Expert Systems With Applications 185 (2021) 115552

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Dependence modeling of multivariate longitudinal hybrid insurance data with dropout

Edward W. Frees [a,b], Catalina Bolancé [c], Montserrat Guillen [c,*], Emiliano A. Valdez [d]

[a] University of Wisconsin-Madison, United States of America
[b] Australian National University, Australia
[c] Universitat de Barcelona, Spain
[d] University of Connecticut, United States of America

ARTICLE INFO

Keywords:
Copula
Generalized method of moments
Joint longitudinal and time-to-event
Insurance lapsation
Tweedie distribution

ABSTRACT

Financial services industries, such as insurance, increasingly use data from their broad cross-section of customers and follow these customers over time. In other areas such as medicine, engineering, and communication systems, it is well known that following subjects over time may result in biased data, for example, the so-called "dropout effect". This paper introduces techniques to address dropout commonly encountered in the insurance domain. Specifically, in the insurance context, multivariate claims outcomes may be related to a customer's dropout or decision to lapse a policy. Insurance claims outcomes are also naturally a hybrid with both discrete and continuous components, which adds complexity to model calibration. Decision makers in the insurance industry will find our work provides helpful guidance in integrating customer loyalty, especially with bundled coverages. This paper introduces a generalized method of moments technique to estimate dependence parameters where associations are represented using copulas. This is especially useful for large data sets. The paper describes how the joint model provides new information that insurers can use to better manage their portfolios of risks. An application to a Spanish insurer data set is presented.

1. Introduction

Modeling of repeated observations of a subject over time, the topic of longitudinal or panel data, has long been an important tool in the biomedical and social sciences. Naturally, the field initially focused attention on the analysis of data in which a single outcome is analyzed. Subsequently, the desire to handle multivariate outcomes has gained prominence, as emphasized in a review by Verbeke et al. (2014). In this paper we study multivariate insurance losses. Specifically, we consider a sample of individuals, followed over time, who have purchased more than one type of insurance policy with claims arising from each contract. To illustrate, we consider people who purchased policies that cover insured losses on an automobile and a home.

It is customary for insurance companies to analyze returns in insurance separating risk premium calculation from customer retention. One of the reasons is that customers have different types of policies, such as motor, home and so on, each one with their own pricing formula. So, companies routinely work with pricing models and retention models separately. However, insurance companies have moved from an approach centered on the retention of each policy contract separately to an approach that considers the customer as a whole. Our contribution shows that there is a significant association between ratemaking and loyalty, and how this association can be addressed. This is important for maximizing profitability by deciding which customers are the most profitable to retain and how to fix prices accordingly. There are few works considering insurance policy renewals and claim costs simultaneously, because these two concepts are considered independent of one another (see Bolancé et al., 2018). For example, Kaishev et al. (2013) looked only at optimization in customer selection for cross-selling.

As emphasized by Verbeke et al. (2014), a multivariate modeling strategy is more complex than analyzing univariate outcomes but also provides important insights not available with univariate analysis. For example, with a multivariate model, a researcher may wish to assess the relation between some covariate and all outcomes simultaneously. Further, multivariate modeling is needed to understand associations among outcomes and how these relationships evolve over time.

A well known limitation of longitudinal data models, with either univariate or multivariate outcomes, is the bias caused by event times that dictate outcome availability where the event is related to the outcome being studied. As an example from the biomedical field, outcomes can be a function of patient medical expenditures which

* Corresponding author.
E-mail address: mguillen@ub.edu (M. Guillen).

https://doi.org/10.1016/j.eswa.2021.115552
Received 14 May 2020; Received in revised form 4 June 2021; Accepted 1 July 2021
Available online 16 July 2021

are associated with survival. Intuitively, it seems plausible that sicker patients are more likely to die (and thus exit a study) and to incur higher medical costs, an outcome of interest, c.f., Liu (2009). To address this, joint models of longitudinal and time-to-event data have been developed in the biomedical literature; for example, Ibrahim et al. (2010) and Tsiatis and Davidian (2004) review the foundations. A more recent overview by Papageorgiou et al. (2019) describes the explosion of the field including the development of several statistical packages within R (R Core Team 2018), SAS/STAT, Stata, WinBUGS, and JAGS that enable innovative applications of the methodology.

Elashoff et al. (2017) describe two basic approaches for modeling the association between longitudinal data and time-to-events, both involving conditioning. The first is the selection model, where the joint distribution between outcomes of interest and an event is factored into a distribution of outcomes and the distribution of events given the outcomes. The second basic approach is the pattern mixture model, where the factorization is reversed. That is, the joint distribution is factored into the distribution of events and the distribution of outcomes given the events.

Other sources, such as Njagi et al. (2014), think of the "shared parameter model" as a separate category (Elashoff et al., 2017 include this approach as a selection model). In the shared parameter model, one or more latent variables may induce an association between survival/drop-out and outcomes (e.g., costs). A strength of this approach is that latent variables are intuitively appealing, one can ascribe attributes to them (e.g., "frailty") to help interpret dependence.

This paper extends the joint multivariate longitudinal and time-to-event data modeling literature in two ways. First, it considers insurance, a new field of application. This new application suggests concerns that are not commonly seen in biomedical applications, including marginal outcome distributions that are hybrid combinations of continuous and discrete components and larger sample sizes (suggesting the need for alternative computational approaches). Considerations of this new application area lead naturally to the introduction of copula models as a tool to understand associations. As noted in the review by Verbeke et al. (2014), copula models have been utilized in the multivariate longitudinal data literature although it is not widely used in biomedical applications; early contributions include Frees and Wang (2005), Lambert (1996), Lambert and Vandenhende (2002), and Meester and Mackay (1994). Second, the copula approach for modeling dependencies has not been explored for joint models of longitudinal and time-to-event data.

As we argue in this paper, there are several strengths of copula models compared to alternatives. First, copulas allow for a unified model of associations while preserving marginal distributions. In complex applications like insurance where marginal distributions are hybrid combinations of continuous and discrete components, this allows analysts to use traditional diagnostic techniques to specify marginal distributions. Second, there is no need to make assumptions about conditioning when specifying joint distributions. In the insurance context, we will think about outcomes of interest as insurance claims and events as lapsation, or dropout, from insurance policies. In advance, we do not know when an insurance claim causes a policyholder to lapse a policy, or whether the decision to lapse influences the outcome of a claims.[1] With a copula approach, we model the observed dependency directly without the need for making decisions about conditioning. Third, we argue that the computational complexity of the copula modeling, with the new estimation procedures introduced in this paper, allow us to handle some large problems that would be difficult, if not infeasible, using alternative available strategies.

[1] Customers may be dissatisfied after a claim and then decide to look for another company and lapse. But, some customers may first decide to lapse and then claim more than they would have if they had renewed the policy, knowing that they would elude penalization. Most companies reward customers with a price rebate when they renew their policy, if they have not reported claims in the year previous to renewal.

Table 1
Expert system in insurance.

Application	References
Risk analysis	Meyer et al. (1992) Hsiao and Whang (2009) Abbasi and Guillen (2013) Hiabu et al. (2016)
Accident analysis	Bayam et al. (2005)
Marketing strategies	Abrahams et al. (2009)
Construction insurance pricing	Imriyas (2009) Cheng et al. (2011)
Cross-selling pricing	Thuring et al. (2012) Kaishev et al. (2013)
Detecting insurance fraud	Pathak et al. (2005) Subelj et al. (2011)

1.1. Literature review: Expert systems applications

The concept of dropout, the momentary or eventual loss of information, can be particularly important to users of expert systems. In medicine, the reasons for patient dropouts in clinical trials are frequently examined because clinical measurement outcomes become missing observations and cannot be ignored (Harrer et al., 2019). In engineering, the input/output data in control systems can be affected by dropouts caused by network failures (see Hou & Bu, 2018). In communication systems, correcting damages caused by digital dropouts are of paramount importance especially when preserving archived media (see Ahn et al., 2015).

Expert systems have provided alternative decision-making strategies for insurance companies, the goal of which is to improve profits. Risk analysis is a fundamental area, in this context, Hsiao and Whang (2009) used a neural network for predicting life insurers insolvency and it is compared with classical discriminant analysis and logistic regression (see also Meyer et al., 1992 for alternative risk management application in life insurance business). Abbasi and Guillen (2013) proposed an expert system-based chart for monitoring value-at-risk in non-life insurance claims (see also Hiabu et al., 2016 for an application in monitoring cash-flow). Focusing on accident analysis, Bayam et al. (2005) used data-mining methods, decision trees and neural networks, for predicting traffic accidents of senior drivers. Using decision trees, Abrahams et al. (2009) analyzed the best marketing strategies for a specific insurance company. Expert systems have also been used for insurance pricing in the construction sector; Imriyas (2009) proposed a strategy to calculate premium for workers' compensation insurance in construction projects and Cheng et al. (2011) focused on calculating the optimal deductible using an evolutionary support vector machine. Optimal cross-selling pricing strategies in auto and home insurance were proposed by Kaishev et al. (2013) and Thuring et al. (2012). Pathak et al. (2005) and Subelj et al. (2011) analyzed fraud detection systems.

In Table 1 we summarize and classify the main contributions in expert system for insurance. Furthermore, in the context of artificial intelligence, Eling et al. (2002) summarize the insurance big data applications of artificial intelligence and evaluate its impact in this industry.

1.2. Dependence and foundations of insurance

For notation, consider the joint distribution of p outcomes, $Y_1, \ldots, Y_p$. Copulas provide a general tool for modeling dependencies and so express the joint distribution as

$$F(y_1, \ldots, y_p) = \Pr(Y_1 \le y_1, \ldots, Y_p \le y_p) = C\left(F_1(y_1), \ldots, F_p(y_p)\right). \quad (1)$$

Here, $F_1(y_1) = \Pr(Y_1 \le y_1)$ is known as the marginal distribution of Y_1 and similarly for the other variables. We assume that there is an associated set of explanatory variables, $\mathbf{x}$, that is available to calibrate the marginal distributions. When the distribution of the marginal distributions depends on covariates, we refer to this as a *copula regression model*. If we think of $Y_1, \ldots, Y_m$ as realizations from the same subject over time (we use m for the number of time points), then this framework reduces to longitudinal data. In the same sense, if a realization at a time point consists of multiple outcomes, then this is multivariate longitudinal data.

Copula regression modeling is ideally suited for applications where there are many variables available to explain outcomes (the regression portion) and where structural dependence among outcomes is critical (the copula portion). Compared to other multivariate techniques, copulas are particularly suitable in insurance applications because there is a lack of theory to support specification of a dependence structure and data-driven methods, such as copula modeling, fare well.

Copula regression modeling is introduced in greater detail in Section 2.1. Traditionally, a barrier to implementing these models in high-dimensional cases (large p) has been the presence of discreteness which substantially increases the computational burden. In this paper, we show how to use the generalized method of moments (GMM), a type of estimating equations approach, to estimate association parameters of this model. Unlike traditional treatments, we emphasize examples where outcome variables may be a hybrid combination of continuous and discrete components. As will be described, this approach provides an alternative to vine copula models that have been developed recently.

Starting with $Y_1, \ldots, Y_m$ vectors, we now allow Y_t to represent a multivariate outcome and an indicator for dropout in Section 2.2. This section is an extension, and not a special case, of Section 2.1 because of potential dependencies between the outcomes and the observation process. To provide intuition, Section 2.3 focuses on the special case of temporal independence (but still accounting for dependency between dropout and outcomes during the same time period).

Section 3 describes an empirical application where we examine outcomes and dropout (lapsation), tracking 40,284 clients over five years. Section 4 concludes.

2. Methods

2.1. Copula regression modeling and GMM estimation

In copula regression applications, it is common to use the *inference for margins* (IFM) procedure, a two stage estimation algorithm due to Joe (2005). The first stage maximizes the likelihood from marginals, and the second stage maximizes the likelihood of dependence parameters with parameters in the marginals held fixed from the first stage. For insurance applications of interest to us where the number of subjects n ranges over the tens of thousands and the time dimension T is limited (typically about 5), this general model fitting strategy works well and will be utilized in the subsequent development.

2.1.1. Hybrid distributions

Assume that the random variables may have both discrete and continuous components. The need for handling both discrete and continuous components was emphasized by Song et al. (2009) in the copula regression literature. They referred to this combination as a "mixture" of discrete and continuous components. In insurance and many other fields, the term "mixture" is used for distributions with different sub-populations that are combined using latent variables. So, we prefer to refer to this as a "hybrid" combination of discrete and continuous components to avoid confusion with mixture distributions. For a random variable Y, let y^d represent a mass point (d for discrete) and let y^c represent a point of continuity where the density is positive.

Likelihood

We now give a general expression for the likelihood. To do so, assume that the first q arguments $y_1, \ldots, y_q$ represent mass points. Further assume that the other arguments $y_{q+1}, \ldots, y_p$ represent points of continuity. Then, the likelihood corresponding to the distribution function in Eq. (1) can be expressed as

$$
\begin{aligned}
& f(y_1, \ldots, y_q, y_{q+1}, \ldots, y_p) = \\
& \quad \sum_{i_1=0}^{1} \cdots \sum_{i_q=0}^{1} \\
& \qquad \times (-1)^{i_1+\cdots+i_q} C_{q+1,\ldots,p}\left(F_1(y_1^{(i_1)}), \ldots, F_q(y_q^{(i_q)}), F_{q+1}(y_{q+1}), \ldots, F_p(y_p)\right) \\
& \qquad \times \prod_{j=q+1}^{p} f_j(y_j), \qquad (2)
\end{aligned}
$$

where $y^{(i)} = y$ if $i = 0$ and $y^{(i)} = y-$ if $i = 1$. The notation $y-$ means evaluate y as a left-hand limit. See, for example, Song et al. (2009), equation (9). Here, $C_{q+1,\ldots,p}$ is the partial derivative of the copula with respect to the arguments in the $q+1, \ldots, p$ positions. In common elliptical families, evaluating partial derivatives of a copula is a tedious, yet straightforward, task. As we will see, a large number of variables with discrete components (q in Eq. (2)) represents a source of difficulty when numerically evaluating likelihoods.

Pairwise Distributions

Particularly for discrete or hybrid outcomes, direct estimation using maximum likelihood (or IFM) can be difficult because the copula distribution function may require a high-dimensional evaluation of an integral for each observation. A generally available alternative is to examine the information from subsets of random variables. For example, focusing on pairs, consider the corresponding bivariate distribution function

$$
\begin{aligned}
F_{jk}(y_j, y_k) &= C\left(\infty, \ldots, \infty, F_j(y_j), \ldots, F_k(y_k), \infty, \ldots, \infty\right) \\
&= C^{(jk)}\left(F_j(y_j), F_k(y_k)\right).
\end{aligned}
$$

Although this expression for the joint distribution function F_{jk} is broadly applicable, it is particular useful for copulas in the elliptical family. If we specify C to be an elliptical copula with association matrix $\boldsymbol{\Sigma}$, then $C^{(jk)}$ is from the same elliptical family with association parameter $\boldsymbol{\Sigma}_{jk}$, corresponding to the jth row and kth column of $\boldsymbol{\Sigma}$. This is the specification used in this paper. For notational purposes, we henceforth drop the superscripts on the bivariate copula and hope that the context makes the definition clear.

Suppose that we wish to evaluate the likelihood of Y_j at mass point y_j^d and of Y_k at point of continuity y_k^c. Then, the joint distribution function has a hybrid probability density/mass function of the form:

$$
\begin{aligned}
f_{jk}(y_j^d, y_k^c) &= \partial_2 \Pr(Y_j = y_j^d, Y_k \le y_k^c) \\
&= \partial_2 \left\{ \Pr(Y_j \le y_j^d, Y_k \le y_k^c) - \Pr(Y_j \le y_j^d-, Y_k \le y_k^c) \right\} \\
&= \partial_2 \left\{ C\left(F_j(y_j^d), F_k(y_k^c)\right) - C\left(F_j(y_j^d-), F_k(y_k^c)\right) \right\} \\
&= \left\{ C_2\left(F_j(y_j^d), F_k(y_k^c)\right) - C_2\left(F_j(y_j^d-), F_k(y_k^c)\right) \right\} f_k(y_k^c).
\end{aligned}
$$

Here, C_2 represents the partial derivative of the copula C with respect to the second argument. For the likelihood, we need to consider two other cases, where y_j and y_k are both points of continuity or both discrete points. These cases are more familiar to most readers; details are available from the authors.

Example: Two Tweedie Variables

In insurance, it is common to refer to a random variable with a mass at zero and a continuous density over the positive reals as a "Tweedie" random variable (corresponding to a Tweedie distribution). Suppose that both Y_j and Y_k are Tweedie random variables. The joint distribution function has a hybrid probability density/mass function of the form:

$$
f_{jk}(y_j, y_k) = \begin{cases}
\Pr(Y_j = 0, Y_k = 0) = F_{jk}(0, 0) & y_j = 0, y_k = 0 \\
C_1\left(F_j(y_j), F_k(0)\right) f_j(y_j) & y_j > 0, y_k = 0 \\
C_2\left(F_j(0), F_k(y_k)\right) f_k(y_k) & y_j = 0, y_k > 0 \\
c\left(F_j(y_j), F_k(y_k)\right) f_j(y_j) f_k(y_k) & y_j > 0, y_k > 0.
\end{cases}
$$

Here, C_1 represents the partial derivative of the copula C with respect to the first argument and c is the corresponding density. To illustrate, when both observations are 0, then the likelihood $F_{jk}(0,0) = C\left(F_j(0), F_k(0)\right)$ requires evaluation of the distribution function C. With a Gaussian copula, this is a two-dimensional integral. In the same way, with m years of data, an observation that has no claims (all 0's) in all years requires a m-dimensional integration. For data sets that have tens of thousands of observations, this becomes cumbersome even when m is small, e.g., $m = 5$ as in our application. This is one motivation for utilizing pairwise distributions in this paper.

2.1.2. Generalized method of moments procedure

As described in Joe (2014), Section 5.5, much of theory needed to justify composite likelihood methods in a copula setting can be provided through estimating equations. We utilize an estimating equation technique that is common in econometrics, *generalized method of moments*, denoted by the acronym *GMM*. Compared to classic estimating equation approaches, this technique has the advantage that a large number of equations may be used and that the resulting estimators enjoy certain optimality properties. The *GMM* procedure is broadly used although applications in the copula context have been limited. For example, Prokhorov and Schmidt (2009) consider *GMM* and copulas in a longitudinal setting although not with discrete outcomes.

GMM Procedure We define the score function

$$g_{\theta,ijk}(Y_{ij}, Y_{ik}) = \partial_\theta \ln f_{ijk}(Y_{ij}, Y_{ik}). \tag{3}$$

This is a mean zero random vector that contains information about θ. It is an unbiased estimator (of zero) and can be used in an estimating equation. Using Eq. (3), we combine several scores from the ith risk as

$$g_{\theta,i}(Y_{i1}, \dots, Y_{ip}) = \begin{cases} g_{\theta,i12}(Y_{i1}, Y_{i2}) \\ \vdots \\ g_{\theta,i1p}(Y_{i1}, Y_{ip}) \\ \vdots \\ g_{\theta,i,p-1,p}(Y_{i,p-1}, Y_{ip}) \end{cases} \tag{4}$$

a column vector with $r\binom{p}{2}$ rows. The sum of these statistics for a sample of size n is $g_\theta = \sum_{i=1}^n g_{\theta,i}(Y_{i1}, \dots, Y_{ip})$. Although g_θ is a mean zero vector containing information about θ, the number of elements in g_θ exceeds the number of parameters and so we use *GMM* to estimate the parameters. Specifically, the *GMM* estimator of θ, say θ_{GMM}, is the minimizer of the expression $g_\theta\left(\text{Var } g_{\hat{\theta}}\right)^{-1} g'_\theta$. To implement this, we use the plug-in estimator of the variance

$$\widehat{\text{Var}}\ g_{\hat{\theta}} = \frac{1}{n}\sum_{i=1}^n g_{\hat{\theta},i}(Y_{i1}, \dots, Y_{ip})\ g_{\hat{\theta},i}(Y_{i1}, \dots, Y_{ip})'. \tag{5}$$

This plug-in estimator requires $\hat{\theta}$, a consistent estimator of θ. We use the pairwise likelihood estimator θ_{PL}.

For asymptotic variances, we use the gradient $G_\theta = \text{E}\ \frac{\partial}{\partial\theta'} g_\theta$, a matrix of dimension $\left(r\binom{p}{2}\right) \times r$. Then, following the usual asymptotic theory (see, for example, Wooldridge, 2010, Chapter 14), we have that θ_{GMM} is asymptotically normal with mean θ and variance $n^{-1}\left(G'_{\hat{\theta}}\left(\text{Var } g_{\hat{\theta}}\right)^{-1} G_{\hat{\theta}}\right)^{-1}$. Further, $n^{-1} g'_{\hat{\theta}}\left(\text{Var } g_{\hat{\theta}}\right)^{-1} g_{\hat{\theta}}$ has a limiting chi-square distribution with $r\binom{p}{2} - r$ degrees of freedom.

Evaluation of GMM Scores

We now evaluate the scores. For discrete y_j^d and continuous y_k^c outcomes, we have

$$\begin{aligned} g_{\theta,ijk}(y_j^d, y_k^c) &= \partial_\theta \ln\left[\left\{C_2\left(F_j(y_j^d), F_k(y_k^c)\right) - C_2\left(F_j(y_j^d-), F_k(y_k^c)\right)\right\} f_k(y_k^c)\right] \\ &= \frac{\partial_\theta\left\{C_2\left(F_j(y_j^d), F_k(y_k^c)\right) - C_2\left(F_j(y_j^d-), F_k(y_k^c)\right)\right\}}{C_2\left(F_j(y_j^d), F_k(y_k^c)\right) - C_2\left(F_j(y_j^d-), F_k(y_k^c)\right)} \end{aligned}$$

For two discrete outcomes, the score can be expressed as

$$\begin{aligned} & g_{\theta,ijk}(y_j^d, y_k^d) \\ &= \partial_\theta \ln f_{ijk}(y_j^d, y_k^d) \\ &= \frac{\partial_\theta\left\{C\left(F_j(y_j^d),F_k(y_k^d)\right) - C\left(F_j(y_j^d-),F_k(y_k^d)\right) - C\left(F_j(y_j^d),F_k(y_k^d-)\right) + C\left(F_j(y_j^d-),F_k(y_k^d-)\right)\right\}}{C\left(F_j(y_j^d),F_k(y_k^d)\right) - C\left(F_j(y_j^d-),F_k(y_k^d)\right) - C\left(F_j(y_j^d),F_k(y_k^d-)\right) + C\left(F_j(y_j^d-),F_k(y_k^d-)\right)}. \end{aligned}$$

For two continuous outcomes, y_j^c and y_k^c, we have

$$\begin{aligned} g_{\theta,ijk}(y_j^c, y_k^c) &= \partial_\theta \ln\left[c(F_{Y_{ij}}(y_j^c), F_{Y_{ik}}(y_k^c)) f_{ij}(y_j^c) f_{ik}(y_k^c)\right] \\ &= \frac{\partial_\theta\ c(F_{Y_{ij}}(y_j^c), F_{Y_{ik}}(y_k^c))}{c(F_{Y_{ij}}(y_j^c), F_{Y_{ik}}(y_k^c))}. \end{aligned}$$

An advantage of restricting ourselves to pairwise distributions is that most of the functions are available from Schepsmeier and Stöber (2012, 2014). We used an additional relationship, from Plackett (1954),

$$\frac{\partial}{\partial\rho} C(u_1, u_2) = \phi_2(z_1, z_2).$$

Here, ϕ_2 is a bivariate normal probability density function, ρ is the correlation parameter and $z_j = \Phi^{-1}(u_j)$, $j = 1, 2$ are the normal scores corresponding to residuals u_j. Further details are available from the authors.

2.1.3. Comparing pairwise likelihood to GMM estimators

This section compares the efficiency of the pairwise likelihood estimator θ_{PL} to the *GMM* estimator θ_{GMM} through large sample approximations and (small sample) simulations.

Asymptotic Comparison. For large sample central limit theorem approximations, we have already stated that the asymptotic variance of θ_{GMM} is $\frac{1}{n}\left(G'_\theta\left(\text{Var } g_\theta\right)^{-1} G_\theta\right)^{-1}$.

For a comparable statement for pairwise likelihoods, first note that the estimating equation can be expressed as

$$g_{PL}(\mathbf{Y}_i; \theta) = \sum_{j=1}^{p-1}\sum_{k=j+1}^{p} g_{\theta,ijk}(Y_{ij}, Y_{ik}) = \left(\mathbf{I}_r \otimes \mathbf{1}'\right) g_{\theta,i}(\mathbf{Y}_i) = \mathbf{B}\ g_{\theta,i}(\mathbf{Y}_i),$$

where $\otimes$ is a Kronecker product, $\mathbf{I}_r$ is an $r \times r$ identity matrix, and $\mathbf{1}'$ is a $1 \times \binom{p}{2}$ vector of ones. From this, the sensitivity matrix is $\mathbf{B}\ G_\theta$, where $G_\theta = \text{E}\ \frac{\partial}{\partial\theta'} g_\theta$. The variance matrix is $\mathbf{B}\ \left(\text{Var } g_\theta\right)^{-1} \mathbf{B}'$. Following the usual estimating equation methodology (c.f. Song, 2007), the asymptotic variance of θ_{PL} is

$$\frac{1}{n}\left(G'_\theta \mathbf{B}'\left(\mathbf{B}\ \text{Var } g_\theta\ \mathbf{B}'\right)^{-1} \mathbf{B}\ G_\theta\right)^{-1}.$$

Not surprisingly, this is larger (in a matrix sense) than the *GMM* estimator.

Small Sample Comparison. We also compared the efficiency of the pairwise likelihood and the *GMM* estimator in a simulation study, the results are available from the authors. Not surprisingly, here we show that the *GMM* estimator has a lower standard error than the pairwise estimator and in this sense is more efficient.

Interestingly, we learned that the pairwise estimator was more robust than the basic GMM estimator when the level of discreteness (proportion of zeros) was extensive and the association among outcomes was large in small sample sizes. Both estimators became more efficient as the sample size increase (in our study, ranging from n=100 to n=2000) and the basic GMM estimator was clearly the better choice for larger sample sizes. Moreover, the flexibility of the GMM estimator allowed us to introduce estimated weights (based on the variance–covariance matrix among scores) that allowed us to propose variations that outperformed the pairwise version even in small sample sizes.

2.2. Multivariate longitudinal modeling with dropout

We now extend the *GMM* estimation of copula regression models to more complex situations where the dependent outcomes include the observation process. This extension is motivated by the lapsation of insurance contracts where insurance customer lapsation is equivalent to biomedical patient dropout.

2.2.1. Joint model specification

Consider the case where we follow individuals (insurance policyholders) over time. For illustration, suppose that there are three outcome variables of interest. Two random variables, $Y_{1,it}$ and $Y_{2,it}$, might represent claims from auto and homeowners coverages, respectively. A third random variable, L_{it}, is a binary variable that represents a policyholder's decision to lapse the policy. Specifically, $L_{it} = 1$ represents the policyholder's decision to not renew the policy (to lapse) and so we do not observe the policy at time $t+1$. If $L_{it} = 0$, then we observe the policy at time $t+1$, subject to limitations on the number of time periods available. Let m represent the maximal number of observations over time.

For each policyholder i at time t, we potentially observe the lapse variable L_{it} and a collection of other outcome variables $\mathbf{Y}_{it} = \{Y_{1,it},\ldots,Y_{p,it}\}$. Associated with each policyholder is a set of (possibly time varying) rating variables $\mathbf{x}_{it}$ for the ith policyholder at time t. Throughout, we assume independence among individuals.

To define the joint distribution among outcomes, we use copulas that are based on transformed outcomes. Specifically, for each ijt, define $v_{ijt} = \Phi^{-1}\left(F_{Y_{ijt}}(Y_{ijt})\right)$ to be the transformed outcome. In the same way, define $v_{iLt} = \Phi^{-1}\left(F_{L_{it}}(L_{it})\right)$. When the outcomes are continuous, the transformed outcomes have a normal distribution. For outcomes that are not continuous, one can utilize the generalized distributional transform of Rüschendorf (2009). With this generalization of the probability integral transform, it is straight-forward to show that a copula exists even when outcomes are not continuous.

Temporal Structure

Although copulas allow for a broad variety of dependence structures, for our longitudinal/panel data it is useful to employ an association structure motivated by traditional time series models as described in the following. In Section 2.3, we focus on the special case of no temporal dependence.

There are many ways of specifying simple parametric structures to capture the dependencies among these transformed outcomes. To capture dependencies, we use a standard multivariate time series framework due to Box and Jenkins (cf., Tiao & Box, 1981). In this framework, there are sequences of latent variables $\{\eta_{ijt}\}$ that are iid. Transformed outcomes can be expressed as $v_{ijt} = \eta_{ijt} + \psi_{j,1}\eta_{ij,t-1} + \cdots + \psi_{j,t-1}\eta_{ij,1}$. For example, we might use a moving average of order one ($MA1$), $v_{ijt} = \eta_{ijt} + \psi_{j,1}\eta_{ij,t-1}$, or an autoregressive of order one ($AR1$) representation, $v_{ijt} = \eta_{ijt} + \psi_{j,1}v_{ij,t-1} = \eta_{ijt} + \psi_{j,1}\eta_{ij,t-1} + \cdots + \psi_{j,1}^{t-1}\eta_{ij,1}$. Using matrix notation, we define $\boldsymbol{v}_{ij} = \left(v_{ij1},\ldots,v_{ijm}\right)'$, $\boldsymbol{\eta}_{ij} = \left(\eta_{ij1},\ldots,\eta_{ijm}\right)'$, and

$$\boldsymbol{\Psi}_j = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \psi_{j,1} & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ \psi_{j,m-1} & \psi_{j,m-2} & \cdots & 1 \end{pmatrix}.$$

Thus, $\boldsymbol{v}_{ij} = \boldsymbol{\Psi}_j \boldsymbol{\eta}_{ij}$ and $\mathrm{Var}\ \boldsymbol{v}_{ij} = \boldsymbol{\Psi}_j \boldsymbol{\Psi}_j'$, for $j = 1,\ldots,p$. In the same way, for lapse we have $\boldsymbol{v}_{iL} = \boldsymbol{\Psi}_L \boldsymbol{\eta}_{iL}$ and $\mathrm{Var}\ \boldsymbol{v}_{iL} = \boldsymbol{\Psi}_L \boldsymbol{\Psi}_L'$.

One can think of the η_{ijt} as a transformed outcome with the temporal dependence filtered out. We further allow contemporaneous correlations among outcomes of the form

$$\mathrm{Cov}(\eta_{ijs},\eta_{ikt}) = \begin{cases} \rho_{jk} & s=t \\ 0 & s \neq t \end{cases} \quad \text{and} \quad \mathrm{Cov}(\eta_{ijs},\eta_{iLt}) = \begin{cases} \rho_{jL} & s=t \\ 0 & s \neq t \end{cases}.$$

This yields $\mathrm{Cov}(\boldsymbol{v}_{ij},\boldsymbol{v}_{ik}) = \rho_{jk}\boldsymbol{\Psi}_j\boldsymbol{\Psi}_k'$ and $\mathrm{Cov}(\boldsymbol{v}_{iL},\boldsymbol{v}_{ij}) = \rho_{Lj}\boldsymbol{\Psi}_L\boldsymbol{\Psi}_j'$. We summarize this using the association matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Psi}_L\boldsymbol{\Psi}_L' & \rho_{L1}\boldsymbol{\Psi}_L\boldsymbol{\Psi}_1' & \cdots & \rho_{Lp}\boldsymbol{\Psi}_L\boldsymbol{\Psi}_p' \\ \rho_{L1}\boldsymbol{\Psi}_1\boldsymbol{\Psi}_L' & \boldsymbol{\Psi}_1\boldsymbol{\Psi}_1' & \cdots & \rho_{1p}\boldsymbol{\Psi}_1\boldsymbol{\Psi}_p' \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{Lp}\boldsymbol{\Psi}_p\boldsymbol{\Psi}_L' & \rho_{1p}\boldsymbol{\Psi}_p\boldsymbol{\Psi}_1' & \cdots & \boldsymbol{\Psi}_p\boldsymbol{\Psi}_p' \end{pmatrix}. \tag{6}$$

To develop intuition, in Section 2.3 we refer to the case where $\boldsymbol{\Psi}_j, j = 1,\ldots,p$, and $\boldsymbol{\Psi}_L$ have elements that are zero off the diagonal, the case of no temporal dependence.

Joint Model

To use Eq. (6) association matrices, we restrict our attention to elliptical copulas and focus applications on the Gaussian copula. Extensions to other types of elliptical copulas (e.g., t-copulas) follow naturally.

It is convenient to introduce T_i, a variable that represents the time that the ith policyholder lapses. Specifically,

$$T_i = \begin{cases} 1 & \text{if } L_{i1} = 1 \\ 2 & \text{if } L_{i1} = 0, L_{i2} = 1 \\ \vdots & \quad \vdots \\ t & \text{if } L_{i1} = 0,\ldots,L_{i,t-1} = 0, L_{i,t} = 1 \\ \vdots & \quad \vdots \\ m & \text{if } L_{i1} = 0,\ldots,L_{i,m-1} = 0, L_{im} = 1 \\ m+1 & \text{if } L_{i1} = 0,\ldots,L_{i,m} = 0. \end{cases}$$

Now, suppose that we observe $T_i = t$ outcome periods. We will condition on the event $\{T_i = t\}$ that has probability

$$\begin{aligned} \Pr(T_i = t) &= \Pr\left(L_{i1} = 0,\ldots,L_{i,t-1} = 0, L_{i,t} = 1\right) \\ &= \begin{cases} C\left(F_{Li,1}(0),\ldots,F_{Li,m}(0)\right) & t = m+1 \\ C\left(F_{Li,1}(0),\ldots,F_{Li,t-1}(0),1,\ldots,1\right) & \\ \quad -C\left(F_{Li,1}(0),\ldots,F_{Li,t}(0),1,\ldots,1\right) & 1 \le t \le m \end{cases} \\ &= \begin{cases} C\left(\mathbf{a}_{im}(F_{Li,m}(0))\right) & t = m+1 \\ C\left(\mathbf{a}_{it}(1)\right) - C\left(\mathbf{a}_{it}(F_{Li,t}(0))\right) & 1 \le t \le m \end{cases}. \end{aligned} \tag{7}$$

The last equality uses some notation $\mathbf{a}_{it}(a) = \left(F_{Li,1}(0),\ldots,F_{Li,t-1}(0),a,1,\ldots,1\right)$ that we introduce to simplify the presentation. Calculation of these joint probabilities are straightforward, although tedious, from the marginals and the copula.

If a policy is renewed for all m periods, then $T_i = m+1$ and the observed likelihood is based on

$$\begin{aligned} &\Pr\left(T_i = m+1, \mathbf{Y}_{i1} \le \mathbf{y}_1,\ldots,\mathbf{Y}_{im} \le \mathbf{y}_m\right) \\ &= \Pr\left(L_{i1} = 0,\ldots,L_{i,m} = 0, \mathbf{Y}_{i1} \le \mathbf{y}_1,\ldots,\mathbf{Y}_{im} \le \mathbf{y}_m\right). \end{aligned} \tag{8}$$

If lapse occurs, then the observed likelihood is based on the distribution function

$$\begin{aligned} &\Pr\left(T_i = t, \mathbf{Y}_{i1} \le \mathbf{y}_1,\ldots,\mathbf{Y}_{it} \le \mathbf{y}_t\right) \\ &= \Pr\left(L_{i1} = 0,\ldots,L_{i,t-1} = 0, L_{i,t} = 1, L_{i,t+1} \le \infty,\ldots,L_{im} \le \infty,\right. \\ &\qquad \left.\mathbf{Y}_{i1} \le \mathbf{y}_1,\ldots,\mathbf{Y}_{it} \le \mathbf{y}_t, \mathbf{Y}_{i,t+1} \le \infty,\ldots,\mathbf{Y}_{im} \le \infty\right). \end{aligned} \tag{9}$$

Note that the evaluation of this likelihood involves a $m(p+1)$ dimensional copula. There are applications where this is computationally prohibitive and so standard maximum likelihood estimation is not available.

2.2.2. Conditional likelihood

Because of the difficulties in computing the observed likelihoods in Eqs. (8) and (9), we focus on the conditional distribution

$$\Pr\left(Y_{j,is} \le y_1, Y_{k,is} \le y_2 | T_i = t\right) = \frac{\Pr\left(T_i = t, Y_{j,is} \le y_1, Y_{k,is} \le y_2\right)}{\Pr\left(T_i = t\right)}. \tag{10}$$

Here, the time point s is chosen so that $s \le \min(t,m)$. The subscripts $\{j,k\}$ represent any pair chosen from $\{1,\ldots,p\}$. We have already discussed the computation of the denominator in Eq. (7). Computation of the numerator is similar; for the general case, it requires evaluation of a $m+2$ dimensional copula. As in Eq. (7),

$$\begin{aligned} &\Pr(T_i = t, Y_{j,is} \le y_1, Y_{k,is} \le y_2) \\ &= \begin{cases} C\left(\mathbf{a}_{im}(F_{Li,m}(0)), F_{j,is}(y_1), F_{k,is}(y_2)\right) & s < t \le m+1 \\ C\left(\mathbf{a}_{it}(1), F_{j,is}(y_1), F_{k,is}(y_2)\right) & \\ \quad -C\left(\mathbf{a}_{it}(F_{Li,t}(0)), F_{j,is}(y_1), F_{k,is}(y_2)\right) & s = t \le m \end{cases}. \end{aligned} \tag{11}$$

We again remark that the copula C in display (11) depends on the variables j and k selected as well as time points s and t. Specifically, we use Eq. (6) to express the association matrix for $\{L_{i1},\ldots,L_{i,m},Y_{j,is},Y_{k,is}\}$ as

$$\boldsymbol{\Sigma}=\begin{pmatrix} \boldsymbol{\Psi}_L\boldsymbol{\Psi}_L' & \rho_{Lj}\boldsymbol{\Psi}_L\boldsymbol{\Psi}_{j,s}' & \rho_{Lk}\boldsymbol{\Psi}_L\boldsymbol{\Psi}_{k,s}' \\ \rho_{Lj}\boldsymbol{\Psi}_{j,s}\boldsymbol{\Psi}_L' & 1 & \rho_{jk}\boldsymbol{\Psi}_{j,s}\boldsymbol{\Psi}_{k,s}' \\ \rho_{Lk}\boldsymbol{\Psi}_{k,s}\boldsymbol{\Psi}_L' & \rho_{jk}\boldsymbol{\Psi}_{j,s}\boldsymbol{\Psi}_{k,s}' & 1 \end{pmatrix}, \tag{12}$$

where $\boldsymbol{\Psi}_{j,s}$ is the jth row of $\boldsymbol{\Psi}_j$ and similarly for $\boldsymbol{\Psi}_k$.

The corresponding conditional hybrid probability density/mass functions follow as in Section 2.1.1. For $s<t\le m+1$, this is

$$f_{ijk,s|m+1}(y_1,y_2)=\frac{1}{\Pr(T_i=m+1)}\times$$
$$\begin{cases} \sum_{i_1=0}^1\sum_{i_2=0}^1(-1)^{i_1+i_2}\,C\left(\mathbf{a}_{im}(F_{Li,m}(0)),F_{j,is}(y_1^{(i_1)}),F_{k,is}(y_2^{(i_2)})\right) & y_1=y_1^d,y_2=y_2^d \\ \left\{\sum_{i_2=0}^1(-1)^{i_2}\,C_{m+1}\left(\mathbf{a}_{im}(F_{Li,m}(0)),F_{j,is}(y_1),F_{k,is}(y_2^{(i_2)})\right)\right\}f_{j,is}(y_1) & y_1=y_1^c,y_2=y_2^d \\ \left\{\sum_{i_1=0}^1(-1)^{i_1}\,C_{m+2}\left(\mathbf{a}_{im}(F_{Li,m}(0)),F_{j,is}(y_1^{(i_1)}),F_{k,is}(y_2)\right)\right\}f_{k,is}(y_2) & y_1=y_1^d,y_2=y_2^c \\ C_{m+1,m+2}\left(\mathbf{a}_{im}(F_{Li,m}(0)),F_{j,is}(y_1),F_{k,is}(y_2)\right)f_{j,is}(y_1)f_{k,is}(y_2) & y_1=y_1^c,y_2=y_2^c \end{cases} \tag{13}$$

Here, we use the notation introduced in Eq. (2) where $y^{(i)}=y$ if $i=0$ and $y^{(i)}=y-$ if $i=1$. Further, $C_{m+1}(u_1,\ldots,u_m,u_{m+1},u_{m+2})=\frac{\partial}{\partial u_{m+1}}C(u_1,\ldots,u_m,u_{m+1},u_{m+2})$ represents the partial derivative of the copula with respect to the second argument and similarly for C_{m+2}. The term $C_{m+1,m+2}$ is a second derivative with respect to the $m+1$st and $m+2$nd arguments. Further, $f_{j,is}$ is the density function corresponding to the distribution function $F_{j,is}$.

The corresponding conditional hybrid probability density/mass functions for $s=t\le m$, follows in the same way

$$f_{ijk,s|t}(y_1,y_2)=\frac{1}{\Pr(T_i=t)}\times$$
$$\begin{cases} \sum_{i_r=0}^1\sum_{i_1=0}^1\sum_{i_2=0}^1(-1)^{i_r+i_1+i_2}\left\{C\left(\mathbf{a}_{it}(F_{Li,t}(0)^{i_r}),F_{j,is}(y_1^{(i_1)}),F_{k,is}(y_2^{(i_2)})\right)\right\} & y_1=y_1^d,y_2=y_2^d \\ \left\{\sum_{i_r=0}^1\sum_{i_2=0}^1(-1)^{i_r+i_2}\left\{C_{m+1}\left(\mathbf{a}_{it}(F_{Li,t}(0)^{i_r}),F_{j,is}(y_1),F_{k,is}(y_2^{(i_2)})\right)\right\}\right\}f_{j,is}(y_1) & y_1=y_1^c,y_2=y_2^d \\ \left\{\sum_{i_r=0}^1\sum_{i_1=0}^1(-1)^{i_r+i_1}\left\{C_{m+2}\left(\mathbf{a}_{it}(F_{Li,t}(0)^{i_r}),F_{j,is}(y_1^{(i_1)}),F_{k,is}(y_2)\right)\right\}\right\}f_{k,is}(y_2) & y_1=y_1^d,y_2=y_2^c \\ \left\{\sum_{i_r=0}^1(-1)^{i_r}\,C_{m+1,m+2}\left(\mathbf{a}_{it}(F_{Li,t}(0)^{i_r}),F_{j,is}(y_1),F_{k,is}(y_2)\right)\right\}f_{j,is}(y_1)f_{k,is}(y_2) & y_1=y_1^c,y_2=y_2^c \end{cases} \tag{14}$$

2.2.3. Dropout GMM procedure

We are now in a position to extend the generalized method of moments, GMM, procedure introduced earlier to incorporate lapse. As before, let θ be an r-dimensional vector that represents the parameters that quantify the association among $\{L_{it},Y_{1,it},\ldots,Y_{p,it}\}$. Given $T_i=t$, the hybrid probability density/mass function of $Y_{j,it}$ and $Y_{k,it}$ is $f_{ijk,s|t}(\cdot,\cdot)$, as specified in Eqs. (13) and (14).

To estimate θ, for $s\le t$, define

$$g_{\theta,i,s,t}(y_1,y_2,T)=\mathrm{I}(T=t)\ \partial_\theta\ln f_{ijk,s|t}(y_1,y_2). \tag{15}$$

This is a mean zero random variable that contains information about θ. To see that it has mean zero,

$$\begin{aligned} \mathrm{E}\ g_{\theta,i,s,t}(Y_{j,is},Y_{k,is},T_i) &= \mathrm{E}\ \left[\mathrm{E}\ (g_{\theta,i,s,t}(Y_{j,is},Y_{k,is},T_i)|T_i=t)\right] \\ &= \mathrm{E}\ \left[\mathrm{I}(T_i=t)\mathrm{E}\ (\partial_\theta\ln f_{ijk,s|t}(Y_{j,is},Y_{k,is}))\right] \\ &= \mathrm{E}\ \left[\mathrm{I}(T_i=t)\cdot 0\right]=0, \end{aligned}$$

using properties of a likelihood. Thus, $g_{\theta,i,s,t}(Y_{j,is},Y_{k,is},T_i)$ is an unbiased estimator (of zero) and can be used in an estimating equation.

With the score function in Eq. (15), we can use the *GMM* procedure. The only thing that remains to do is to evaluate the score functions in terms of copula-based functions.

Dropout Score Evaluation

The dropout scores are similar to the scores introduced before except now we have a $m+2$ dimensional copula to evaluate instead of a 2-dimensional one. To provide additional intuition, we restrict ourselves to two random variables Y_1 and Y_2 that both follow a Tweedie distribution. These are non-negative random variables with continuous densities on the positive axis with a mass point at zero.

Thus, for $s<t\le m+1$ and two zero outcomes, this can be expressed as

$$\partial_\theta\ln f_{i12,s|m+1}(0,0)=\frac{\partial_\theta\ C\left(\mathbf{a}_{im}(F_{Li,m}(0)),F_{1,is}(0),F_{2,is}(0)\right)}{C\left(\mathbf{a}_{im}(F_{Li,m}(0)),F_{1,is}(0),F_{2,is}(0)\right)}-\frac{\partial_\theta\ C\left(\mathbf{a}_{im}(F_{Li,m}(0))\right)}{C\left(\mathbf{a}_{im}(F_{Li,m}(0))\right)}.$$

For a single positive outcome, $y>0$, we have

$$\partial_\theta\ln f_{i12,s|m+1}(y,0)=\frac{\partial_\theta C_{m+1}\left(\mathbf{a}_{im}(F_{Li,m}(0)),F_{1,is}(y),F_{2,is}(0)\right)}{C_{m+1}\left(\mathbf{a}_{im}(F_{Li,m}(0)),F_{1,is}(y),F_{2,is}(0)\right)}-\frac{\partial_\theta\ C\left(\mathbf{a}_{im}(F_{Li,m}(0))\right)}{C\left(\mathbf{a}_{im}(F_{Li,m}(0))\right)}$$

For two positive outcomes, $y_1>0$ and $y_2>0$, we have

$$\partial_\theta\ln f_{i12,s|m+1}(y_1,y_2)=\frac{\partial_\theta\ C_{m+1,m+2}\left(\mathbf{a}_{im}(F_{Li,m}(0)),F_{1,is}(y_1),F_{2,is}(y_2)\right)}{C_{m+1,m+2}\left(\mathbf{a}_{im}(F_{Li,m}(0)),F_{1,is}(y_1),F_{2,is}(y_2)\right)}-\frac{\partial_\theta\ C\left(\mathbf{a}_{im}(F_{Li,m}(0))\right)}{C\left(\mathbf{a}_{im}(F_{Li,m}(0))\right)}.$$

Scores for $s\le t=m$ follow in a similar way.

2.3. Dropout modeling with temporal independence

In this section, we assume no temporal dependence for dropout and outcomes. The purpose is to provide a context that is still helpful in applications and allows us to develop more intuition. To aid with developing intuition, we focus on the case where $p=2$. Even though there is no temporal dependence, we can use a copula function to capture dependence among dropout and outcomes. This specification permits, for example in our insurance setting, large claims to influence the tendency to lapse a policy or a latent variable to simultaneously influence both lapse and claims outcomes.

2.3.1. Joint model specification

With no temporal dependence, we assume that the random variables in $\{L_{it},\mathbf{Y}_{it}\}$ are independent over i and t. With the independence over time, the joint distribution function of the potentially observed variables is

$$\Pr\left(L_{i1}\le r_1,\ldots,L_{im}\le r_m,\mathbf{Y}_{i1}\le\mathbf{y}_1,\ldots,\mathbf{Y}_{im}\le\mathbf{y}_m\right)=\prod_{t=1}^m C\left(F_{Lit}(r_t),F_{1,it}(y_{1t}),F_{2,it}(y_{2t})\right).$$

As before, F_{Lit} and $F_{j,it}$ represent the marginal distributions of L_{it} and $Y_{j,it}$, $j=1,2$.

In the case of temporal independence, display (7) reduces to

$$\Pr(T_i=t)=\begin{cases} \prod_{s=1}^m F_{Li,s}(0) & t=m+1 \\ \left(1-F_{Li,t}(0)\right)\prod_{s=1}^{t-1}F_{Li,s}(0) & 1\le t\le m \end{cases}.$$

From this, we will be able to use dropouts to estimate the marginal distribution in the usual fashion. Intuitively, this is because the claims outcomes do not affect our ability to observe lapse outcomes. The converse is not true, lapses do affect our ability to observe claims.

However, this is not true of first period claims. These are always observed in this model and so provide the basis for consistent estimates of the claims marginal distributions.

If a policy is renewed for all m periods, then $T_i = m+1$ and with (8), the observed likelihood is based on

$$\Pr\left(T_i = m+1, \mathbf{Y}_{i1} \le \mathbf{y}_1, \dots, \mathbf{Y}_{im} \le \mathbf{y}_m\right) = \prod_{s=1}^{m} C\left(F_{Lis}(0), F_{1,is}(y_{1s}), F_{2,is}(y_{2s})\right).$$

If lapse occurs, then $T_i \le m$ and with (9), the observed likelihood is based on the distribution function

$$\Pr\left(T_i = t, \mathbf{Y}_{i1} \le \mathbf{y}_1, \dots, \mathbf{Y}_{it} \le \mathbf{y}_t\right) = \left\{C\left(1, F_{1,it}(y_{1t}), F_{2,it}(y_{2t})\right) - C\left(F_{Lit}(0), F_{1,it}(y_{1t}), F_{2,it}(y_{2t})\right)\right\} \prod_{s=1}^{t-1} C\left(F_{Lis}(0), F_{1,is}(y_{1s}), F_{2,is}(y_{2s})\right).$$

Thus, the corresponding conditional distribution function is

$$\Pr\left(Y_{1,is} \le y_1, Y_{2,is} \le y_2 | T_i = t\right) = \begin{cases} \frac{C\left(F_{Lis}(0), F_{1,is}(y_1), F_{2,is}(y_2)\right)}{F_{Lis}(0)} & \text{for } s < t \le m+1 \\ \frac{C\left(1, F_{1,is}(y_1), F_{2,is}(y_2)\right) - C\left(F_{Lis}(0), F_{1,is}(y_1), F_{2,is}(y_2)\right)}{1 - F_{Lis}(0)} & \text{for } s = t \le m \end{cases}. \tag{16}$$

This is more intuitive than the general expressions given in Section 2.2.2. Note that the evaluation of this distribution function involves a 3 dimensional copula.

Expressions for the conditional density/mass functions and the *GMM* scores follow a similar pattern and are omitted here.

2.3.2. Missing at random

The no temporal dependence assumption provides one additional benefit; the decision to lapse turns out to be *ignorable* in the sense that the response mechanism does not affect inference about claims. This is in spite of the fact that we still allow for a dependency between claims and response; we note that this is not consistent with the usual biomedical literature on joint models of longitudinal and time-to-event data, c.f., Elashoff et al. (2017). Intuitively, this is because, in the insurance setting, we postulate a dependency between a claim during the tth year, Y_t, and a decision lapse or renew **at** time t, L_t. In contrast, in biomedical applications, the concern is for dependencies between whether an outcome (claims) is observed and a variable to indicate whether it is observed. In our notation, if Y_t is the random variable in question, then L_{t-1} indicates whether or not it is observed (lapse/renewal at time $t-1$).

More formally, consider a joint probability mass/density function of the form $Like^* = f(Y, L|\mathbf{X}, \theta)$ where $\mathbf{X}$ are covariates and θ are parameters. By conditioning, write this as $Like^* = \prod_{s=1}^{m} f(Y_s, L_s|\mathbf{X}, \theta, H_s)$ with history $H_s = \{Y_1, \dots, Y_{s-1}, L_1, \dots, L_{s-1}\}$. On the event $\{T = t\} = \{L_1 = 1, \dots, L_{t-1} = 0, L_t = 1, \dots, L_m = 1\}$, we have

$$\begin{aligned} Like^* &= \left(\prod_{s=1}^{t-1} f(Y_s, L_s = 0|\mathbf{X}, \theta, H_s)\right) f(Y_t, L_t = 1|\mathbf{X}, \theta, H_s) \times \left(\prod_{s=t+1}^{m} f(Y_s, L_s = 1|\mathbf{X}, \theta, H_s)\right) \\ &= \left(\prod_{s=1}^{t-1} f(Y_{obs,s}, L_s = 0|\mathbf{X}, \theta, H_s)\right) f(Y_{obs,t}, Y_{mis,t}, L_t = 1|\mathbf{X}, \theta, H_s) \left(\prod_{s=t+1}^{m} f(Y_{mis,s}, L_s = 1|\mathbf{X}, \theta, H_s)\right), \end{aligned}$$

where Y_{obs}, Y_{mis} signals that the claim is observed or missing (unobserved), respectively. In the last term, the event $\{L_s = 1\}$ is known given H_s due to the monotonicity of lapsation. Further, because of independence, the last term is

$$\prod_{s=t+1}^{m} f(Y_{mis,s}, L_s = 1|\mathbf{X}, \theta, H_s) = \prod_{s=t+1}^{m} f(Y_{mis,s}|\mathbf{X}, \theta)$$

which can be integrated out to get the (observed) likelihood. In our insurance sampling scheme, the middle term is $f(Y_{obs,t}, Y_{mis,t}, L_t = 1|\mathbf{X}, \theta, H_s) = f(Y_{obs,t}, L_t = 1|\mathbf{X}, \theta, H_s)$. Thus, the integrated likelihood is

$$Like = \left(\prod_{s=1}^{t-1} f(Y_{obs,s}, L_s = 0|\mathbf{X}, \theta, H_s)\right) f(Y_{obs,t} L_t = 1|\mathbf{X}, \theta, H_s)$$

which is the observed likelihood. This establishes our claim that the lapse decision is ignorable in the case of temporal independence.

2.3.3. Comparing trivariate likelihood to GMM estimators

Similar to the simulation study cited at the end of Section 2.1.3, we consider a sample of n policyholders that are potentially observed over 5 years. In addition to the claims, we have five rating (explanatory) variables: (a) x_1, a binary variable that indicates whether or not an attribute holds, (b) x_2, x_3, x_4, are generic continuous explanatory variables, and (c) x_5 is a time trend ($x_{it} = t$).

For the claims variables, we used a logarithmic link to form the mean claims $\mu_{j,it} = \exp\left(\mathbf{x}_{it}'\boldsymbol{\beta}_j\right)$, $j = 1, 2$. For the lapse variable, $\pi_{it} = \exp\left(\mathbf{x}_{it}'\boldsymbol{\beta}_L\right) / \left(1 + \exp\left(\mathbf{x}_{it}'\boldsymbol{\beta}_L\right)\right)$, is the expected value, a common form for the logit model. We use a negative coefficient associated with the time trend variable to reflect the fact that lapse probabilities tend to decrease with policyholder duration.

Each type of claims was simulated using the Tweedie distribution, a mean, and two other parameters, ϕ_j (for dispersion) and P_j (the "power" parameter). Recall, for a Tweedie distribution, that the variance is $\phi_j \mu^P$; we use $P = 1.67$ in this study.

Dependence among claims was taken to be a Gaussian copula with the following structure

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \rho_{L1} & \rho_{L2} \\ \rho_{L1} & 1 & \rho_{12} \\ \rho_{L2} & \rho_{12} & 1 \end{pmatrix}.$$

For example, we might use positive values for the association between lapse and claims (ρ_{L1} and ρ_{L2}) so that large claims are associated with higher lapse. We might use a positive value for the association between claim types (ρ_{12}).

Table 2 summarizes the performance of the *GMM* estimators lapse estimator by varying the sample size and dispersion parameters, ϕ_1, and ϕ_2. For this table, we used $\rho_{L1} = -0.2$, $\rho_{L2} = 0.2$, and $\rho_{12} = 0.1$ for the association parameters, these being comparable to the results of our empirical work. For smaller samples, $n = 100, 250$, we used 500 simulations to make sure that the bias was being determined appropriately. This was less of a concern with larger sample sizes, $n = 500, 1000, 2000$, and so for convenience we used 100 simulations in this study.

Some aspects of the results are consistent with Section 2.1.3 small sample study which compares the pairwise likelihood and GMM estimators study (without lapse). As the dispersion parameters ϕ increase, there are more discrete observations resulting in larger biases and standard errors for all sample sizes. The magnitude of biases suggests that our general procedure may not be suitable for sample sizes as small as $n = 100$. However, even for $n = 250$ (and above), we deem their performance acceptable on the bias criterion.

For the standard error criterion, we view the smaller sample sizes $n = 100, 250$ as unacceptable. For example, if $n = 100$, $\phi_1 = \phi_2 = 500$, and $\rho_{L1} = -0.2$, it is hard to imagine recommending a procedure where the average standard error is 0.158. Only for nearly continuous data, when $\phi_1 = \phi_2 = 2$, do the standard errors seem desirable with $n = 500$. In general, for more discrete data where $\phi_1 = \phi_2 = 500$, we recommend samples sizes of $n = 2,000$ and more. Most users that we work with are

Table 2
Summary of the *GMM* Lapse Estimators.

Num Sim	n	$\phi_1 = \phi_2$	Bias $L1$	Bias $L2$	Bias 12	Standard error $L1$	Standard error $L2$	Standard error 12
500	100	2	0.000	0.003	−0.001	0.092	0.091	0.049
500	100	42	−0.006	−0.011	−0.002	0.096	0.094	0.053
500	100	500	−0.023	−0.006	−0.035	0.158	0.139	0.135
500	250	2	−0.001	−0.002	0.001	0.059	0.058	0.031
500	250	42	0.005	0.002	0.000	0.062	0.060	0.034
500	250	500	−0.004	−0.012	−0.011	0.107	0.091	0.086
100	500	2	−0.001	0.000	−0.002	0.042	0.041	0.022
100	500	42	−0.002	−0.005	−0.003	0.044	0.043	0.024
100	500	500	−0.011	−0.009	−0.008	0.076	0.064	0.061
100	1000	2	−0.002	0.002	0.000	0.030	0.029	0.015
100	1000	42	0.009	0.001	−0.002	0.031	0.030	0.017
100	1000	500	−0.002	−0.003	−0.003	0.054	0.045	0.044
100	2000	2	0.000	0.001	0.001	0.021	0.021	0.011
100	2000	42	0.004	−0.005	0.001	0.022	0.021	0.012
100	2000	500	0.004	0.000	−0.003	0.038	0.032	0.031

primarily interested in point estimates but also want to say something about statistical significance.

3. Results

3.1. Insurance lapsation: a case study

Life insurers enter into contracts that can last many years (e.g., a 20 year old purchasing a policy who dies at age 100 has an 80 year contract). In contrast, property and casualty (also known as general and as property and liability) insurers write contracts of shorter durations, typically six months or a year. In both cases, the company anticipates policyholders will retain a relationship with the insurer for many years and they track when and why policyholders leave, or lapse. Although policyholders may depart at any time, lapsation tends to occur at a periodic premium payment times and so we follow standard industry practice and treat time as discrete.

Why do insurers worry about lapsation? They seek to (a) retain profitable customers for a longer period of time, (b) achieve profit margins on new customers within short time, and (c) achieve higher profit margins on existing customers. See, for example, Brockett et al. (2008), Guillén et al. (2003), and Guillén et al. (2012).

There is a natural struggle between the price of an insurance risk and loyalty. The higher the price, the higher is the probability to lapse the policy. Despite this relationship, insurers typically separate the processes of calculating price and renewal prospects. They use standard event-time models to understand characteristics of policyholders that drive renewal or lapsation.

It does make sense that a large claim in one period would imply a large claim in the next period. However, for our data, as well as many on personal insurance (auto and homeowners data), with about 80–95 percent zeros, limitations imposed by deductibles and upper limits, it is simply very difficult to calibrate time dependent associations. For this reason, in this section we present the results of the trivariate model assuming time independence.

3.2. Data and variable descriptions

This paper examines longitudinal data from a major Spanish insurance company that offers automobile and homeowners insurance. As in many countries, vehicle owners in Spain are obliged to have some minimum form of insurance coverage for personal injury to third parties. Homeowners insurance, on the other hand, is optional. The data set tracks 40,284 clients over five years, between 2010 and 2015, who subscribed to both automobile and homeowners insurance.

Table 3
Number of Policies and Lapse by year.

	2010	2011	2012	2013	2014
Number of customers at the beginning of the period	40284	29818	22505	17044	13284
Number of customers that cancel at least one policy per period	10466	7313	5461	3760	2296
Rate of lapsation for at least one policy (%)	26%	25%	24%	22%	17%

Table 4
Variable descriptions.

Variable	Description
Year	
Gender	1 for male, 0 for female
Age_client	Age of the customer
Client_Seniority	The number of years with the company
metro_code	1 for urban or metropolitan, 0 for rural
Car_power_M	Power of the car
Car_2ndDriver_M	Presence of a second driver
Policy_PaymentMethodA	1 for annual payment, 0 for monthly payment
Insuredcapital_continent_re	Value of the property
appartment	1 for apartment, 0 for houses or semi-attached houses
Policy_PaymentMethodH	1 for annual payment, 0 for monthly payment

From the unbalanced panel of policyholders over 5 years, there are $N = 122,935$ observations in the data set. Table 3 summarizes lapse behavior. The Spanish market is competitive; the table shows 29,296 lapses, for a lapse rate of 23.8%.

Our database includes variables that are commonly used for determining prices and understanding lapse behavior. These include (a) customer characteristics such as age and gender, (b) vehicle characteristics for auto insurance, (c) information on property for homeowners, and (d) renewal information such as the date of renewal. Table 4 provides variable descriptions.

Table 5 gives basic frequency and severity summary statistics for auto and homeowners claims. For type 1 (auto) claims, we have 1,967 or about 1.6% claims. For type 2 (home) claims, we have 2,189 or about 1.78% claims.

3.3. Marginal model fits

After extensive diagnostic testing, we fit standard generalized linear models to the three outcome variables, comparable to insurance industry practice. Specifically, we fit a logistic model for lapse and Tweedie regression models for auto and homeowners claims. Table 6 summarizes estimates of marginal model fits.

3.4. Interpreting the results

Using the marginal fitted regression models as inputs, we then computed the generalized method of moments association parameters. The results are summarized in Table 7.

These results show that even after controlling for the effects of the characteristics of the client, vehicles and homes, there is still evidence of relationship between claims for auto insurance, claims for home insurance and the renewal behavior.

Specifically, for a customer with a claim, there is a higher tendency to lapse. From the consumers perspective, a claim may induce a search for a new company (possibly for fear of experience rating induced premium increases). With no claims, insureds may be content to remain loyal to a company (perhaps simply inertia). From the company perspective, a claim is an alert for non-renewal, something that the insurer can take action upon should they wish to.

Table 5
Claim summary statistics.

	2010	2011	2012	2013	2014
Clients with positive claims - Auto	769	547	318	209	124
Average number of claims - Auto	0.04	0.03	0.03	0.02	0.02
Average claim amount - Auto (Euros)	1539.99	1689.84	2031.2	1629.18	1222.13
Clients with positive claims - Home	660	531	448	310	240
Average number of claims - Home	0.03	0.03	0.04	0.03	0.03
Average claim amount - Home (Euros)	447.85	501.59	410.73	348.1	508.86

We also note that the association between auto and home claims has implications for portfolio management and reserving practices — these two risks are not independent.

Academic pricing models are anchored to the notion that personal lines is a short-term business because contracts typically are 6 months or a year. These pricing models are based on costs of insurance so from an economics perspective can be thought of as "supply-side" models. In contrast, insurers consider non-renewals a critical disruption for their business — losing a customer implies that they will not see benefits in renewal years. Although the contract is only for 6 months or a year, customers are likely to renew and this is factored into models of insurer profitability. In fact, some insurers explicitly seek to maximize what is known as "lifetime customer value" where renewal is a key feature. Because renewal is about who buys insurance, from an economics perspective it can be thought of as "demand-side" modeling. Our joint model of renewal and insurance risk may play a key role in developing models that balance demand and supply side considerations of personal insurance lines.

3.5. Effect on the insurance premium

The premium for the next period of an insurance policy is calculated assuming the insured will renew their policy. However, the renewal of the policy is not a certain event, as we show in this paper, it has an associated probability. Therefore, to support marketing strategies, the premium should be obtained by considering (i) the cost and (ii) the fact that the insured may or may not renew the policy, i.e. the premium would have to be calculated as $\mathrm{E}\left(Y_{j,it}\left(1-L_{j,it}\right)\right)$ for $j = Auto,\ Home$. For given distributions of random variables $Y_{j,it}$ and $L_{j,it}$, factorizing the previous expression requires independence between these two variables. Directly, if $Y_{j,it}$ and $L_{j,it}$ are uncorrelated then the pure premium is equal to the product of expectation, i.e. $\mathrm{E}\left(Y_{j,it}\right)\mathrm{E}\left(1-L_{j,it}\right)$. On the contrary if the correlation (ρ_{jL}) is different from zero, the pure premium is equal to:

$$\mathrm{E}\left(Y_{j,it}\right)\mathrm{E}\left(1-L_{j,it}\right)-\rho_{jL}\mathrm{SD}_{Y_{j,it}}\sqrt{\mathrm{E}\left(L_{j,it}\right)\mathrm{E}\left(1-L_{j,it}\right)},$$

where $\mathrm{SD}_{Y_{j,it}}$ is the standard deviation of the total claim cost random variable for the jth line of business. A difficulty that we can find when the previous expression is applied to calculate a premium is that negative values can be obtained. In this case, cost dispersion is very high and it could indicate that, even if we reduce the premium, we may not be able to retain the insured and, furthermore, we would increase our losses.

Using the results for marginal distributions that were shown in Section 3.3, the premiums for auto and home insurance policies are calculated assuming that the correlation between lapse and total cost is zero and with the estimated correlations that are shown in Table 7. We calculated the premium for three profiles of policyholders: with low risk, medium risk and high risk, respectively. To calculate the probability of lapse, we assume that gender is male, but for females the results are similar, the coefficient of gender is not significant.

We show the results for home insurance data. The quotient between pure premiums with and without correlation and the values of the explanatory variables for the three profiles are given in Table 8. The values of the quotient in the last row of Table 8 indicate that the low risk needs a greater discount for retaining the policyholder. For auto data, the premiums calculated considering correlation between cost and lapse probability would become negative, showing that correlation between premiums and loyalty is strong for this line of insurance. This effect gets exacerbated by a large standard deviation in the claim cost variable.

Premium estimation results indicate that a good strategy that an insurance company could use with policyholders that have several policies, such as auto and home, simultaneously would be to reduce premiums of home insurance for low risk profiles. This could ensure that they renew their home policy and, indirectly, they would renew their auto policy.

Table 6
Marginal models fits.

	Logistic – Lapse		Tweedie – Auto		Tweedie – Home	
	Estimate	t value	Estimate	t value	Estimate	t value
(Intercept)	0.324	9.39	21.132	5.58	−2.794	−2.58
Year	−0.078	−15.19	−1.179	−2.68	−0.015	−0.32
Gender	0.097	5.89				
Age_client	−0.023	−41.18	−0.421	−7.59	0.013	2.60
Client_Seniority	−0.006	−4.50	0.170	0.63	−0.004	−0.33
metro_code	0.163	9.13	−4.356	−2.25	0.261	1.57
Car_power_M			0.122	23.54		
Car_2ndDriver_M			−2.354	−0.49		
Policy_PaymentMethodA			3.542	2.68		
Insuredcapital_continent_re					0.348	4.25
appartment					1.097	7.16
Policy_PaymentMethodH					−0.765	−3.40

Table 7
GMM association estimates.

	Estimate	Std Error	t value
Lapse-Auto	0.101	0.007	14.14
Lapse-Home	0.069	0.011	6.11
Auto-Home	0.118	0.029	4.10

Table 8
Risk profiles and quotient between premium with/without correlation in home insurance.

	Low risk	Medium risk	High risk
Age_client	30	60	20
Client_Seniority	5	15	1
metro_code	1	0	1
Insuredcapital_continent_re	8	15	3
appartment	1	0	1
Policy_PaymentMethodH	1	1	0
Quotient	0.44	0.68	1

4. Conclusions

Motivated by insurance applications, this paper has introduced the *GMM* procedure to estimate association parameters in complex copula regression modeling situations where the marginals may be a hybrid combination of discrete and continuous components. Because the *GMM* scores use only a low-dimensional subset of the data, this procedure is available in high-dimensional situations. Thus, it provides an alternative to the vine copula methodology, c.f., Panagiotelis et al.

(2012). Compared to the *GMM* approach, the advantage of vines is the flexibility in model specification where many different sub-copula models may be joined to represent the data. The comparative advantage of the *GMM* approach may be the relative efficiency. For example, we have seen in Section 2.1 that the *GMM* approach is more efficient than the pairwise likelihood. A strength of the *GMM* approach is that, in Section 2, it uses scores based on pairs of relationships and so is able to take advantage of technologies developed for paired copulas, e.g., Schepsmeier and Stöber (2014).

Moreover, the *GMM* approach was extended in Sections 2.2–3 to handle problems of data attrition. Here, one needs to condition on the observation process and so at least three observations are required to calibrate associations among claims. Section 2.2 provides a detailed theory on handling general longitudinal data with explicit expressions in Section 2.3 for temporally independent, yet still dependent censoring, of observations. In work available from the authors, we also extend some of the pairwise technologies to multiple dimensions and make connections with an earlier literature on sensitivity of association parameters based on work of Plackett (1954).

We believe that the artificial intelligence community will play an enormous role in the analysis of insurance customers, with the perspective presented in this paper. It is currently a challenge for insurers to find the true value of their business portfolios because they need to combine (i) what they call the "technical value", meaning expected returns based solely on the difference between premiums and claims, possibly also considering additional managerial, marketing and regulatory expenses, and (ii) the possibility to retain the customer or even cross-sell additional products. Our paper serves to present this case study as an area where more contributions are expected in the near future.

Insurance analysts employ regression techniques to address the heterogeneity of customers. Copulas provide an interpretable way of incorporating dependence, that is fundamental to insurance operations, while preserving marginal models of claims outcomes. As in many social science and biomedical fields where subjects are followed longitudinally, problems of attrition arise that, in the insurance context, are known as lapsation. In this paper, we argue that a copula approach that allows one to handle attrition, in addition to many other potential sources of dependence, is a natural modeling strategy for analysts to adopt.

CRediT authorship contribution statement

Edward W. Frees: Conceptualization, Methodology, Software, Investigation, Writing – original draft, Writing – review & editing. **Catalina Bolancé:** Data curation, Writing – original draft, Methodology, Validation, Writing – review & editing. **Montserrat Guillen:** Conceptualization, Data curation, Writing – original draft, Investigation, Funding acquisition, Visualization, Writing – review & editing. **Emiliano A. Valdez:** Supervision, Conceptualization, Methodology, Investigation, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

C.B. and M.G. thank the Spanish Ministry of Science, grant PID2019-105986GB-C21- M.G. thanks ICREA Academia.

References

Abbasi, Babak, & Guillen, Montserrat (2013). Bootstrap control charts in monitoring value at risk in insurance. *Expert Systems with Applications*, *40*(15), 6125–6135.

Abrahams, Alan S., Becker, Adrian B., Sabido, Daniel, D'Souza, Rosskyn, Makriyiannis, George, & Krasnodebski, Michal (2009). Inducing a marketing strategy for a new pet insurance company using decision trees. *Expert Systems with Applications*, *36*(2, Part 1), 1914–1921.

Ahn, Kiok, Hoque, Md. Monirul, Song, Gihun, Abdullah-Al-Wadud, M., & Chae, Oksam (2015). Learning discriminant DCT coefficients driven block descriptor for digital dropout detection system in degraded archived media. *Expert Systems with Applications*, *42*, 5811–5821.

Bayam, Evrim, Liebowitz, Jay, & Agresti, William (2005). Older drivers and accidents: A meta analysis and data mining application on traffic accident data. *Expert Systems with Applications*, *29*(3), 598–629.

Bolancé, Catalina, Guillen, Montserrat, Nielsen, Jens Perch, & Thuring, Fredrik (2018). Price and profit optimization for financial services. *Risks*, *6*(1), 1–12.

Brockett, Patrick L., Golden, Linda L., Guillen, Montserrat, Nielsen, Jens Perch, Parner, Jan, & Perez-Marin, Ana Maria (2008). Survival analysis of a household portfolio of insurance policies: How much time do you have to stop total customer defection? *The Journal of Risk and Insurance*, *75*(3), 713–737.

Cheng, Min-Yuan, Peng, Hsien-Sheng, Wu, Yu-Wei, & Liao, Yi-Hung (2011). Decision making for contractor insurance deductible using the evolutionary support vector machines inference model. *Expert Systems with Applications*, *38*(6), 6547–6555.

Elashoff, Robert M., Li, Gang, & Li, Ning (2017). *Joint modeling of longitudinal and time-to-event data*. CRC Press.

Eling, M., Nuessle, D., & Staubli, J. (2002). The impact of artificial intelligence along the insurance value chain and on the insurability of risks. *The Geneva Papers on Risk and Insurance - Issues and Practice*, *8*.

Frees, Edward W., & Wang, Ping (2005). Credibility using copulas. *North American Actuarial Journal*, *9*(2), 31–48.

Guillén, Montserrat, Nielsen, Jens Perch, Scheike, Thomas H., & Pérez-Marín, Ana Maria (2012). Time-varying effects in the analysis of customer loyalty: A case study in insurance. *Expert Systems with Applications*, *39*(3), 3551–3558.

Guillén, Montserrat, Parner, Jan, Densgsoe, Chresten, & Perez-Marin, Ana M. (2003). Using logistic regression models to predict and understand why customer leave an insurance company. In *Intelligent techniques in the insurance industry: Theory and applications* (pp. 465–490). London: World Scientific.

Harrer, Stefan, Shah, Pratik, Antony, Bhavna, & Jianying, Hu (2019). Artificial intelligence for clinical trial design. *Trends in Pharmacological Sciences*, *40*(8), 577–591.

Hiabu, Munir, Margraf, Carolin, Martínez-Miranda, Maraí Dolores, & Nielsen, Jens Perch (2016). Cash flow generalisations of non-life insurance expert systems estimating outstanding liabilities. *Expert Systems with Applications*, *45*, 400–409.

Hou, Zhongsheng, & Bu, Xuhui (2018). Model free adaptive control with data dropouts. *Expert Systems with Applications*, *38*, 10709–10717.

Hsiao, Shu-Hua, & Whang, Thou-Jen (2009). A study of financial insolvency prediction model for life insurers. *Expert Systems with Applications*, *36*(3, Part 2), 6100–6107.

Ibrahim, Joseph G., Chu, Haitao, & Chen, Liddy M. (2010). Basic concepts and methods for joint models of longitudinal and survival data. *Journal of Clinical Oncology*, *28*(16), 2796–2801.

Imriyas, Kamardeen (2009). An expert system for strategic control of accidents and insurers' risks in building construction projects. *Expert Systems with Applications*, *36*(2, Part 2), 4021–4034.

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

Joe, Harry (2014). *Dependence modeling with copulas*. CRC Press.

Kaishev, Vladimir K., Nielsen, Jens Perch, & Thuring, Fredrik (2013). Optimal customer selection for cross-selling of financial services products. *Expert Systems with Applications*, *40*, 1748–1757.

Lambert, Philippe (1996). Modelling irregularly sampled profiles of non-negative dog triglyceride responses under different distributional assumptions. *Statistics in Medicine*, *15*(15), 1695–1708.

Lambert, Philippe, & Vandenhende, Francois (2002). A copula-based model for multivariate non-normal longitudinal data: Analysis of a dose titration safety study on a new antidepressant. *Statistics in Medicine*, *21*(21), 3197–3217.

Liu, Lei (2009). Joint modeling longitudinal semi-continuous data and survival, with application to longitudinal medical cost data. *Statistics in Medicine*, *28*(6), 972–986.

Meester, Steven G., & Mackay, Jock (1994). A parametric model for cluster correlated categorical data. *Biometrics*, *50*, 954–963.

Meyer, Marc H., De Tore, Arthur, Siegel, Stephen F., & Curley, Kathleen F. (1992). The strategic use of expert systems for risk management in the insurance industry. *Expert Systems with Applications*, *5*(1), 15–24.

Njagi, Edmund Njeru, Molenberghs, Geert, Kenward, Michael G., Verbeke, Geert, & Rizopoulos, Dimitris (2014). A characterization of missingness at random in a generalized shared-parameter joint modeling framework for longitudinal and time-to-event data, and sensitivity analysis. *Biometrical Journal*, *56*(6), 1001–1015.

Panagiotelis, Anastasios, Czado, Claudia, & Joe, Harry (2012). Pair copula constructions for multivariate discrete data. *Journal of the American Statistical Association*, *107*(499), 1063–1072.

Papageorgiou, Grigorios, Mauff, Katya, Tomer, Anirudh, & Rizopoulos, Dimitris (2019). An overview of joint modeling of time-to-event and longitudinal outcomes. *Annual Review of Statistics and its Application*, *6*, 223–240.

Pathak, J., Vidyarthi, N., & Summers, S. L. (2005). A fuzzy-based algorithm for auditors to detect elements of fraud in settled insurance claims. *Managerial Auditing Journal*, *6*, 632–644.

Plackett, Robin L. (1954). A reduction formula for normal multivariate integrals. *Biometrika*, *41*(3/4), 351–360.

Prokhorov, Artem, & Schmidt, Peter (2009). Likelihood-based estimation in a panel setting: Robustness, redundancy and validity of copulas. *Journal of Econometrics*, *153*(1), 93–104.

Rüschendorf, Ludger (2009). On the distributional transform, Sklar's theorem, and the empirical copula process. *Journal of Statistical Planning and Inference*, *139*(11), 3921–3927.

Schepsmeier, Ulf, & Stöber, Jakob (2012). *Web supplement: Derivatives and Fisher information of bivariate copulas*: *Technical report*, TU München.

Schepsmeier, Ulf, & Stöber, Jakob (2014). Derivatives and Fisher information of bivariate copulas. *Statistical Papers*, *55*(2), 525–542.

Song, Xue-Kun (Peter) (2007). *Correlated data analysis: Modeling, analytics, and applications*. Springer Science & Business Media.

Song, Xue-Kun (Peter), Li, Mingyao, & Yuan, Ying (2009). Joint regression analysis of correlated data using Gaussian copulas. *Biometrics*, *65*(1), 60–68.

Subelj, Lovro, Furlan, Stefan, & Bajec, Marko (2011). An expert system for detecting automobile insurance fraud using social network analysis. *Expert Systems with Applications*, *38*(1), 1039–1052.

Thuring, F., Nielsen, J. P., Guillen, M., & Bolancé, C. (2012). Selecting prospects for cross-selling financial products using multivariate credibility. *Expert Systems with Applications*, *39*(10), 8809–8816.

Tiao, George C., & Box, George E. P. (1981). Modeling multiple time series with applications. *Journal of the American Statistical Association*, *76*(376), 802–816.

Tsiatis, Anastasios A., & Davidian, Marie (2004). Joint modeling of longitudinal and time-to-event data: An overview. *Statistica Sinica*, 809–834.

Verbeke, Geert, Fieuws, Steffen, Molenberghs, Geert, & Davidian, Marie (2014). The analysis of multivariate longitudinal data: A review. *Statistical Methods in Medical Research*, *23*(1), 42–59.

Wooldridge, Jeffrey M. (2010). *Econometric analysis of cross section and panel data*. MIT Press.