Joint Models of Insurance Lapsation and Claims

Edward W. Frees, Catalina Bolancé, Montserrat Guillen, Emiliano Valdez *

October 11, 2018

Abstract

This paper introduces a generalized method of moments technique to estimate dependence parameters in a multivariate copula regression framework. This framework is motivated by insurance applications, where it is common to have outcomes, such as insurance claims, that may be from a discrete, a long-tailed continuous, or a hybrid combination of discrete and continuous distributions. Copula methods are natural for modeling dependence when the outcome distribution may be complicated especially in the presence of covariates.

The paper also considers extensions to handle insurance lapsation. In the insurance context, claims outcomes may be related to a policyholder’s decision to lapse (or the converse, renew) a policy. Motivated by insurance, this paper introduces a copula model where longitudinal outcomes may be related to lapsation, a time-to-event random variable, that accounts for this dependence. A simulation study demonstrates the viability of the approach. Also considered is a sample from a Spanish insurer that includes auto and homeowners claims. The paper describes how the joint model provides new information that insurers can use to better manage their portfolios of risks.

Keywords and Phrases: Copula, joint longitudinal and time-to-event, generalized method of moments

*Edward W. Frees, University of Wisconsin-Madison. Catalina Bolancé, Universitat de Barcelona. Montserrat Guillen, Universitat de Barcelona. Emiliano Valdez, University of Connecticut.
1 Introduction

In traditional insurance modeling, the focus has been on understanding the distribution of an insurance claim $Y$ which can be complicated. It is complicated in part because there may be many predictor or rating variables $x$ that can be used to calibrate the distribution. For example, in auto insurance, analysts may have available characteristics of the driver (e.g., age, sex), of the vehicle (e.g., make/brand/model of car, cubic capacity), and geographic information (e.g., where the car is garaged). Another complication is that claims distributions are typically skewed and long-tailed so that there is a significant probability of large losses. Moreover, distributions are often a hybrid combination of continuous and discrete components. For example, analysts commonly examine claims distribution where there is a mass at zero, corresponding to no claims. Further, claims may have an upper bound due to policy coverage limits; this right censoring induces mass probabilities at the policy limits. In this same way, information may not be available due to no information about claims less than a deductible; this also induces discreteness in a likelihood calculation. Handling these many complications is now a routine aspect of statistical analysis of claims distribution. See, for example, Frees et al. (2014).

Copula Modeling

Less routine is modeling dependence. At a basic level, modeling dependence is critical for insurance. Insurance systems are predicated on the pooling of risks. Insurers pool risks in order to enjoy the benefits of diversification; but, those benefits depend upon relationships among risks.

For notation, suppose that the insurer wishes to understand the joint distribution of $p$ risks, $Y_1, \ldots, Y_p$. In this paper, we explore copulas, a general tool for modeling dependencies
and so express the joint distribution as

\[ F(y_1, \ldots, y_p) = \Pr(Y_1 \leq y_1, \ldots, Y_p \leq y_p) = C(F_1(y_1), \ldots, F_p(y_p)). \] (1)

Here, \( F_1(y_1) = \Pr(Y_1 \leq y_1) \) 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 rating variables, \( x \), that is available to calibrate the marginal distributions.

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 with motivating insurance applications in Online Supplement 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.
Insurance Lapsation

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, Guillén et al. (2003), Brockett et al. (2008), 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.

Joint Models of Longitudinal and Time-to-Event Data

A special case of multivariate regression is when subjects are followed over time and the $p$ outcomes refer to observations at time points $t = 1, \ldots, p$. This case is known as longitudinal or panel data. In the longitudinal/panel data case, it is common to observe unbalanced data. Typically, the assumption is that missingness is due to a process unrelated to the outcomes, that is, the data are said to be “missing at random.”
However, there are important situations where outcomes may influence the observation process. As an example from the biomedical field, assume that outcomes are a function of patient medical expenditures which 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 (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, Tsiatis and Davidian (2004) and Ibrahim et al. (2010) provide overviews and Rizopoulos et al. (2010) introduces an R package.

Much of this literature is centered around one or more (time-constant) latent variables that induce an association between survival/drop-out and amounts (e.g., costs). A strength of this approach is that latent variables are intuitively appealing, one can ascribe an attribute to them (e.g., “frailty”) to help interpret dependence. A limitation of this approach is that the marginal model structure is not maintained (in the non-linear case), making it more difficult to use data for model identification. As an alternative, we explore copula modeling.

**Plan of the Paper.** Section 3 extends the copula regression approach where the censoring mechanism, lapsation, may depend on other outcomes. This section is an extension, and not a special case, of Section 2 because now we allow the observation process to depend on outcomes as do the joint models of longitudinal and time-to-event data. Section 4 focuses on the special case of temporal independence. Section 5 describes an empirical application. Section 6 concludes.

## 2 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, this general model fitting strategy works well and will be utilized in the subsequent development.

2.1 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 $d$ arguments $y_1, \ldots, y_d$ represent mass points. Further assume that the other arguments $y_{d+1}, \ldots, y_p$ represent points of continuity. Then, the likelihood corresponding to the distribution function in equation (1) can be expressed as
\[ f(y_1, \ldots, y_d, y_{d+1}, \ldots, y_p) = \]
\[ \sum_{i_1=0}^{1} \ldots \sum_{i_d=0}^{1} (-1)^{i_1+\cdots+i_d} C_{d+1,\ldots,p} \left( F_1(y_1^{(i_1)}), \ldots, F_d(y_d^{(i_d)}), F_{d+1}(y_{d+1}), \ldots, F_p(y_p) \right) \]
\[ \times \prod_{j=d+1}^{p} f_j(y_j), \]  

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, \( C_{d+1,\ldots,p} \) is the partial derivative of the copula with respect to the arguments in the \( d+1,\ldots,p \) 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 (\( d \) in equation (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

\[ F_{jk}(y_j, y_k) = C(\infty, \ldots, \infty, F_j(y_j), \ldots, F_k(y_k), \infty, \ldots, \infty) = C^{(jk)}(F_j(y_j), F_k(y_k)). \]

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 \( \Sigma \), then \( C^{(jk)} \) is from the same elliptical family with association
parameter $\Sigma_{jk}$, corresponding to the $j$th row and $k$th column of $\Sigma$. 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 $y^{d}_j$ and of $Y_k$ at point of continuity $y^{c}_k$. Then, the joint distribution function has a hybrid probability density/mass function of the form:

$$f_{jk}(y^{d}_j, y^{c}_k) = \partial_2 \Pr(Y_j = y^{d}_j, Y_k \leq y^{c}_k) = \partial_2 \left\{ \Pr(Y_j \leq y^{d}_j, Y_k \leq y^{c}_k) - \Pr(Y_j \leq y^{d}_j - , Y_k \leq y^{c}_k) \right\} = \partial_2 \left\{ C(F_j(y^{d}_j), F_k(y^{c}_k)) - C(F_j(y^{d}_j - ), F_k(y^{c}_k)) \right\} = \left\{ C_2(F_j(y^{d}_j), F_k(y^{c}_k)) - C_2(F_j(y^{d}_j - ), F_k(y^{c}_k)) \right\} f_k(y^{c}_k).$$

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 in our Online Supplement 3.6.

**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:

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

To illustrate, when both observations are 0, then the likelihood $F_{jk}(0,0) = C (F_j(0), F_k(0))$ requires evaluation of the distribution function $C$. With a Gaussian copula, this is a two-dimensional integral. In the same way, with $m = 5$ years of data, an observation that has no claims (all 0’s) in all years requires a five-dimensional integration. For datasets that have tens of thousands of observations, this becomes cumbersome. This is one motivation for utilizing pairwise distributions in this paper.

2.2 Pairwise Likelihood-Based Inference

When the number of discrete components is small, then full maximum likelihood estimation typically is feasible, c.f., Song et al. (2009). For copula likelihoods, the presence of a large number of discrete components means high-dimensional integration is needed to evaluate the full likelihood which can be computationally prohibitive. Instead, analysts look to (weighted sums of) low dimensional marginals that have almost as much information as a full likelihood and that are much easier to evaluate. Elliptical copulas naturally lend themselves to this approach as structure is preserved when examining smaller dimensions. As described in Joe (2014), Section 5.6, the strategy of basing inference on low-dimensional marginals is a special case of the “composite likelihood” approach.

As one special case, we have already mentioned the IFM technique. As another, consider
a “pairwise” (logarithmic) likelihood that has the form
\[
L = \sum_{i=1}^{n} \left\{ \sum_{j=1}^{p-1} \sum_{k=j+1}^{p} \ln f_{ijk}(y_{ij}, y_{ik}) \right\}.
\]

This assumes independence among policyholders but permits dependence among risks. The pairwise likelihood procedure works well and has been used successfully in connection with copulas, c.f., Shi (2014) and Jiryaie et al. (2016).

We can express pairwise likelihood procedure in terms of an estimating, or inference function, equation that will allow us to see why pairwise likelihood works. Specifically, assume that the parameters of the marginal distribution have been fit. Now, we wish to estimate a vector of association parameters \( \theta \), a \( r \times 1 \) vector, that captures dependencies.

To estimate \( \theta \), we define the score function
\[
g_{\theta,ijk}(Y_{ij}, Y_{ik}) = \partial_{\theta} \ln f_{ijk}(Y_{ij}, Y_{ik}).
\]

This is a mean zero random vector that contains information about \( \theta \). It is an unbiased estimator (of zero) and can be used in an estimating equation. Under mild regularity conditions, c.f. Song (2007), minimizing the pairwise likelihood yields the same estimate as solving the estimating equation
\[
\sum_{i=1}^{n} \left\{ \sum_{j=1}^{p-1} \sum_{k=j+1}^{p} g_{\theta,ijk}(Y_{ij}, Y_{ik}) \right\} = 0.
\]

Denote this estimator as \( \theta_{PL} \).

### 2.3 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.

### 2.3.1 GMM Procedure

Using equation (3), we combine several scores from the *i*th risk as

\[ g_{\theta,i}(Y_{i1}, \ldots, Y_{ip}) = \begin{cases} 
  g_{\theta,12}(Y_{i1}, Y_{i2}) \\
  \vdots \\
  g_{\theta,1p}(Y_{i1}, Y_{ip}) \\
  \vdots \\
  g_{\theta,i,p-1,p}(Y_{i,p-1}, Y_{ip}) 
\end{cases} \]  

(4)
a column vector with \( r \binom{p}{2} \) rows. The sum of these statistics for a sample of size \( n \) is \( g_{\theta} = \sum_{i=1}^{n} g_{\theta,i}(Y_{i1}, \ldots, Y_{ip}) \). Although \( g_{\theta} \) is a mean zero vector containing information about \( \theta \), the number of elements in \( g_{\theta} \) exceeds the number of parameters and so we use *GMM* to estimate the parameters. Specifically, the *GMM* estimator of \( \theta \), say \( \hat{\theta}_{GMM} \), is the minimizer of the expression \( g_{\theta} (\text{Var} \ g_{\hat{\theta}})^{-1} g_{\theta}' \). To implement this, we use the plug-in estimator of the variance

\[ \hat{\text{Var}} \ g_{\hat{\theta}} = \frac{1}{n} \sum_{i=1}^{n} g_{\hat{\theta},i}(Y_{i1}, \ldots, Y_{ip})' g_{\hat{\theta},i}(Y_{i1}, \ldots, Y_{ip}) \]  

(5)
This plug-in estimator requires \( \hat{\theta} \), a consistent estimator of \( \theta \). We use \( \theta_{PL} \), defined in Section 2.2.

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

### 2.3.2 Evaluation of GMM Scores

We now evaluate the scores. For discrete \( y^d_j \) and continuous \( y^c_k \) outcomes, we have

\[
g_{\theta,ijk}(y^d_j, y^c_k) = \partial_\theta \ln \left\{ C_2 \left( F_j(y^d_j), F_k(y^c_k) \right) - C_2 \left( F_j(y^d_j), F_k(y^c_k) \right) \right\} f_k(y^c_k)
\]

\[
= \frac{\partial_\theta \left\{ C_2 \left( F_j(y^d_j), F_k(y^c_k) \right) - C_2 \left( F_j(y^d_j), F_k(y^c_k) \right) \right\}}{C_2 \left( F_j(y^d_j), F_k(y^c_k) \right) - C_2 \left( F_j(y^d_j), F_k(y^c_k) \right)}
\]

For two discrete outcomes, the score can be expressed as

\[
g_{\theta,ijk}(y^d_j, y^d_k) = \partial_\theta \ln f_{ijk}(y^d_j, y^d_k)
\]

\[
= \frac{\partial_\theta \left\{ C_j(y^d_j, F_k(y^d_k)) - C_j(F_j(y^d_j), F_k(y^d_k)) - C_j(F_j(y^d_j), F_k(y^d_k)) - C_j(F_j(y^d_j), F_k(y^d_k)) \right\}}{
C_j(F_j(y^d_j), F_k(y^d_k)) - C_j(F_j(y^d_j), F_k(y^d_k)) - C_j(F_j(y^d_j), F_k(y^d_k)) - C_j(F_j(y^d_j), F_k(y^d_k)) \right\}}
\]

For two continuous outcomes, \( y^c_j \) and \( y^c_k \), we have

\[
g_{\theta,ijk}(y^c_j, y^c_k) = \partial_\theta \ln \left\{ c(F_{Y_j}(y^c_j), F_{Y_k}(y^c_k)) f_{ij}(y^c_j) f_{ik}(y^c_k) \right\}
\]

\[
= \frac{\partial_\theta c(F_{Y_j}(y^c_j), F_{Y_k}(y^c_k))}{c(F_{Y_j}(y^c_j), F_{Y_k}(y^c_k))}
\]

In the Appendix and Online Supplement 2, we give explicit formulas for Gaussian copula derivatives. An advantage of restricting ourselves to pairwise distributions is that most of the functions are available from Schepsmeier and Stöber (2012, 2014). We use an additional
relationship, from Plackett (1954), that is developed in the Appendix,

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

Here, $\phi_2$ is a bivariate normal probability density function and $z_j = \Phi^{-1}(u_j), j = 1, 2$ are the normal scores corresponding to residuals $u_j$.

### 2.4 Comparing Pairwise Likelihood to GMM Estimators

This section compares the efficiency of the pairwise likelihood estimator $\theta_{PL}$ to the GMM estimator $\theta_{GMM}$ 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 $\theta_{GMM}$ is

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

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

$$g_{PL}(Y_i; \theta) = \sum_{j=1}^{p-1} \sum_{k=j+1}^{p} g_{\theta,ijk}(Y_{ij}, Y_{ik}) = (I_r \otimes 1') g_{\theta,i}(Y_i) = B g_{\theta,i}(Y_i),$$

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

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

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 summarized in Online Supplement 3. 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.

3 Lapse Modeling

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.

3.1 Joint Model Specification

Consider the case where we follow 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 \( Y_{it} = \{Y_{1,it}, \ldots, Y_{p,it}\} \). Associated with each policyholder is a set of (possibly time varying) rating variables \( x_{it} \) for the \( i \)th policyholder at time \( t \).

Throughout, we assume independence among policyholders.

To define the joint distribution among outcomes, we use copulas that are based on transformed outcomes. Specifically, for each \( ijt \), define \( \nu_{ijt} = \Phi^{-1} \left( F_{Y_{ijt}}(Y_{ijt}) \right) \) to be the trans-
formed outcome. In the same way, define \( \nu_{iLt} = \Phi^{-1}(F_{iLt}(L_{iLt})) \). 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 straightforward to show that a copula exists even when outcomes are not continuous.

3.1.1 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 4, 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 and Box (1981)). In this framework, there are sequences of latent variables \( \{\eta_{ijt}\} \) that are iid. Transformed outcomes can be expressed as \( \nu_{ijt} = \eta_{ijt} + \psi_{j,1}\eta_{ij,t-1} + \cdots + \psi_{j,t-1}\eta_{ij,1} \). For example, we might use a moving average of order one (MA1), \( \nu_{ijt} = \eta_{ijt} + \psi_{j,1}\eta_{ij,t-1} \), or an autoregressive of order one (AR1) representation, \( \nu_{ijt} = \eta_{ijt} + \psi_{j,1}\nu_{ij,t-1} + \cdots + \psi_{j,t-1}\eta_{ij,1} \).

Using matrix notation, we define \( \nu_{ij} = (\nu_{ij1}, \ldots, \nu_{ijm})' \), \( \eta_{ij} = (\eta_{ij1}, \ldots, \eta_{ijm})' \), and

\[
\Psi_j = \begin{pmatrix}
1 & 0 & \cdots & 0 \\
\psi_{j,1} & 1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
\psi_{j,m-1} & \psi_{j,m-2} & \cdots & 1
\end{pmatrix}.
\]

Thus, \( \nu_{ij} = \Psi_j \eta_{ij} \) and \( \text{Var } \nu_{ij} = \Psi_j \Psi_j' \), for \( j = 1, \ldots, p \). In the same way, for lapse we have
\( \nu_{iL} = \Psi_L \eta_{iL} \) and \( \text{Var} \nu_{iL} = \Psi_L \Psi'_L \).

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

\[
\text{Cov}(\eta_{ij}, \eta_{ikt}) = \begin{cases} 
\rho_{jk} & s = t \\
0 & s \neq t 
\end{cases}
\text{ and } \text{Cov}(\eta_{ij}, \eta_{iL}) = \begin{cases} \rho_{jL} & s = t \\
0 & s \neq t \end{cases}.
\]

This yields \( \text{Cov}(\nu_{ij}, \nu_{ik}) = \rho_{jk} \Psi_j \Psi'_k \) and \( \text{Cov}(\nu_{iL}, \nu_{ij}) = \rho_{jL} \Psi_L \Psi'_j \). We summarize this using the association matrix

\[
\Sigma = \begin{pmatrix}
\Psi_L \Psi'_L & \rho_{L1} \Psi_L \Psi'_1 & \ldots & \rho_{Lp} \Psi_L \Psi'_p \\
\rho_{L1} \Psi'_1 & \Psi'_1 \Psi'_1 & \ldots & \rho_{1p} \Psi'_1 \Psi'_p \\
\vdots & \vdots & \ddots & \vdots \\
\rho_{Lp} \Psi'_p & \rho_{1p} \Psi'_p \Psi'_1 & \ldots & \Psi'_p \Psi'_p \\
\end{pmatrix}.
\quad (6)
\]

To develop intuition, in Section 4 we refer to the case where \( \Psi_j, j = 1, \ldots, p \), and \( \Psi_L \) have elements that are zero off the diagonal, the case of no temporal dependence.

### 3.1.2 Joint Model

To use the equation (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 \( i \)th policy-
holder lapses. Specifically,

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

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

\[
\Pr(T_i = t) = \Pr(L_{i1} = 0, \ldots, L_{i,t-1} = 0, L_{it} = 1) = \begin{cases} 
C(F_{Li,1}(0), \ldots, F_{Li,m}(0)) & t = m + 1 \\
C(F_{Li,1}(0), \ldots, F_{Li,t-1}(0), 1, \ldots, 1) - C(F_{Li,1}(0), \ldots, F_{Li,t}(0), 1, \ldots, 1) & 1 \leq t \leq m \\
C(a_{im}(F_{Li,m}(0))) & t = m + 1 \\
C(a_{it}(1)) - C(a_{it}(F_{Li,t}(0))) & 1 \leq t \leq m.
\end{cases}
\] (7)

The last equality uses some notation \(a_{it}(a) = (F_{Li,1}(0), \ldots, F_{Li,t-1}(0), a, 1, \ldots, 1)\) 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

\[16\]
based on

\[ \Pr(T_i = m + 1, Y_{i1} \leq y_1, \ldots, Y_{im} \leq y_m) \]

\[ = \Pr(L_{i1} = 0, \ldots, L_{i,m} = 0, Y_{i1} \leq y_1, \ldots, Y_{im} \leq y_m). \] (8)

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

\[ \Pr(T_i = t, Y_{i1} \leq y_1, \ldots, Y_{it} \leq y_t) \]

\[ = \Pr(L_{i1} = 0, \ldots, L_{i,t-1} = 0, L_{i,t} = 1, L_{i,t+1} \leq \infty, \ldots, L_{im} \leq \infty, \]

\[ Y_{i1} \leq y_1, \ldots, Y_{it} \leq y_t, Y_{i,t+1} \leq \infty, \ldots, Y_{im} \leq \infty). \] (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.

### 3.2 Conditional Likelihood

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

\[ \Pr(Y_{j,is} \leq y_1, Y_{k,is} \leq y_2 | T_i = t) = \frac{\Pr(T_i = t, Y_{j,is} \leq y_1, Y_{k,is} \leq y_2)}{\Pr(T_i = t)}. \] (10)

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

$$\Pr(T_i = t, Y_{j, is} \leq y_1, Y_{k, is} \leq y_2) = \begin{cases} 
C(a_{mi}(F_{Li,m}(0)), F_{j, is}(y_1), F_{k, is}(y_2)) & s < t \leq m + 1 \\
C(a_{it}(1), F_{j, is}(y_1), F_{k, is}(y_2)) - C(a_{it}(F_{Li,t}(0)), F_{j, is}(y_1), F_{k, is}(y_2)) & s = t \leq m 
\end{cases}.$$  

(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 equation (6) to express the association matrix for \{L_{i1}, \ldots, L_{im}, Y_{j, is}, Y_{k, is}\} as

$$\Sigma = \begin{pmatrix} 
\Psi_L \Psi'_L & \rho_{Lj} \Psi'_L \Psi_{j, is} & \rho_{Lk} \Psi'_L \Psi_{k, is} \\
\rho_{Lj} \Psi_{j, is} \Psi'_L & 1 & \rho_{jk} \Psi_{j, is} \Psi_{k, is} \\
\rho_{Lk} \Psi_{k, is} \Psi'_L & \rho_{jk} \Psi_{j, is} \Psi'_L & 1 
\end{pmatrix},$$  

(12)

where $\Psi_{j, is}$ is the $j$th row of $\Psi_j$ and similarly for $\Psi_k$.

The corresponding conditional hybrid probability density/mass functions follow as in Section 2.1. For $s < t \leq m + 1$, this is

$$f_{ijk,s|m+1}(y_1, y_2) = \frac{1}{\Pr(T_i=m+1)} \times \sum_{i_2=0}^{1} \sum_{i_1=0}^{1} (-1)^{i_1+i_2} C\left(a_{im}(F_{Li,m}(0)), F_{j, is}(y_1^{(i_1)}), F_{k, is}(y_2^{(i_2)})\right) f_{j, is}(y_1) f_{k, is}(y_2)$$

$$= \begin{cases} 
\sum_{i_2=0}^{1} \sum_{i_1=0}^{1} (-1)^{i_1+i_2} C_{m+1} \left(a_{im}(F_{Li,m}(0)), F_{j, is}(y_1^{(i_1)}), F_{k, is}(y_2^{(i_2)})\right) & y_1 = y_1^d, y_2 = y_2^d \\
\sum_{i_2=0}^{1} \sum_{i_1=0}^{1} (-1)^{i_1+i_2} C_{m+2} \left(a_{im}(F_{Li,m}(0)), F_{j, is}(y_1^{(i_1)}), F_{k, is}(y_2^{(i_2)})\right) & y_1 = y_1^d, y_2 = y_2^d \\
C_{m+1,m+2} \left(a_{im}(F_{Li,m}(0)), F_{j, is}(y_1), F_{k, is}(y_2)\right) & y_1 = y_1^d, y_2 = y_2^d 
\end{cases}.$$  

(13)

Here, we use the notation introduced in equation (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 $F_{j, is}$. 

18
The corresponding conditional hybrid probability density/mass functions for \( s = t \leq m \), follows in the same way

\[
f_{ijk,s|t}(y_1, y_2) = \frac{1}{Pr(T_i = t)} \times 
\begin{align*}
&\sum_{i_r=0}^{1} \sum_{i_1=0}^{1} \sum_{i_2=0}^{1} (-1)^{i_r+i_1+i_2} \left\{ C \left( a_{it}(F_{Li,t}(0)^{i_r}), F_{j,is}(y_1^{(i_1)}), F_{k,is}(y_2^{(i_2)}) \right) \right\} \\
&\sum_{i_r=0}^{1} \sum_{i_1=0}^{1} \sum_{i_2=0}^{1} (-1)^{i_r+i_1+i_2} \left\{ C_{m+1} \left( a_{it}(F_{Li,t}(0)^{i_r}), F_{j,is}(y_1), F_{k,is}(y_2^{(i_2)}) \right) \right\} f_{j,is}(y_1) \\
&\sum_{i_r=0}^{1} \sum_{i_1=0}^{1} (-1)^{i_r+i_1+i_2} \left\{ C_{m+2} \left( a_{it}(F_{Li,t}(0)^{i_r}), F_{j,is}(y_1^{(i_1)}), F_{k,is}(y_2) \right) \right\} f_{k,is}(y_2) \\
&\sum_{i_r=0}^{1} (-1)^{i_r} C_{m+1,m+2} \left( a_{it}(F_{Li,t}(0)^{i_r}), F_{j,is}(y_1), F_{k,is}(y_2) \right) \left\{ f_{j,is}(y_1) f_{k,is}(y_2) \right\}
\end{align*}
\]

\( (14) \)

### 3.3 Lapse 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 \( \theta \) 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 equations (13) and (14).

To estimate \( \theta \), for \( s \leq t \), define

\[
g_{\theta,i,s,t}(y_1, y_2, T) = I(T = t) \partial_\theta \ln f_{ijk,s|t}(y_1, y_2). \quad (15)
\]

This is a mean zero random variable that contains information about \( \theta \). To see that it has
mean zero,

\[
E g_{\theta, i, s, t}(Y_{j, i, s, T_i}) = E [E (g_{\theta, i, s, t}(Y_{j, i, s, Y_{k, i, s}, T_i})|T_i = t)] \\
= E [I(T_i = t) E (\partial_\theta \ln f_{i j k, s|t}(Y_{j, i, s}, Y_{k, i, s}))] \\
= E [I(T_i = t) \cdot 0] = 0,
\]

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

With the score function in equation (15), we can use the GMM procedure defined in Section 2.3.1. The only thing that remains to do is to evaluate the score functions in terms of copula-based functions.

### 3.3.1 Lapse Score Evaluation

The lapse scores are similar to the scores introduced in Section 2.3.2 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 \( s < t \leq m + 1 \) and two zero outcomes, this can be expressed as

\[
\partial_\theta \ln f_{i12, s|m+1}(0, 0) = \frac{\partial_\theta C (a_{im}(F_{Li,m}(0)), F_{1,i,s}(0), F_{2,i,s}(0))}{C (a_{im}(F_{Li,m}(0)), F_{1,i,s}(0), F_{2,i,s}(0))} - \frac{\partial_\theta C (a_{im}(F_{Li,m}(0)))}{C (a_{im}(F_{Li,m}(0)))}.
\]

For a single positive outcome, \( y > 0 \), we have

\[
\partial_\theta \ln f_{i12, s|m+1}(y, 0) = \frac{\partial_\theta C_{m+1} (a_{im}(F_{Li,m}(0)), F_{1,i,s}(y), F_{2,i,s}(0))}{C_{m+1} (a_{im}(F_{Li,m}(0)), F_{1,i,s}(y), F_{2,i,s}(0))} - \frac{\partial_\theta C (a_{im}(F_{Li,m}(0)))}{C (a_{im}(F_{Li,m}(0)))}
\]
For two positive outcomes, \( y_1 > 0 \) and \( y_2 > 0 \), we have

\[
\partial \ln f_{i12,s|m+1}(y_1, y_2) = \frac{\partial \theta C_{m+1,m+2}(a_{im}(F_{Li,m}(0)), F_{1, is}(y_1), F_{2, is}(y_2))}{C_{m+1,m+2}(a_{im}(F_{Li,m}(0)), F_{1, is}(y_1), F_{2, is}(y_2))} - \frac{\partial \theta C(a_{im}(F_{Li,m}(0)))}{C(a_{im}(F_{Li,m}(0)))}.
\]

Scores for \( s \leq t = m \) follow in a similar way. In Online Supplement 2, we give explicit formulas for Gaussian copula derivatives.

## 4 Lapse Modeling with Temporal Independence

In this section, we assume no temporal dependence for both the lapse and claims 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 lapse and claims outcomes. This specification permits, for example, large claims to influence the tendency to lapse a policy or a latent variable to simultaneously influence both lapse and claims outcome.

### 4.1 Joint Model Specification

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

\[
\Pr(L_{i1} \leq r_1, \ldots, L_{im} \leq r_m, Y_{i1} \leq y_1, \ldots, Y_{im} \leq y_m)
= \prod_{t=1}^{m} C(F_{Lit}(r_t), F_{1, it}(y_{1t}), F_{2, it}(y_{2t})).
\]

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

\[
\Pr(T_i = t) = \begin{cases} 
\prod_{s=1}^m F_{Li,s}(0) & t = m + 1 \\
(1 - F_{L1,t}(0)) \prod_{s=1}^{t-1} F_{Li,s}(0) & 1 \leq t \leq m 
\end{cases}
\]

From this, we will be able to use lapse outcomes to estimate the marginal lapse outcomes 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

\[
\Pr(T_i = m + 1, Y_{i1} \leq y_1, \ldots, Y_{im} \leq y_m) = \prod_{s=1}^m C (F_{Li,s}(0), F_{1,is}(y_{1s}), F_{2,is}(y_{2s})) .
\]

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

\[
\Pr(T_i = t, Y_{i1} \leq y_1, \ldots, Y_{it} \leq y_t)
= \{ C (1, F_{1,it}(y_{1t}), F_{2,it}(y_{2t})) - C (F_{L1,t}(0), F_{1,it}(y_{1t}), F_{2,it}(y_{2t})) \}
\prod_{s=1}^{t-1} C (F_{Li,s}(0), F_{1,is}(y_{1s}), F_{2,is}(y_{2s})) .
\]
Thus, the corresponding conditional distribution function is

\[
\Pr (Y_{1,i,s} \leq y_1, Y_{2,i,s} \leq y_2 | T_i = t) =
\begin{cases}
\frac{C(F_{L,i,s}(0), F_{1,i,s}(y_1), F_{2,i,s}(y_2))}{F_{L,i,s}(0)} & \text{for } s < t \leq m + 1 \\
\frac{C(1, F_{1,i,s}(y_1), F_{2,i,s}(y_2)) - C(F_{L,i,s}(0), F_{1,i,s}(y_1), F_{2,i,s}(y_2))}{1 - F_{L,i,s}(0)} & \text{for } s = t \leq m
\end{cases}
\] (16)

This is more intuitive than the general expressions given in Section 3.2. 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.

### 4.2 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 \(t\)th 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 \(\text{Like}^* = f(Y, L | X, \theta)\) where \(X\) are covariates and \(\theta\) are parameters. By conditioning, write this as \(\text{Like}^* = \prod_{s=1}^{m} f(Y_s, L_s | X, \theta, H_s)\) with history \(H_s = \{Y_1, \ldots, Y_{s-1}, L_1, \ldots, L_{s-1}\}\). On the
event \( \{ T = t \} = \{ L_1 = 1, \ldots, L_{t-1} = 0, L_t = 1, \ldots, L_m = 1 \} \), we have

\[
\text{Like}^* = \left( \prod_{s=1}^{t-1} f(Y_s, L_s = 0|X, \theta, H_s) \right) f(Y_t, L_t = 1|X, \theta, H_s) \left( \prod_{s=t+1}^{m} f(Y_s, L_s = 1|X, \theta, H_s) \right)
\]

\[
= \left( \prod_{s=1}^{t-1} f(Y_{\text{obs},s}, L_s = 0|X, \theta, H_s) \right) f(Y_{\text{obs},t}, Y_{\text{mis},t}, L_t = 1|X, \theta, H_s)
\]

\[
\left( \prod_{s=t+1}^{m} f(Y_{\text{mis},s}, L_s = 1|X, \theta, H_s) \right),
\]

where \( Y_{\text{obs}}, Y_{\text{mis}} \) 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

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

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

Thus, the integrated likelihood is

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

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

### 4.3 Comparing Trivariate Likelihood to GMM Estimators

Similar to Section 2.4, 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 takes on values 1 or 2 depending on 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 \( \mu_{j, it} = \exp(x'_{it}\beta_j) \), \( j = 1, 2 \). For the lapse variable, the expected value is \( \pi_{it} = \exp(x'_{it}\beta_L) / (1 + \exp(x'_{it}\beta_L)) \), 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, \( \phi_j \) (for dispersion) and \( P_j \) (the “power” parameter). Recall, for a Tweedie distribution, that the variance is \( \phi_j\mu^P \) with \( P = 1.67 \) in this study.

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

\[
\Sigma = \begin{pmatrix}
1 & \rho_{LA} & \rho_{LH} \\
\rho_{LA} & 1 & \rho_{AH} \\
\rho_{LH} & \rho_{AH} & 1
\end{pmatrix}.
\]

For example, we might use positive values for the association between lapse and claims (\( \rho_{LA} \) and \( \rho_{LH} \)) so that large claims are associated with higher lapse. We might use a positive value for the association between claim types (\( \rho_{AH} \)).

This section compares the efficiency of the pairwise likelihood estimator \( \theta_{PL} \) to the GMM estimator \( \theta_{GMM} \).

Table 1 summarizes the performance of the GMM estimators lapse estimator by varying the sample size and dispersion parameters, \( \phi_1 \) (for auto), and \( \phi_2 \) (for home). For this table, we used \( \rho_{LA} = -0.2, \rho_{LH} = 0.2, \) and \( \rho_{AH} = 0.1 \) 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 our Table 1 which compares pairwise likelihood and GMM estimators study (without lapse). As the dispersion parameters \( \phi \) 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 \( \rho_{LA} = -0.2 \), 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 primarily interested in point estimates but also want to say something about statistical significance.

We use the pairwise estimators as starting values in the more complete GMM estimators. As documented in Online Supplement 4, the GMM estimators are marginally more efficient than the pairwise estimators. Moreover, we investigated the performance of the GMM estimators by varying the association parameters and learned that their performance was robust to the choice of different sets of association parameters.
Table 1: Summary of the GMM Lapse Estimators

| Num | n  | φ₁ = φ₂ | Bias LA | Bias LH | Bias AH | Standard Error LA | Standard Error LH | Standard Error AH |
|-----|----|---------|---------|---------|---------|-------------------|-------------------|-------------------|
| 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            |

5 Empirical Application

Introduction and Variable Descriptions

This paper examines longitudinal data from a major Spanish insurance company that offers automobile and homeowners insurance. As in many countries, in Spain vehicle owners 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 dataset tracks 40,284 clients over five years, between 2010 and 2015, who subscribed to both automobile and homeowners insurance. From the unbalanced panel of policyholders over 5 years, there are \( N = 122,935 \) observations in the data set. Table 2 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 home-
Table 2: 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 non-renewals for at least one policy (%) | 26% | 25% | 24% | 22% | 17% |

owners, and (d) renewal information such as the date of renewal. Table 3 provide variable descriptions.

Table 3: 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 | 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                                 |

Table 4 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. Readers interested in details of this dataset may look to the Online Supplement 5 which gives many more summary statistics.
Marginal Model Fits

After extensive diagnostic testing described in the *Online Supplement* 5, 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 5 summarizes estimates of marginal model fits.

### Table 5: 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    | 0.348 4.25    |
| Insuredcapital_continent_reappartment | |                | 1.097 7.16 |
| Policy_PaymentMethodH | | -0.765 -3.40 |               |

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 6.
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.

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

|         | 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    |
models that balance demand and supply side considerations of personal insurance lines.

6 Summary and 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 is the relative efficiency. For example, we have seen in Section 2 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 3-5 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 3 provides a detailed theory on handling general longitudinal data with explicit expressions in Section 4 for temporally independent, yet still dependent censoring, of observations. The supporting Appendices 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).

Insurance analysts employ regression techniques to address the heterogeneity of cus-
tomers. 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.

References

Brockett, Patrick L., Linda L. Golden, Montserrat Guillén, Jens Perch Nielsen, Jan Parner, and Ana Maria Perez-Marín (2008). “Survival analysis of a household portfolio of insurance policies: how much time do you have to stop total customer defection?” *Journal of Risk and Insurance*, Vol. 75, pp. 713–737.

Elashoff, Robert M., Gang Li, and Ning Li (2017). *Joint Modeling of Longitudinal and Time-to-Event Data*. CRC Press.

Frees, Edward W, Richard A. Derrig, and Glenn Meyers (2014). *Predictive Modeling Applications in Actuarial Science*, Vol. 1. Cambridge University Press.

Gassmann, H. I. (2003). “Multivariate normal probabilities: implementing an old idea of Plackett’s.” *Journal of Computational and Graphical Statistics*, Vol. 12, pp. 731–752.

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

Guillén, Montserrat, Jan Parner, Chresten Densgsoe, and Ana M Perez-Marin (2003). “Using logistic regression models to predict and understand why customer leave an insurance company,” *Intelligent Techniques in the Insurance Industry: Theory and Applications. London: World Scientific*.

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

Jiryaie, F., N. Withanage, B. Wu, and A.R. De Leon (2016). “Gaussian copula distributions for mixed data, with application in discrimination,” *Journal of Statistical Computation and Simulation*, Vol. 86, pp. 1643–1659.

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

——— (2014). *Dependence Modeling with Copulas*. CRC Press.

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

Panagiotelis, Anastasios, Claudia Czado, and Harry Joe (2012). “Pair copula constructions for multivariate discrete data,” *Journal of the American Statistical Association*, Vol. 107, pp. 1063–1072.
7 Appendix. Derivatives with Respect to Association Parameters

7.1 Bivariate Gaussian Copula

For derivatives with respect to $\rho$, we start with the distribution function and note a result that, according to Plackett (1954), was well known even in the mid-1950’s,

$$\frac{\partial}{\partial \rho} \Phi_2(z_1, z_2) = \phi_2(z_1, z_2).$$

For Gaussian copulas, this immediately yields

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

For other details in the bivariate case, we use the work of Schepsmeier and Stöber (2012, 2014). For example, they calculate the partial derivative

$$\frac{\partial C_2(u_1, u_2)}{\partial \rho} = \phi \left( \frac{\Phi^{-1}(u_1) - \rho \Phi^{-1}(u_2)}{\sqrt{1 - \rho^2}} \right) \cdot \frac{\rho \Phi^{-1}(u_1) - \Phi^{-1}(u_2)}{(1 - \rho^2)^{3/2}}.$$
7.2 Classic Multivariate Result

In the following, we sometimes require the correlation matrix associated with $\Sigma$, defined as

$$R = R(\Sigma) = (\text{diag}(\Sigma))^{-1/2} \Sigma (\text{diag}(\Sigma))^{-1/2}.$$  

Further, following standard notation, the subscripts in $R_{ij}$ refer to the $i$th row and $j$th column of the matrix $R$.

For a general multivariate approach, we cite a result due to Plackett (1954). Partition the correlation matrix as

$$R = \left( \begin{array}{cc} R_{1:2,1:2} & R_{1:2,3:d} \\ R'_{1:2,3:d} & R_{3:d,3:d} \end{array} \right), \quad R_{1:2,1:2} = \left( \begin{array}{cc} 1 & \rho \\ \rho & 1 \end{array} \right),$$

so that $R_{1:2,1:2}$ is the submatrix for the first two elements and $\rho$ is the corresponding correlation coefficient. Then, from Plackett (1954), we have

$$h^{(1:2)}_{2,d}(x_1, \ldots, x_d; \Sigma) = \frac{\partial}{\partial \rho} \Phi_d(x_1, \ldots, x_d; R) \quad \text{(17)}$$

where

$$x^* = \begin{pmatrix} x_3 \\ \vdots \\ x_d \end{pmatrix} - R'_{1:2,3:d} R^{-1}_{1:2,1:2} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \quad \text{and} \quad R_{\{3:d,3:d\}-\{1:2\}} = R_{3:d,3:d} - R'_{1:2,3:d} R^{-1}_{1:2,1:2} R_{1:2,3:d}.$$  

In the same way, define $h^{(ij)}_{2,d}$ by interchanging the first and second random variables with the $(i, j)$ pair of random variables.

See also Gassmann (2003) for a more recent commentary on this interesting result.

7.3 Association Parameter Derivatives

7.3.1 Association Parameter Derivatives of $C_d$

We now seek to compute the partial derivative with respect association parameter of the partial copula function (the partial copula function is reviewed in Online Supplement 2). This a function of $d - 1$ arguments of the form $z_j - \mu_{1:2,j}$ and an additional $\binom{d}{2}$ arguments in $R$. For the first set of arguments, we use the expression

$$h^{(1)}_{1,d}(x_1, \ldots, x_d; R) = \frac{\partial}{\partial x_1} \Phi_d(x_1, \ldots, x_d; R) \quad \text{(18)}$$

where

$$x^*_1 = x^*_1(R) = R_{2:d,1} x_1 \quad \text{and} \quad R_{2:1} = R_{2:d,2:d} - R_{2:d,1} R'_{2:d,1}.$$  

We define $h^{(j)}_{1,d}$ in the same way by interchanging the first and the $j$th random variables.
Here, we seek to evaluate the partial copula function with respect to a generic association parameter $\rho$. With this notation, we have

$$\frac{\partial}{\partial x_1} \Pr(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_d \leq x_d) = f_{X_1}(x_1) \Pr(X_2 \leq x_2 \ldots, X_d \leq x_d | X_1 = x_1).$$

The second is that, for $X_1, \ldots, X_d$ is normal with mean zero and variance-covariance matrix $\mathbf{R}$, we have $X_2, \ldots, X_d$ given $X_1$ is normal with mean $x_1^*$ and variance-covariance $\mathbf{R}_{2:1}$.

Define the rescaled arguments $z_j^* = (z_j - \mu_{\{1, \ldots, d-1\},d,j}) \Sigma_{\{1, \ldots, d-1\},d,ij}^{-1/2}$. With this and equation (17), the partial derivative of the partial copula function with respect to a generic association parameter $\rho$ is

$$\frac{\partial}{\partial \rho} C_d(u_1, \ldots, u_d) = \frac{\partial}{\partial \rho} \Phi_{d-1} (z_1 - \mu_{\{1, \ldots, d-1\},d,1}, \ldots, z_{d-1} - \mu_{\{1, \ldots, d-1\},d,d-1}; \Sigma_{\{1, \ldots, d-1\},d}) = \frac{\partial}{\partial \rho} \Phi_{d-1} (z_1^*, \ldots, z_{d-1}^*; \mathbf{R}_{\{1, \ldots, d-1\},d})$$

$$= \sum_{k=1}^{d-1} h_{1,d-1}^{(k)} (z_1^*, \ldots, z_{d-1}^*; \mathbf{R}_{\{1, \ldots, d-1\},d}) \frac{\partial}{\partial \rho} z_k^* + \sum_{i<j} h_{2,d-1}^{(ij)} (z_1^*, \ldots, z_{d-1}^*; \mathbf{R}_{\{1, \ldots, d-1\},d}) \frac{\partial}{\partial \rho} \mathbf{R}_{\{1, \ldots, d-1\},d,ij}$$

(19)

Trivariate Case: Association Parameter Derivatives of $C_3$

To illustrate, consider the trivariate case (more details available in Online Supplement 2.1). Here, we seek to evaluate $\frac{\partial}{\partial \rho} \Phi_2 (z_1 - \mu_{12,3,1}, z_2 - \mu_{12,3,2}; \Sigma_{12,3})$.

With the matrix $\Sigma_{12,3}$, to simplify notation, we define

$$\Sigma_{12,3,11} = 1 - \rho_{13}^2 = \sigma_1^2$$

$$\Sigma_{12,3,22} = 1 - \rho_{23}^2 = \sigma_2^2$$

$$\Sigma_{12,3,12} = \rho_{12} - \rho_{13} \rho_{23} = \rho_X \sigma_1 \sigma_2.$$

With this notation, we have

$$\Phi_2 (z_1 - \mu_{12,3,1}, z_2 - \mu_{12,3,2}; \Sigma_{12,3}) = \Phi_2 \left( \frac{z_1 - \mu_{12,3,1}}{\sigma_1}, \frac{z_2 - \mu_{12,3,2}}{\sigma_2}; \rho_x \right) = \Phi_2 (z_1^*, z_2^*; \rho_x)$$

where $z_1^* = (z_1 - \mu_{12,3,1}) / \sigma_1$ and similarly for $z_2^*$.

From equation (18), we have

$$h_{1,2}^{(1)}(x_1, x_2; \mathbf{R}) = \phi(x_1) \Phi \left( \frac{x_2 - \mathbf{R}_{12} x_1}{\sqrt{1 - \mathbf{R}_{12}^2}} \right),$$

With this, we have

$$\frac{\partial}{\partial \rho} \Phi_2 (z_1 - \mu_{12,3,1}, z_2 - \mu_{12,3,2}; \Sigma_{12,3}) = \frac{\partial}{\partial \rho} \Phi_2 (z_1^*, z_2^*; \rho_x)$$

$$= h_{1,2}^{(1)}(z_1^*, z_2^*; \rho_x) \frac{\partial}{\partial \rho} z_1^* + h_{1,2}^{(1)}(z_2^*, z_1^*; \rho_x) \frac{\partial}{\partial \rho} z_2^* + \phi_2 (z_1^*, z_2^*; \rho_x) \frac{\partial}{\partial \rho} \rho_x,$$
We also need
\[
\frac{\partial}{\partial \rho^*} z^*_1 = \frac{\partial}{\partial \rho} \left( \frac{z_1 - \mu_{12, 3, 1}}{\sigma_1} \right) = \frac{\partial}{\partial \rho} \left( \frac{z_1 - z_3 \rho_{13}}{\sqrt{1 - \rho_{13}^2}} \right) = \begin{cases} 
0 & \rho = \rho_{12} \\
\frac{z_{13} - z_3}{(1 - \rho_{13}^2)^{3/2}} & \rho = \rho_{13} \\
0 & \rho = \rho_{23}
\end{cases}
\]

In the same way
\[
\frac{\partial}{\partial \rho^*} z^*_2 = \frac{\partial}{\partial \rho} \left( \frac{z_2 - \mu_{12, 3, 2}}{\sigma_2} \right) = \frac{\partial}{\partial \rho} \left( \frac{z_2 - z_3 \rho_{23}}{\sqrt{1 - \rho_{23}^2}} \right) = \begin{cases} 
0 & \rho = \rho_{12} \\
\frac{z_{23} - z_3}{(1 - \rho_{23}^2)^{3/2}} & \rho = \rho_{13} \\
0 & \rho = \rho_{23}
\end{cases}
\]

Similarly,
\[
\frac{\partial}{\partial \rho} \rho_x = \frac{\partial}{\partial \rho} \left( \frac{\rho_{12} - \rho_{13} \rho_{23}}{\sigma_1 \sigma_2} \right) = \frac{\partial}{\partial \rho} \left( \frac{\rho_{12} - \rho_{13} \rho_{23}}{(1 - \rho_{13}^2)^{1/2}(1 - \rho_{23}^2)^{1/2}} \right) = \begin{cases} 
\frac{1}{(1 - \rho_{13}^2)^{1/2}(1 - \rho_{23}^2)^{1/2}} & \rho = \rho_{12} \\
\rho_{12} \rho_{13} \rho_{23} & \rho = \rho_{13} \\
\rho_{12} \rho_{23} \rho_{13} & \rho = \rho_{23}
\end{cases}
\]

This provides needed details to compute \( \frac{\partial}{\partial \rho} C_3 \).

### 7.3.2 Association Parameter Derivatives of \( C_{d-1, d} \)

We next derive the partial derivative with respect association parameter of the copula function \( C_{d-1, d}(u_1, \ldots, u_d) \). We start with
\[
\frac{\partial}{\partial \rho} C_{d-1, d}(u_1, \ldots, u_d) = C \left( u_1, \ldots, u_{d-2}, u_{d-1}, u_d \right) \frac{\partial}{\partial \rho} c(u_{d-1}, u_d) \\
+ \left( \frac{\partial}{\partial \rho} C \left( u_1, \ldots, u_{d-2}, u_{d-1}, u_d \right) \right) c(u_{d-1}, u_d),
\]
where
\[
C \left( u_1, \ldots, u_{d-2}, u_{d-1}, u_d \right) = \Phi_{d-2} \left( z_1 - \mu_{1, \ldots, d-2}, \{d-1, d\}, \ldots, z_{d-2} - \mu_{1, \ldots, d-2}, \{d-1, d\}; \Sigma_{\{1, \ldots, d-2\}, \{d-1, d\}} \right).
\]

Further, as in equation (19)
\[
\frac{\partial}{\partial \rho} \Phi_{d-2} \left( z_1 - \mu_{1, \ldots, d-2}, \{d-1, d\}, \ldots, z_{d-2} - \mu_{1, \ldots, d-2}, \{d-1, d\}; \Sigma_{\{1, \ldots, d-2\}, \{d-1, d\}} \right) \\
= \frac{\partial}{\partial \rho} \Phi_{d-2} \left( z_1^*, \ldots, z_{d-2}^*; \{1, \ldots, d-2\}, \{d-1, d\} \right) \\
= \sum_{k=1}^{d-2} h_{1, d-2}^{(k)} \left( z_1^*, \ldots, z_{d-2}^*; \{1, \ldots, d-2\}, \{d-1, d\} \right) \frac{\partial}{\partial \rho} z_k^* \\
+ \sum_{i<j}^{d-2} h_{2, d-2}^{(ij)} \left( z_1^*, \ldots, z_{d-2}^*; \{1, \ldots, d-2\}, \{d-1, d\} \right) \frac{\partial}{\partial \rho} R_{\{1, \ldots, d-2\}, \{d-1, d\}, i, j}.
\]

Here, the rescaled arguments are defined as \( z_j^* = \left( z_j - \mu_{1, \ldots, d-2}, \{d-1, d\} \right) \Sigma_{\{1, \ldots, d-2\}, \{d-1, d\}}^{-1} \).

Derivatives of the bivariate density \( c(u_{d-1}, u_d) \) are in Section 7.1.
Trivariate Case: Association Parameter Derivatives of $C_{23}$

To illustrate, for $d = 3$, we have

$$
\frac{\partial}{\partial \rho} C_{23}(u_1, u_2, u_3) = C(u_1 | u_2, u_3) \frac{\partial}{\partial \rho} c(u_2, u_3) + \left( \frac{\partial}{\partial \rho} C(u_1 | u_2, u_3) \right) c(u_2, u_3)
$$

where $C(u_1 | u_2, u_3) = \Phi(z_1 - \mu_{1:23}; \sigma_{1:23})$. Further,

$$
\frac{\partial}{\partial \rho} C(u_1 | u_2, u_3) = \frac{\partial}{\partial \rho} \Phi(z_1 - \mu_{1:23}; \sigma_{1:23})
$$

$$
= \phi \left( \frac{z_1 - \mu_{1:23}}{\sqrt{\sigma_{1:23}}} \right) \frac{\partial}{\partial \rho} \left( \frac{z_1 - \mu_{1:23}}{\sqrt{\sigma_{1:23}}} \right)
$$

$$
= -\phi \left( \frac{z_1 - \mu_{1:23}}{\sqrt{\sigma_{1:23}}} \right) \frac{1}{\sigma_{1:23}} \left( \sigma_{1:23} \frac{\partial}{\partial \rho} \mu_{1:23} + \frac{1}{2}(z_1 - \mu_{1:23}) \frac{\partial}{\partial \rho} \sigma_{1:23} \right).
$$

Recall $\frac{\partial}{\partial \rho} \Sigma^{-1} = -\Sigma^{-1} \left( \frac{\partial}{\partial \rho} \Sigma \right) \Sigma^{-1}$. Now, with $\mu_{1:23} = \left( \begin{array}{c} \rho_{12} \\ \rho_{13} \end{array} \right)$ and $\Sigma^{-1} = \left( \begin{array}{cc} 1 & \rho_{23} \\ \rho_{23} & 1 \end{array} \right)^{-1} = \left( \begin{array}{cc} 1 & \rho_{23} \\ \rho_{23} & 1 \end{array} \right)$, we have

$$
\frac{\partial}{\partial \rho} \mu_{1:23} = \left\{ \begin{array}{c}
( 1 0 ) \left( \begin{array}{cc} 1 & \rho_{23} \\ \rho_{23} & 1 \end{array} \right)^{-1} \left( \begin{array}{c} 1 \\ \rho_{23} \end{array} \right) \\
( 0 1 ) \left( \begin{array}{cc} 1 & \rho_{23} \\ \rho_{23} & 1 \end{array} \right)^{-1} \left( \begin{array}{c} 1 \\ \rho_{23} \end{array} \right) \\
\left( \begin{array}{cc} \rho_{12} & \rho_{13} \end{array} \right) \left( \begin{array}{cc} 1 & \rho_{23} \\ \rho_{23} & 1 \end{array} \right)^{-1} \left( \begin{array}{c} 1 \\ \rho_{23} \end{array} \right)
\end{array} \right\} \rho = \rho_{12} \quad \rho = \rho_{13} \quad \rho = \rho_{23}
$$

Further, with $\sigma_{1:23} = 1 - \left( \begin{array}{cc} \rho_{12} & \rho_{13} \end{array} \right) \left( \begin{array}{c} 1 \\ \rho_{23} \end{array} \right)$, we have

$$
\frac{\partial}{\partial \rho} \sigma_{1:23} = \left\{ \begin{array}{cc}
-2 \left( \begin{array}{cc} 1 & 0 \end{array} \right) \left( \begin{array}{cc} 1 & \rho_{23} \\ \rho_{23} & 1 \end{array} \right)^{-1} \left( \begin{array}{c} 1 \\ \rho_{23} \end{array} \right) \\
-2 \left( \begin{array}{cc} 0 & 1 \end{array} \right) \left( \begin{array}{cc} 1 & \rho_{23} \\ \rho_{23} & 1 \end{array} \right)^{-1} \left( \begin{array}{c} 1 \\ \rho_{23} \end{array} \right) \\
\left( \begin{array}{cc} \rho_{12} & \rho_{13} \end{array} \right) \left( \begin{array}{cc} 1 & \rho_{23} \\ \rho_{23} & 1 \end{array} \right)^{-1} \left( \begin{array}{c} 1 \\ \rho_{23} \end{array} \right)
\end{array} \right\} \rho = \rho_{12} \quad \rho = \rho_{13} \quad \rho = \rho_{23}
$$

This provides needed details to compute $\frac{\partial}{\partial \rho} C_{23}$.