discreteness in the marginals on the other hand (see, for example, Genest and Nešlehová (2007)). Note that the contours of the two copulas are almost identical in the high dependence case, which explains the low coverage probability of the survival Joe in Table 16.

References

Aas, K., C. Czado, A. Frigessi, and H. Bakken (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics 44*(2), 182–198.

Bedford, T. and R. M. Cooke (2001). Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial intelligence 32*(1), 245–268.

Bedford, T. and R. M. Cooke (2002). Vines–a new graphical model for dependent random variables. *Annals of Statistics 30*(4), 1031–1068.

Brechmann, E. C. and C. Czado (2015). COPAR–multivariate time series modeling using the copula autoregressive model. *Applied Stochastic Models in Business and Industry 31*(4), 495–514.

Brechmann, E. C., C. Czado, and K. Aas (2012). Truncated regular vines in high dimensions with application to financial data. *Canadian Journal of Statistics 40*(1), 68–85.

Bühlmann, H. (1967). Experience rating and credibility. *ASTIN Bulletin: The Journal of the International Actuarial Association 4*(03), 199–207.

Diggle, P., P. Heagerty, K.-Y. Liang, and S. Zeger (2002). *Analysis of Longitudinal Data* (2nd ed.). Oxford University Press.

Dißmann, J., E. C. Brechmann, C. Czado, and D. Kurowicka (2013). Selecting and estimating regular vine copulae and application to financial returns. *Computational Statistics & Data Analysis 59*, 52–69.

Dunn, P. K. and G. K. Smyth (2005). Series evaluation of Tweedie exponential dispersion model densities. *Statistics and Computing 15*(4), 267–280.

Dunn, P. K. and G. K. Smyth (2008). Series evaluation of Tweedie exponential dispersion model densities by Fourier inversion. *Statistics and Computing 18*(1), 73–86.

Frees, E. (2014). Frequency and severity models. In E. Frees, G. Meyers, and R. Derrig (Eds.), *Predictive Modeling Applications in Actuarial Sciences: Volume I, Predictive Modeling Techniques*, pp. 138–166. Cambridge University Press: Cambridge.

Frees, E. and E. Valdez (2008). Hierarchical insurance claims modeling. *Journal of the American Statistical Association 103*(484), 1457–1469.

Frees, E. and P. Wang (2006). Copula credibility for aggregate loss models. *Insurance: Mathematics and Economics 38*(2), 360–373.

Frees, E. W., G. Meyers, and A. D. Cummings (2012). Summarizing insurance scores using a gini index. *Journal of the American Statistical Association 106*(495).

Frees, E. W. and P. Wang (2005). Credibility using copulas. *North American Actuarial Journal 9*(2), 31–48.

Frees, E. W., V. R. Young, and Y. Luo (1999). A longitudinal data analysis interpretation of credibility models. *Insurance: Mathematics and Economics 24*(3), 229–247.

Frees, E. W. J., G. Meyers, and A. D. Cummings (2014). Insurance ratemaking and a gini index. *Journal of Risk and Insurance 81*(2), 335–366.

Genest, C. and J. Nešlehová (2007). A primer on copulas for count data. *ASTIN Bulletin: The Journal of the International Actuarial Association 37*(02), 475–515.

Gruber, L., C. Czado, et al. (2015). Sequential bayesian model selection of regular vine copulas. *Bayesian Analysis 10*(4), 937–963.

Haff, I. H., K. Aas, and A. Frigessi (2010). On the simplified pair-copula constructionsimply useful or too simplistic? *Journal of Multivariate Analysis 101*(5), 1296–1310.

Hintze, J. L. and R. D. Nelson (1998). Violin plots: a box plot-density trace synergism. *The American Statistician 52*(2), 181–184.

Joe, H. (2005). Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis 94*(2), 401–419.

Joe, H. (2014). *Dependence Modeling with Copulas*. New York: Chapman & Hall.

Jørgensen, B. and M. de Souza (1994). Fitting tweedies compound poisson model to insurance claims data. *Scandinavian Actuarial Journal 1*(1), 69–93.

Kurowicka, D. and R. M. Cooke (2006). *Uncertainty Analysis with High Dimensional Dependence Modelling*. John Wiley & Sons.

Liang, K.-Y. and S. L. Zeger (1986). Longitudinal data analysis using generalized linear models. *Biometrika 73*(1), 13–22.

Masarotto, G., C. Varin, et al. (2012). Gaussian copula marginal regression. *Electronic Journal of Statistics 6*, 1517–1549.

McDonald, J. (1984). Some generalized functions for the size distribution of income. *Econometrica 52*(3), 647–63.

McDonald, J. and Y. Xu (1995). A generalization of the beta distribution with applications. *Journal of Econometrics 66*(1-2), 133–152.

Mowbray, A. H. (1914). How extensive a payroll exposure is necessary to give a dependable pure premium. In *Proceedings of the Casualty Actuarial Society*, Volume 1, pp. 24–30.

Olsen, M. K. and J. L. Schafer (2001). A two-part random-effects model for semicontinuous longitudinal data. *Journal of the American Statistical Association 96*(454), 730–745.

Panagiotelis, A., C. Czado, and H. Joe (2012). Pair copula constructions for multivariate discrete data. *Journal of the American Statistical Association 107*(499), 1063–1072.

Panagiotelis, A., C. Czado, H. Joe, and J. Stöber (2015). Model selection for discrete regular vine copulas. *Working Paper*.

Ruppert, D., M. P. Wand, and R. J. Carroll (2003). *Semiparametric Regression*. Cambridge University Press.

Shi, P. (2014). Fat-tailed regression models. In E. Frees, G. Meyers, and R. Derrig (Eds.), *Predictive Modeling Applications in Actuarial Sciences: Volume I, Predictive Modeling Techniques*, pp. 138–166. Cambridge University Press: Cambridge.

Shi, P., X. Feng, and J.-P. Boucher (2016). Multilevel modeling of insurance claims using copulas. *Annals of Applied Statistics 10*(2), 834–863.

Shi, P. and W. Zhang (2015). Private information in healthcare utilization: specification of a copula-based hurdle model. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 178*(2), 337–361.

Smith, M., A. Min, C. Almeida, and C. Czado (2010). Modeling longitudinal data using a pair-copula decomposition of serial dependence. *Journal of the American Statistical Association 105*(492), 1467–1479.

Smith, M. S. (2015). Copula modelling of dependence in multivariate time series. *International Journal of Forecasting 31*(3), 815–833.

Smyth, G. and B. Jørgensen (2002). Fitting Tweedie's compound poisson model to insurance claims data: dispersion modelling. *ASTIN Bulletin: The Journal of the International Actuarial Association 32*(1), 143–157.

Stöber, J. (2013). *Regular vine copulas with the simplifying assumption, time-variation, and mixed discrete and continuous margins*. Ph. D. thesis, Technische Universität München.

Stöber, J., H. G. Hong, C. Czado, and P. Ghosh (2015). Comorbidity of chronic diseases in the elderly: Patterns identified by a copula design for mixed responses. *Computational Statistics & Data Analysis 88*, 28–39.

Stoeber, J., H. Joe, and C. Czado (2013). Simplified pair copula constructionslimitations and extensions. *Journal of Multivariate Analysis 119*, 101–118.

Whitney, A. W. (1918). Theory of experience rating. In *Proceedings of the Casualty Actuarial Society*, Volume 4, pp. 274–292.

Zhang, Y. (2013). Likelihood-based and Bayesian methods for Tweedie compound Poisson linear mixed models. *Statistics and Computing 23*(6), 743–757.