Abstract

We propose inference tools to analyse the ordering of concordance of random vectors. The analysis in the bivariate case relies on tests for upper and lower quadrant dominance of the true distribution by a parametric or semiparametric model, i.e. for a parametric or semiparametric model to give a probability that two variables are simultaneously small or large at least as great as it would be were they left unspecified. Tests for its generalisation in higher dimensions, namely joint lower and upper orthant dominance, are also analysed. The parametric and semiparametric setting are based on the copula representation for multivariate distribution, which allows for disentangling behaviour of margins and dependence structure. We propose two types of testing procedures for each setting. The first procedure is based on a formulation of the dominance concepts in terms of values taken by random variables, while the second procedure is based on a formulation in terms of probability levels. For each formulation a distance test and an intersection-union test for inequality constraints are developed depending on the definition of null [...]

Reference

CEBRIAN, Ana C., DENUIT, Michel, SCAILLET, Olivier. Testing for Concordance Ordering. 

DOI: 10.1017/S0515036100013933
TESTING FOR CONCORDANCE ORDERING

Ana C. Cebrián
Facultad de Ciencias
Universidad de Zaragoza
CP. Cerbuna, 12
S-Zaragoza 50009, Spain
acebrian@posta.unizar.es

Michel Denuit
Institut de Statistique
Université Catholique de Louvain
Voie du Roman Pays, 20
B-1348 Louvain-la-Neuve, Belgium
denuit@stat.ucl.ac.be

Olivier Scaillet
HEC Genève and FAME
Université de Genève
Bd Carl Vogt, 102
CH - 1211 Genève 4, Suisse
olivier.scaillet@hec.unige.ch

February 21, 2002
Abstract

We propose inference tools to analyse the ordering of concordance of random vectors. The analysis in the bivariate case relies on tests for upper and lower quadrant dominance of the true distribution by a parametric or semiparametric model, i.e. for a parametric or semiparametric model to give a probability that two variables are simultaneously small or large at least as great as it would be were they left unspecified. Tests for its generalisation in higher dimensions, namely joint lower and upper orthant dominance, are also analysed. The parametric and semiparametric setting are based on the copula representation for multivariate distribution, which allows for disentangling behaviour of margins and dependence structure. We propose two types of testing procedures for each setting. The first procedure is based on a formulation of the dominance concepts in terms of values taken by random variables, while the second procedure is based on a formulation in terms of probability levels. For each formulation a distance test and an intersection-union test for inequality constraints are developed depending on the definition of null and alternative hypotheses. An empirical illustration is given for US insurance claim data.

Key words and phrases: Nonparametric, Concordance Ordering, Quadrant Dominance, Orthant Dominance, Copula, Inequality Constraint Tests, Risk Management, Loss Severity Distribution.

JEL Classification: C12, D81, G10, G21, G22.

1 Introduction

Random variables are concordant if they tend to be all large together or small together. Concordance of random variables conveys the idea of clustering of large and small events. An ordering of concordance was initially considered for two random variables by Yanagimoto and Okamoto (1960), Cambanis, Simons and Stout (1976) and Tchen (1980), and then extended by Joe (1990) to the multivariate case. This ordering corresponds to a natural notion of stochastic dominance between two distributions functions with fixed marginals. Large and small values will tend to be more often associated under the distribution which dominates the other one.

Detection of concordant behaviour is especially important in risk management of large portfolios of insurance contracts or financial assets. In these portfolios the main risk is the occurrence of many joint default events or simultaneous downside evolution of prices. An accurate knowledge of concordance between claims or financial asset prices will help to assess this risk of loss clustering and there from allows for taking appropriate action to ensure that the risk incurred by the financial institution remains within its stated risk appetite. Clearly the presence of concordance affects risk measures and asset allocations resulting from optimal portfolio selection. Analysis of concordance cannot be neglected and reveals much of the danger associated to a given position.

Modelling of concordance can be fully parametric or semiparametric. In the first case specific parametric forms are selected for the dependence structure and the margins, while in the second case margins are left unspecified. The dependence structure is expressed by means of a parametric copula function. Once these models estimated a natural concern of a risk manager ought to be: does the adopted modelling reflect the dependence structure present in the data safely enough? The aim of this paper is to provide inference tools to answer this question. These tools will be tests for the ordering of concordance for random variables. In fact we propose testing procedures for concordance order between the chosen model and the empirical distribution. This allows for checking whether the estimated parametric or semiparametric model gives a safe picture of the association between small and large observed losses.

The paper is organized as follows. In Section 2, we formally define concordance and its ordering. We also recall the definition of copula functions, as well as the classical Sklar’s representation theorem for multivariate distributions. Section 3 is devoted to illustrations of the practical relevance of the concordance order. Sections 4 and 5 are devoted to inference. The first one deals with the parametric setting, and the second one with the semiparametric setting. For each setting we develop inference formulated in terms of values taken by the random variables, i.e. loss or return levels, and in terms of probability levels. In Section 6 we develop the testing procedures, and describe the null and alternative hypotheses we are interested in. These procedures are closely related to inference tools for traditional first order and second order stochastic dominance, or for positive quadrant dependence. These tools also rely on distance and intersection-union tests for inequality constraints (see Davidson and Duclos (2000), Denuit and Scaillet (2001) and the references therein). An empirical illustration on US insurance claim data is proposed in Section 7. Section 8 contains some concluding remarks. Proofs are gathered in an appendix.
2 Concordance order

Let \( F \) and \( G \) denote an \( n \)-dimensional cdf, and \( \bar{F} \) and \( \bar{G} \) their corresponding survival function. Then \( G \) is more concordant than \( F \), written \( F \preceq_c G \), if

\[
F(y) \leq G(y) \quad \text{and} \quad \bar{F}(y) \leq \bar{G}(y), \quad \forall y \in \mathbb{R}^n.
\]  

(2.1)

The first inequality \( F(y) \leq G(y) \) corresponds to lower orthant dominance, while the second one \( \bar{F}(y) \leq \bar{G}(y) \) corresponds to upper orthant dominance. If both inequalities hold, large and small values will tend to be more often associated under \( G \) than \( F \). Condition (2.1) implies that \( F \) and \( G \) have the same \( j \)-th univariate marginal distribution (\( j = 1, \ldots, n \)), and that all bivariate and higher dimensional marginals of \( G \) are more concordant than the corresponding ones for \( F \). Hence we see that the ordering of concordance of variables is derived from comparisons of pairs of distributions with identical marginals (Yanagimoto and Okamoto (1960), Cambanis, Simons and Stout (1976), Tchen (1980), Joe (1990)). Besides the concordance order is equivalent to a partial order among the parameters for elliptically contoured distributions, such as the multivariate normal and student distributions. Henceforth we will freely use \( \preceq_c \) between cdf’s or random vectors to indicate that (2.1) holds.

At this stage, it is interesting to stress the very particular nature of the bivariate case (compared to dimension \( \geq 3 \)). Indeed, consider \( F \) and \( G \) with identical marginals. Then, it is easily seen that anyone of the two inequalities in (2.1) implies the other. It will be seen further in the paper that the bivariate case is really particular, mainly because the concordance order coincides with the supermodular order for random couples (this is not the case for random vectors of larger dimension).

The concordance order can also be characterised in terms of copulas. The marginal pdf and cdf of each element \( Y_j \) of \( Y \) at point \( y_j \), \( j = 1, \ldots, n \), will be written \( f_j(y_j) \), and \( F_j(y_j) \), respectively. How the joint distribution \( F \) is “coupled” to its univariate margins \( F_j \), can be described by a copula. While the joint distribution \( F \) provides complete information concerning the behaviour of \( Y \), copulas allow for separating dependence and marginal behaviour of the elements constituting \( Y = (Y_1, \ldots, Y_n)' \). Before defining formally a copula, we would like to refer the reader to Nelsen (1999) and Joe (1997) for more extensive theoretical treatments.

A \( n \)-dimensional copula \( C \) is simply (the restriction to \([0, 1]^n\) of) an \( n \)-dimensional cdf with unit uniform marginals. The reason why a copula is useful in revealing the link between the joint distribution and its margins transpires from the following theorem.

**Theorem 2.1. (Sklar’s Theorem)**

Let \( F \) be an \( n \)-dimensional cdf with margins \( F_1, \ldots, F_n \). Then there exists an \( n \)-copula \( C \) such that for all \( y \) in \( \mathbb{R}^n \),

\[
F(y) = C(F_1(y_1), \ldots, F_n(y_n)).
\]  

(2.2)

If \( F_1, \ldots, F_n \) are all continuous, then \( C \) is uniquely defined. Otherwise, \( C \) is uniquely determined on range \( F_1 \times \ldots \times \text{range } F_n \). Conversely, if \( C \) is an \( n \)-copula and \( F_1, \ldots, F_n \) are cdf’s, then the function \( F \) defined by (2.2) is an \( n \)-dimensional cdf with margins \( F_1, \ldots, F_n \).
Although copulas constitute a less well-known approach to describing dependence than correlation, they offer the best understanding of the general concept of dependency (see Embrechts, McNeil and Straumann (2000) for implications on risk management). In particular, copulas share the nice property that strictly increasing transformations of the underlying random variables result in the transformed variables having the same copula (what is not true for linear correlation).

As an immediate corollary of Sklar’s Theorem, we have

\[ C(u) = F(F_1^{-1}(u_1), ..., F_n^{-1}(u_n)) \]  

for any \( u \in [0,1]^n \). From Expression (2.3), we may observe that the dependence structure embodied by the copula can be recovered from the knowledge of the joint cdf \( F \) and its margins \( F_j \).

Let us now state the definition (2.1) of concordance in terms of copulas. We denote by \( \bar{C} \) the survival copula associated with \( C \), and use subscripts \( F \) and \( G \) to reflect their associated distribution. If \( F \) and \( G \) share the same univariate margins, \( G \) is more concordant than \( F \) if

\[ C_F(u) \leq C_G(u) \quad \text{and} \quad \bar{C}_F(u) \leq \bar{C}_G(u), \quad \forall u \in [0,1]^n. \]  

(2.4)

Note that the density \( c \) associated with the copula \( C \) is given by:

\[ c(u_1, \ldots, u_n) = \frac{\partial^n C(u_1, \ldots, u_n)}{\partial u_1 \ldots \partial u_n}. \]

The density \( f \) of \( F \) can be expressed in terms of the copula density \( c \) and the product of the univariate marginal densities \( f_j \):

\[ f(y_1, \ldots, y_n) = c(F_1(y_1), \ldots, F_n(y_n)) \prod_{j=1}^n f_j(y_i). \]

Obviously the latter equality does not depend on a parametric assumption for the multivariate distribution. In the inference part of this paper we will consider two cases. The first case will put some parametric assumption \( f_j(y_j; \beta_j) \) on the margins, while the second will not. The copula will be parameterised according to \( c(u_1, \ldots, u_n; \theta) \) in both cases.

Let us remark that independence between random variables can be characterised through copulas. Indeed, \( n \) random variables are independent if, and only if, their copula is \( C(u) = C^\perp(u) = \prod_{j=1}^n u_j \), for all \( u \in [0,1]^n \). \( C^\perp \) is further referred to as the independence copula.

Before proceeding further with tests for concordance ordering we illustrate its practical relevance for insurance and finance in the next section.

### 3 Applications of the concordance order

Stochastic orderings are binary relations defined on classes of probability distributions. They aim to mathematically translate intuitive ideas like “being larger” or “being more variable” for random quantities. They thus extend the classical mean-variance approach to compare riskiness.
Let us define the following utility classes. Let $U_1$ contain all non-decreasing utilities $u : \mathbb{R} \to \mathbb{R}$. Let $U_2$ be the restriction of $U_1$ to its concave elements. In what follows, we assume that decision-makers maximize a von Neumann-Morgenstern expected utility (but we mention that results involving $U_1$ and $U_2$ still hold in dual theories for choice under risk, see e.g. Denuit, Dhaene and VanWouwe (1999) for further information).

Let $Y_1$ and $Y_2$ be two random variables such that $Eu(Y_1) \leq Eu(Y_2)$ holds for all $u \in U_1$ (resp. $u \in U_2$), provided the expectations exist. Then $Y_1$ is said to be smaller than $Y_2$ in the stochastic dominance (resp. increasing concave order), denoted as $Y_1 \preceq_d Y_2$ (resp. $Y_1 \preceq_{icv} Y_2$). From the very definitions of $\preceq_d$ and $\preceq_{icv}$, we see that these stochastic orderings express the common preferences of the classes of profit-seeking decision-makers, and of profit-seeking risk-avers, respectively. This provides an intuitive meaning to rankings in the common preferences of the classes of profit-seeking decision-makers, and of profit-seeking risk-avers). For instance, since the statistic inference for stochastic orderings, see e.g. the review papers by Kroll and Levy (1980) and Levy (1992), the classified bibliography by Mosler and Scarsini (1993) and the book by Shaked and Shanthikumar (1994).

Statistical inference for $\preceq_d$ and $\preceq_{icv}$ is investigated in vast details in Davidson and Duclos (2000), where connections with economic and social welfare in different populations are explicated.

The bivariate version of $\preceq_c$ coincides with the supermodular order. This yields a host of useful results for random couples (which are no more valid in dimensions $\geq 3$ because concordance and supermodular orders are then strongly distinct, see e.g. Müller (1997) for further details on this issue).

The main interest of the concordance order among random couples comes from the following result of which we provide a short proof in Appendix A. It is a straightforward adaptation of the result of Dhaene and Goovaerts (1996) established in the convex actuarial setting.

**Proposition 3.1.** If $X \preceq_c Y$ then $Y_1 + Y_2 \preceq_{icv} X_1 + X_2$.

This means that when $X \preceq_c Y$ holds, every risk-averter agrees to say that $Y_1 + Y_2$ is less favourable than $X_1 + X_2$. Consequently, most insurance premiums and risk measures will be larger for $Y_1 + Y_2$ than for $X_1 + X_2$ (since the principles used to calculate such quantities are in accordance with the common preferences of risk-avers). For instance, since the function $x \mapsto -(x - \kappa)_+$, with $(\cdot)_+ = \max\{0, \cdot\}$, is concave for any $\kappa \in \mathbb{R}$, the inequality $E(X_1 + X_2 - \kappa)_+ \leq E(Y_1 + Y_2 - \kappa)_+$ holds true for all $\kappa$. The quantity $E(Y_1 + Y_2 - \kappa)_+$ is referred to as the stop-loss premium relating to the risk portfolio $Y_1 + Y_2$ in actuarial science ($\kappa$ is called the deductible). In finance, when appropriately discounted, it can be regarded as the price of a basket option with $Y_1$ and $Y_2$ as underlying assets and $\kappa$ as strike price.

In particular, we see that if $X \preceq_c Y$ then

$$\text{Var}[\alpha_1 X_1 + \alpha_2 X_2] \leq \text{Var}[\alpha_1 Y_1 + \alpha_2 Y_2]$$

for all $\alpha_1, \alpha_2 > 0$,

that is, the variance of a linear combination with positive weights of each coordinate will be lower when computed under $F$ than under $G$. In finance this means that portfolios with short sales constraints will be considered as more efficient in the mean-variance sense under $F$ than under $G$. 

4
For every $X$ with marginals $F_1$ and $F_2$, the stochastic inequalities

\[ (F_1^{-1}(U), F_2^{-1}(1-U)) \preceq_c (F_1^{-1}(U), F_2^{-1}(U)) \]

are valid, where $U$ stands for a unit uniform random variables. The random vectors involved in (3.1) are referred to as the Fréchet bounds; they represent perfect positive and negative dependence, respectively.

A powerful closure property of the concordance order is given next (it is easily established coming back to the definition (2.4) of $\preceq_c$ in terms of copulas). For all non-decreasing functions $\phi$ and $\psi$, the implication

\[(X_1, X_2) \preceq_c (Y_1, Y_2) \Rightarrow (\phi(X_1), \psi(X_2)) \preceq_c (\phi(Y_1), \psi(Y_2)),\]

holds true. So, $\preceq_c$ has a functional invariance property. As a simple illustration of the relevance of this result, suppose that we have a probability model (multivariate distribution) for dependent insurance losses of various kinds. If we decide that our interest now lies in modelling the logarithm of these losses, the $\preceq_c$ ranking will not change. Similarly if we change from a model of percentage returns on several financial assets to a model of logarithmic returns. This also clearly shows that an ordering in the $\preceq_c$ sense only depends on the underlying copula once the marginals have been fixed.

It is also known that the concave order is closely related to the Lorenz order. Let us recall that the Lorenz order is defined by means of pointwise comparison of Lorenz curves. The latter is used in economics to measure the inequality of incomes (see Beach and Davidson (1983), Dardanoni and Forcina (1999) for related inference). More precisely, let $Y$ be a non-negative random variable with cdf $F$. The Lorenz curve $L$ associated with $Y$ is then defined by

\[ L(p) = \frac{1}{EY} \int_{t=0}^{p} F^{-1}(u) du, \quad p \in [0, 1]. \]

When $Y$ represents the income of the individuals in some population, $L$ maps $p \in [0, 1]$ to the proportion of the total income of the population which accrues to the poorest $100p\%$ of the population.

Consider two non-negative random variables $Y_1$ and $Y_2$ with finite expectations. Then, $Y_1$ is said to be smaller than $Y_2$ in the Lorenz order, henceforth denoted by $Y_1 \preceq_{\text{Lorenz}} Y_2$, when $L_1(p) \geq L_2(p)$ for all $p \in [0, 1]$. When $Y_1 \preceq_{\text{Lorenz}} Y_2$ holds, $Y_1$ does not exhibit more inequality in the Lorenz sense than does $Y_2$. A standard reference for $\preceq_{\text{Lorenz}}$ is Arnold (1987).

Provided $EY_1 = EY_2$, it can be shown $Y_1 \preceq_{\text{Lorenz}} Y_2 \iff Y_2 \preceq_{\text{cv}} Y_1$. Hence, if $X \preceq_c Y$ then $X_1 + X_2 \preceq_{\text{Lorenz}} Y_1 + Y_2$ in virtue of Proposition 3.1.

The usefulness of this relation can be illustrated as follows. Let $(X_1, X_2)$ represent the incomes of married couples ($X_1$ for husband and $X_2$ for his wife) in some population. Assume that for some sociological or legal reason, $(X_1, X_2)$ is replaced with $(Y_1, Y_2)$ such that $(X_1, X_2) \preceq_c (Y_1, Y_2)$ (in words, this means that the incomes of husband and wife become more positively dependent, but the marginal incomes for married men and women remain unchanged). Then, this increases the inequality of incomes at the couples level since $X_1 + X_2 \preceq_{\text{Lorenz}} Y_1 + Y_2$ holds.
Let us now provide an illustration in relation with insurance premium calculation principles. Consider an insurance company with initial wealth $w$ and with a utility function $u \in U_2$. The company covers a collective risk with an aggregate claim amount $S$. It wonders whether it should cover a new risk $X$, and if affirmative how to set the premium amount for this new risk. Of course, $X$ is correlated to $S$ (at least to some extent). The amount of premium $\pi(X)$ is determined following the adoption of an economic decision principle. We assume here that the insurance company sets its price for coverage $\pi(X)$ as the solution of the equation

$$Eu(w - S + \pi(X) - X) = u(w - S).$$

(3.2)

This way of computing premiums is classical in actuarial science; see e.g. Goovaerts and al. (1990) for more details. Condition (3.2) expresses that the premium $\pi(X)$ is fair in terms of utility: the right-hand side of (3.2) represents the utility of not issuing the contract; the left-hand side of (3.2) represents the expected utility of the insurer assuming the random financial loss $X$. Therefore (3.2) means that the expected utility of wealth with the contract is equal to the utility without the contract. This type of pricing principle is sometimes also used in finance to set prices of derivative assets (see e.g. Davis (1997), Karatzas and Kou (1996)).

In such a case, it is possible to show that the premium $\pi(X)$ should increase with the positive dependence existing between $X$ and $S$. Indeed, the implication

$$(S, X) \preceq_c (S, Y) \Rightarrow \pi(X) \leq \pi(Y)$$

holds true, meaning that the safety loading increases with the dependence existing between the new risk and those already written by the company.

4 Inference under parametric specification

Now that the relevant theoretical concepts and applications have been presented, we may turn our attention to inference. We consider a setting made of i.i.d. observations \(\{Y_t; t = 1, ..., T\}\) of a random vector $Y$ taking values in $\mathbb{R}^n$. These data may correspond to either observed individual losses on $n$ insurance contracts, amounts of claims reported by a given policyholder on $n$ different guarantees in a multiline product or observed returns of $n$ financial assets. We begin with fully parametric specifications, and analyse two cases. The first one is based on a grid of loss or return levels, while the second one is based on a grid of probability levels.

4.1 Inference based on loss levels

Let us start with the parametric family

$$\{F(y; \nu) = C(F_1(y_1; \beta_1), ..., F_n(y_n; \beta_n); \theta), \nu = (\beta', \theta')' \in \Psi \subset \mathbb{R}^{q+p}\}.$$ 

This parametric family is specified in terms of a parametric copula $C(u; \theta)$ and parametric margins $F_j(y_j; \beta_j)$, $j = 1, \ldots, n$. The $q$-dimensional vector $\beta = (\beta_1', \ldots, \beta_n')'$ and the $p$-dimensional vector $\theta$ forming $\nu$ are jointly estimated by pseudo maximum likelihood. The
estimator \( \hat{\nu} \) is derived from
\[
\max_{\beta, \theta} \frac{1}{T} \sum_{t=1}^{T} \ln c(F_1(Y_{1t}; \beta_1), \ldots, F_n(Y_{nt}; \beta_n); \theta) \sum_{j=1}^{n} \ln f_i(Y_{it}; \beta_i),
\]
and its limit, i.e. the pseudo true value, is denoted by \( \nu_0 = (\beta'_0, \theta'_0)' \). We wish to check whether \( F(\cdot; \nu_0) \) is more concordant than the true distribution function \( F_0(\cdot) \), namely
\[
F_0(y) \leq F(y; \nu_0) \quad \text{and} \quad \bar{F}_0(y) \leq \bar{F}(y; \nu_0), \quad \forall y \in \mathbb{R}^n.
\]

As in traditional stochastic dominance tests or positive quadrant dependence tests we use a version of the conditions defining concordance on a predetermined grid, and only consider fixed number of distinct points, say \( d \) points \( \mathbf{y}_i = (y_{i1}, \ldots, y_{in})' \) in \( \mathbb{R}^n \). These points will typically span the whole range of possible values. We define \( D_i^1 = F(y_i; \nu_0) - F_0(y_i), \) and \( \bar{D}_i^1 = \bar{F}(y_i; \nu_0) - \bar{F}(y_i), \) and set \( \mathbf{D}_1 = (D_1^1, \ldots, D_d^1)' \), and \( \mathbf{D}_1 = (\bar{D}_1^1, \ldots, \bar{D}_d^1)' \). The testing procedures described in Section 6 will be built from the empirical counterpart \( \hat{D}_F, \) resp. \( \hat{\bar{D}}_F, \) resp. \( \hat{D}_F \), obtained by substituting estimated and empirical distributions for the unknown parametric and true distributions. The joint empirical distribution is given by
\[
\hat{F}(y_i) = \frac{1}{T} \sum_{t=1}^{T} \prod_{j=1}^{n} \mathbb{I}[Y_{jt} \leq y_{ij}], \quad i = 1, \ldots, d.
\]

The empirical counterparts are thus equal to:
\[
\hat{D}_i^1 = C(F_1(y_{i1}; \hat{\beta}_1), \ldots, F_n(y_{im}; \hat{\beta}_n); \hat{\theta}) - \hat{F}(y_i),
\]
and
\[
\hat{\bar{D}}_i^1 = \bar{C}(F_1(y_{i1}; \hat{\beta}_1), \ldots, F_n(y_{im}; \hat{\beta}_n); \hat{\theta}) - \bar{F}(y_i).
\]

The following proposition characterizes the joint asymptotic distribution of \( \hat{\mathbf{D}}_1 \) and \( \hat{\bar{\mathbf{D}}}_1 \). The subindex in \( E_0 \) and \( \mathbb{C}ov_0 \) refers to integration w.r.t. the true distribution \( F_0 \).

**Proposition 4.1.** The random vector \( \sqrt{T}(\hat{\mathbf{D}}_1 - \mathbf{D}_1) \), resp. \( \sqrt{T}(\hat{\bar{\mathbf{D}}}_1 - \bar{\mathbf{D}}_1) \), converges in distribution to a \( d \)-dimensional normal random variable with mean zero and covariance matrix \( \mathbf{V}_1, \) resp. \( \bar{\mathbf{V}}_1, \) whose elements are
\[
v_{1,kl} = \lim_{T \to \infty} T \mathbb{C}ov_0 \left[ \hat{D}_k^l, \hat{D}_l^k \right] = B_{v_0}^{kl} \mathbb{C}ov_0 \left[ S_{v_0}^k, S_{v_0}^l \right] B_{v_0}^{lk}, \quad k, l = 1, \ldots, d,
\]
with
\[
B_{v_0}^{kl} = \left( \nabla_{v_0}^l C' J_{v_0}^{-1}, 1 \right)',
\]
and
\[
S_{v_0}^i = \left( \frac{\partial}{\partial v_i} \log f(Y; \nu_0), \mathbb{I}[Y \leq y_i] \right)',
\]
and
\[
\bar{v}_{1,kl} = \lim_{T \to \infty} T \mathbb{C}ov_0 \left[ \hat{\bar{D}}_k^l, \hat{\bar{D}}_l^k \right] = \bar{B}_{v_0}^{kl} \mathbb{C}ov_0 \left[ \bar{S}_{v_0}^k, \bar{S}_{v_0}^l \right] \bar{B}_{v_0}^{lk}, \quad k, l = 1, \ldots, d,
\]
with
\[ \bar{B}^i_{\nu_0} = (\nabla^i_{\nu_0} \bar{C}', J^{-1}_{\nu_0}, 1)' \, , \]
\[ S^i_{\nu_0} = \left( \frac{\partial}{\partial \nu'} \log f(Y; \nu_0), \mathbb{I}[Y > y_i] \right)' \, , \]

where
\[ J_{\nu_0} = E_0 \left[ -\frac{\partial^2}{\partial \nu \partial \nu'} \log f(Y; \nu_0) \right] \, , \]

while
\[ \lim_{T \to \infty} T \text{Cov}_{\nu_0} \left[ \hat{D}^k_1, \hat{D}^l_1 \right] = B^{kl}_{\nu_0} \text{Cov}_{\nu_0} \left[ S^k_{\nu_0}, S^l_{\nu_0} \right] \bar{B}^l_{\nu_0}, \quad k, l = 1, \ldots, d. \]

A consistent estimate of each covariance can be obtained by replacing expectations with empirical averages, and unknown parameter values by their estimates.

### 4.2 Inference based on probability levels

Let us now proceed with the analogous quantities when we use probability levels instead of loss levels, and take \( d \) points \( u_i = (u_{i1}, \ldots, u_{in})' \), with \( u_{ij} \in (0, 1) \), \( i = 1, \ldots, d, \ j = 1, \ldots, n \). We assume hereafter that the cdf \( F_{j0} \) is such that the equation \( F_{j0}(y) = u_{ij} \) admits a unique solution denoted \( \zeta_{ij} \), \( i = 1, \ldots, d, \ j = 1, \ldots, n \), while \( f_{j0}(\zeta_{ij}) > 0 \) at each quantile \( \zeta_{ij} \). We denote the stack of the univariate quantiles \( \zeta_{ij} \) by \( \zeta_i \).

We may then define \( D^2_i = C(u_i; \theta_0) - F_0(\zeta_i) \), and \( D^2 = (D^2_1, \ldots, D^2_d)' \). The survival quantities will be \( \hat{D}^2_i = \hat{C}(u_i; \hat{\theta}) - \hat{F}(\hat{\zeta}_i) \), and \( \hat{D}^2 = (\hat{D}^2_1, \ldots, \hat{D}^2_d)' \). The empirical counterparts are then \( \hat{D}^2_i = C(u_i; \hat{\theta}) - \hat{F}(\hat{\zeta}_i) \), and \( \hat{D}^2_i = \hat{C}(u_i; \hat{\theta}) - \hat{F}(\hat{\zeta}_i) \), where \( \hat{\zeta}_i = (\hat{\zeta}_{i1}, \ldots, \hat{\zeta}_{in})' \) is made of the empirical univariate quantiles \( \hat{\zeta}_{ij} \). The main difference when compared with the previous case is that the loss levels are no more given deterministic values, but quantiles estimated on the basis of sample information, and thus random quantities.

**Proposition 4.2.** The random vector \( \sqrt{T}(\hat{D}_2 - D^2) \), resp. \( \sqrt{T}(\hat{D}_2 - \bar{D}_2) \), converges in distribution to a \( d \)-dimensional normal random variable with mean zero and covariance matrix \( V_2 \), resp. \( \bar{V}_2 \), whose elements are

\[ \nu_{2,kl} = \lim_{T \to \infty} T \text{Cov}_{\nu_0} \left[ \hat{D}^k_2, \hat{D}^l_2 \right] = B^{kl}_{\theta_0} \text{Cov}_{\nu_0} \left[ S^k_{\theta_0}, S^l_{\theta_0} \right] B^l_{\theta_0}, \quad k, l = 1, \ldots, d, \]

with
\[ B^i_{\theta_0} = \left( \nabla^i_{\theta_0} C' J_{\theta_0}^{-1}, -1, \frac{\partial F_0(\zeta_i)}{f_{00}(\zeta_{i1})}, \ldots, \frac{\partial F_0(\zeta_i)}{f_{00}(\zeta_{in})} \right)' \, , \]
\[ S^i_{\theta_0} = \left( \frac{\partial}{\partial \theta'} \log f(Y; \nu_0), \mathbb{I}[Y \leq \zeta_{i1}], \mathbb{I}[Y_1 \leq \zeta_{i1}], \ldots, \mathbb{I}[Y_n \leq \zeta_{in}] \right)' \, , \]

and
\[ \bar{v}_{2,kl} = \lim_{T \to \infty} T \text{Cov}_{\nu_0} \left[ \bar{D}^k_2, \bar{D}^l_2 \right] = \bar{B}^{kl}_{\theta_0} \text{Cov}_{\nu_0} \left[ \bar{S}^k_{\theta_0}, \bar{S}^l_{\theta_0} \right] \bar{B}^l_{\theta_0}, \quad k, l = 1, \ldots, d, \]
with

\[ B_{\theta_0}^i = \left( \nabla_{\theta_0}^i C^t J_{\theta_0}^{-1}, -1, \frac{\partial F_0(\zeta_1)}{\partial x_1}, ..., \frac{\partial F_0(\zeta_m)}{\partial x_n} \right)^t, \]

\[ S_{\theta_0}^i = \left( \frac{\partial}{\partial \theta^t} \log f(Y; \nu_0), \mathbb{I}[Y > \zeta_i], \mathbb{I}[Y_1 < \zeta_{i1}], ..., \mathbb{I}[Y_n \leq \zeta_{in}] \right)^t, \]

where

\[ J_{\nu_0} = E_0 \left[ -\frac{\partial^2}{\partial \theta \partial \theta^t} \log f(Y; \nu_0) \right], \]

while the elements of the cross covariance matrix \( CV \) are

\[ cv_{2,kl} = \lim_{T \to \infty} T \text{Cov} \left[ \hat{D}_k^l, \hat{D}_2^l \right] = B_{\theta_0}^{k, l} \text{Cov} \left[ S_{\theta_0}^k, S_{\theta_0}^l \right] B_{\theta_0}^0, \quad k, l = 1, ..., d. \]

Some of the asymptotic covariances involve derivatives of \( F_0 \) and the univariate densities \( f_{j0} \). These quantities may be estimated by standard kernel methods (see e.g. Scott (1992)) in order to deliver a consistent covariance estimate. For example we may take a Gaussian kernel and different bandwidth values \( h_j \) in each dimension, which leads to:

\[ \frac{\partial \hat{F}(\zeta_i)}{\partial x_j} = (Th_j)^{-1} \sum_{t=1}^T \varphi \left( \frac{Y_{jt} - \hat{\zeta}_{ij}}{h_j} \right) \prod_{l \neq j} \Phi \left( \frac{Y_{lt} - \hat{\zeta}_{il}}{h_l} \right), \]

\[ \hat{f}_j(\zeta_{ij}) = (Th_j)^{-1} \sum_{t=1}^T \varphi \left( \frac{Y_{jt} - \hat{\zeta}_{ij}}{h_j} \right), \]

where \( \varphi \) and \( \Phi \) denote the pdf and cdf of a standard Gaussian variable. In the empirical section of the paper, we opt for the standard choice (rule of thumb) for the bandwidths \( h_j \), that is \( 1.05T^{-1/5} \) times the estimated standard deviation of \( Y_j \).

5 Inference under semiparametric specification

5.1 Inference based on loss levels

The previous section was devoted to the fully parametric specification. If we wish to be less restrictive a priori on the univariate margins, we may leave them unspecified, and use the family

\[ \{ F(y; \theta) = C(F_1(y_1), ..., F_n(y_n); \theta), \theta \in \Theta \subset \mathbb{R}^p \}. \]

Hence we get a semiparametric setting only parameterised through \( C(u; \theta) \). The estimator \( \hat{\theta} \) of \( \theta \) is obtained by

\[ \max_{\theta} \frac{1}{T} \sum_{t=1}^T \ln c(\hat{F}_1(Y_{1t}), ..., \hat{F}_n(Y_{nt}); \theta), \]

where

\[ \hat{F}_j(y) = \frac{1}{T} \sum_{t=1}^T \mathbb{I}[Y_{jt} \leq y], \quad j = 1, ..., n. \]
Its limit is denoted by $\theta_0^*$, and will correspond to $\theta_0$ (the true value) if both copula and margins are well specified in the parametric case. The asymptotic distribution of $\hat{\theta}$ under correct specification is given in Genest, Ghouci and Rivest (1995) and Shih and Louis (1995). Again we wish to check whether $F(\cdot; \theta_0^*)$ is more concordant than the true distribution function $F_0(\cdot)$, namely

$$F_0(y) \leq F(y; \theta_0^*) \quad \text{and} \quad F_0(y) \leq \hat{F}(y; \theta_0^*), \quad \forall y \in \mathbb{R}^n. \quad (5.2)$$

Along the same lines as in the parametric setting we define

$$\hat{D}_3 = C(\hat{F}_1(y_{i1}), ..., \hat{F}_n(y_{in}); \hat{\theta}) - \hat{F}(y_i),$$

and

$$\hat{D}_3 = C(\hat{F}_1(y_{i1}), ..., \hat{F}_n(y_{in}); \hat{\theta}) - \hat{F}(y_i),$$

together with their corresponding stacks $\hat{D}_3$ and $\hat{D}_3$.

**Proposition 5.1.** The random vector $\sqrt{T}(\hat{D}_3 - D_3)$, resp. $\sqrt{T}(\hat{D}_3 - D_3)$, converges in distribution to a $d$-dimensional normal random variable with mean zero and covariance matrix $V_3$, resp. $V_3$, whose elements are

$$v_{3,kl} = \lim_{T \to \infty} T \text{Cov}_0 \left[ \hat{D}_3, \hat{D}_3 \right] = B_{00}^{kl} \text{Cov}_0 \left[ S_{05}^{k}, S_{05}^{l} \right] B_{00}^{kl}, \quad k, l = 1, ..., d,$$

with

$$B_{00}^{kl} = \left( \nabla_{\theta_0} C_{\theta_0}^{-1} \nabla_{u_1} C, ..., \nabla_{u_n} C, -1 \right)'$$

$$S_{05}^{k} = \left( U_{05}', [Y_1 \leq y_{i1}], ..., [Y_n \leq y_{in}], [Y \leq y_i] \right)'$$

and

$$\bar{v}_{3,kl} = \lim_{T \to \infty} T \text{Cov}_0 \left[ \hat{D}_3, \hat{D}_3 \right] = \bar{B}_{00}^{kl} \text{Cov}_0 \left[ S_{05}^{k}, S_{05}^{l} \right] \bar{B}_{00}^{kl}, \quad k, l = 1, ..., d,$$

with

$$\bar{B}_{00}^{kl} = \left( \nabla_{\theta_0} C_{\theta_0}^{-1}, \nabla_{u_1} C, ..., \nabla_{u_n} C, -1 \right)'$$

$$\bar{S}_{05}^{k} = \left( U_{05}', [Y_1 \leq y_{i1}], ..., [Y_n \leq y_{in}], [Y > y_i] \right)'$$

where

$$J_{05} = E_0 \left[ -\frac{\partial^2}{\partial \theta_0 \partial \theta_0} \log c(F_{10}(Y_1), ..., F_{n0}(Y_n); \theta_0^*) \right]$$

and

$$U_{05} = \frac{\partial}{\partial \theta_0} \log c(F_{10}(Y_1), ..., F_{n0}(Y_n); \theta_0^*)$$

$$+ \sum_{j=1}^{n} \int_{\mathbb{R}^n} [Y_j \leq z_j] \frac{\partial^2}{\partial \theta_0 \partial u_j} \log c(F_{10}(z_1), ..., F_{n0}(z_n); \theta_0^*) dF_0(z_1, ..., z_n)$$

while

$$\lim_{T \to \infty} T \text{Cov}_0 \left[ \hat{D}_3, \hat{D}_3 \right] = B_{05}^{kl} \text{Cov}_0 \left[ S_{05}^{k}, S_{05}^{l} \right] \bar{B}_{00}^{kl}, \quad k, l = 1, ..., d.$$
5.2 Inference based on probability levels

If we prefer to set probability levels, we will use

$$\hat{D}^i_4 = C(u_i; \tilde{\theta}) - \hat{F}(\hat{\zeta}_i),$$

and

$$\hat{\bar{D}}^i_4 = \bar{C}(u_i; \tilde{\theta}) - \hat{F}(\hat{\zeta}_i).$$

This leads to the following proposition which is equivalent to Proposition 2 of the parametric setting.

**Proposition 5.2.** The random vector $\sqrt{T}(\hat{D}_4 - D_4)$, resp. $\sqrt{T}(\hat{\bar{D}}_4 - \bar{D}_4)$, converges in distribution to a $d$-dimensional normal random variable with mean zero and covariance matrix $V_4$, resp. $\bar{V}_4$, whose elements are

$$v_{4,kl} = \lim_{T \to \infty} T \text{Cov}_0 \left[ \hat{D}^k_4, \hat{D}^l_4 \right] = B^k_{\theta_0^*} \text{Cov}_0 \left[ S^k_{\theta_0^*}, S^l_{\theta_0^*} \right] B^l_{\theta_0^*}, \quad k, l = 1, \ldots, d,$$

with

$$B^i_{\theta_0^*} = \left( \nabla^i_{\theta_0^*} C' J_{\theta_0^*}^{-1}, -1, \frac{\partial F_0(\zeta_1)}{\partial x_1}, \ldots, \frac{\partial F_0(\zeta_n)}{\partial x_n} \right)'$$

and

$$S^i_{\theta_0^*} = \left( U_{\theta_0^*}, \mathbb{I}[Y \leq \zeta_i], \mathbb{I}[Y_1 \leq \zeta_i], \ldots, \mathbb{I}[Y_n \leq \zeta_i] \right)'$$

and

$$\bar{v}_{4,kl} = \lim_{T \to \infty} T \text{Cov}_0 \left[ \hat{\bar{D}}^k_4, \hat{\bar{D}}^l_4 \right] = \bar{B}^k_{\theta_0^*} \text{Cov}_0 \left[ \bar{S}^k_{\theta_0^*}, \bar{S}^l_{\theta_0^*} \right] \bar{B}^l_{\theta_0^*}, \quad k, l = 1, \ldots, d,$$

with

$$\bar{B}^i_{\theta_0^*} = \left( \nabla^i_{\theta_0^*} \bar{C}' J_{\theta_0^*}^{-1}, -1, \frac{\partial F_0(\zeta_1)}{\partial x_1}, \ldots, \frac{\partial F_0(\zeta_n)}{\partial x_n} \right)'$$

and

$$\bar{S}^i_{\theta_0^*} = \left( U_{\theta_0^*}, \mathbb{I}[Y > \zeta_i], \mathbb{I}[Y_1 \leq \zeta_i], \ldots, \mathbb{I}[Y_n \leq \zeta_i] \right)'$$

where

$$J_{\theta_0^*} = E_0 \left[ -\frac{\partial^2}{\partial \theta \partial \theta} \log c(F_{10}(Y_1), \ldots, F_{n0}(Y_n) ; \theta_0^*) \right]$$

and

$$U_{\theta_0^*} = \frac{\partial}{\partial \theta} \log c(F_{10}(Y_1), \ldots, F_{n0}(Y_n) ; \theta_0^*)$$

$$+ \sum_{j=1}^n \int_{\mathbb{R}^n} \mathbb{I}[Y_j \leq z_j] \frac{\partial^2}{\partial \theta \partial u_j} \log c(F_{10}(z_1), \ldots, F_{n0}(z_n) ; \theta_0^*) dF_0(z_1, \ldots, z_n)$$

while

$$\lim_{T \to \infty} T \text{Cov}_0 \left[ \hat{D}^k_4, \hat{\bar{D}}^l_4 \right] = B^k_{\theta_0^*} \text{Cov}_0 \left[ S^k_{\theta_0^*}, S^l_{\theta_0^*} \right] B^l_{\theta_0^*}, \quad k, l = 1, \ldots, d.$$
6 Testing procedures

The distributional results of Propositions 4.1-4.2 and 5.1-5.2 are the building blocks of the testing procedures. Let $Z_k$, resp. $\hat{Z}_k$, be the stack of $D_k$, resp. $\hat{D}_k$, and $\bar{D}_k$, resp. $\bar{\hat{D}}_k$, $k = 1, \ldots, 4$.

The null hypothesis of a test for concordance may be written as

$$H^0_k = \{ Z_k : Z_k \geq 0 \},$$

with alternative hypothesis:

$$H^1_k = \{ Z_k : Z_k \text{ unrestricted} \}.$$ 

To examine these hypotheses we will use the usual distance tests for inequality constraints, initiated in the multivariate one-sided hypothesis literature for positivity of the mean (Bartholomew (1959a,b)). They are relevant when one or several components of $\hat{Z}_k$ are found to be negative (in such a case one wants to know whether this invalidates concordance).

Let $\tilde{Z}_k$, be solution of the constrained quadratic minimisation problem:

$$\inf_{Z} T( Z - \tilde{Z}_k)' \Sigma_k^{-1} ( Z - \tilde{Z}_k ) \quad s.t. \quad Z \geq 0, \quad (6.1)$$

where $\hat{\Sigma}_k$ is a consistent estimate of the asymptotic covariance matrix $\Sigma_k$ of $\sqrt{T}Z_k$, and put

$$\hat{\xi}_k = T( \tilde{Z}_k - \hat{Z}_k)' \Sigma_k^{-1} ( \tilde{Z}_k - \hat{Z}_k ).$$

Roughly speaking, $\tilde{Z}_k$ is the closest point to $\hat{Z}_k$ under the null in the distance measured in the metric of $\hat{\Sigma}_k$, and the test statistic $\hat{\xi}_k$ is the distance between $\tilde{Z}_k$ and $\hat{Z}_k$. The idea is to reject $H^0_k$ when this distance becomes too large.

The asymptotic distribution of $\hat{\xi}_k$ under the null (see e.g. Gouriéroux, Holly and Monfort (1982), Kodde and Palm (1986), Wolak (1989a,b)) is such that for any positive $x$:

$$P[\hat{\xi}_k \geq x] = \sum_{i=1}^{d} P[\chi_i^2 \geq x]w(d, d - i, \hat{\Sigma}_k),$$

where the weight $w(d, d - i, \hat{\Sigma}_k)$ is the probability that $\tilde{Z}_k$ has exactly $d - i$ positive elements.

Computation of the solution $\tilde{Z}_k$ can be performed by a numerical optimisation routine for constrained quadratic programming problems available in most statistical softwares. Closed form solutions for the weights are available for $d \leq 4$ (Kudo (1963)). For higher dimensions one usually relies on a simple Monte Carlo technique as advocated in Gouriéroux, Holly and Monfort (1982) (see also Wolak (1989a)). Indeed it is enough to draw a given large number of realisations of a multivariate normal with mean zero and covariance matrix $\hat{\Sigma}_k$. Then use these realisations as $\tilde{Z}_k$ in the above minimisation problem (6.1), compute $\tilde{Z}_k$, and count the number of elements of the vector greater than zero. The proportion of draws such that $\tilde{Z}_k$ has exactly $d - i$ elements greater that zero gives a Monte Carlo estimate of $w(d, d - i, \hat{\Sigma}_k)$. If one wishes to avoid this computational burden, the upper and lower bound critical values of Kodde and Palm (1986) can be adopted.
Let us now turn our attention to the second testing procedure aimed to test for non-concordance. It is based on the null hypothesis:

$$\tilde{H}_k^0 = \{ Z_k : Z^l_k \leq 0 \text{ for some } l, l = 1, \ldots, 2d \},$$

and the alternative hypothesis:

$$\tilde{H}_k^1 = \{ Z_k : Z^l_k > 0 \text{ for all } l \}.$$

These hypotheses will be tested through intersection-union tests based on the minimum of a $t$-statistic. They are used when all components of $\tilde{Z}_k$ are found to be positive. The question is then whether this suffices to ensure concordance.

Let $\tilde{\gamma}^l_k = \sqrt{T} \tilde{Z}^l_k / \sqrt{\hat{\sigma}_{k,l}}$, where $\hat{\sigma}_{k,l}$ is a consistent estimate of the asymptotic standard deviation of $\sqrt{T} \tilde{Z}^l_k$, $l = 1, \ldots, 2d$. Then under $\tilde{H}_k^0$, the limit of $P[\inf \tilde{\gamma}^l_k > z_{1-\alpha}]$ will be less or equal to $\alpha$, and exactly equal to $\alpha$ if $Z^l_k = 0$ for a given $l$ and $Z^s_k > 0$ for $s \neq l$, while its limit is one under $\tilde{H}_k^1$. Hence the test consisting of rejecting $\tilde{H}_k^0$ when $\inf \tilde{\gamma}^l_k$ is above the $(1 - \alpha)$-quantile $z_{1-\alpha}$ of a standard normal distribution has an upper bound $\alpha$ on the asymptotic size and is consistent (see e.g. HOWES (1993), KAUR, PRAKASA RAO and SINGH (1994)).

Power issues are extensively studied for stochastic dominance and nondominance tests in DARDANONI and FORCINA (1999) (see also the comments in DAVIDSON and DUCLOS (2000)). They carry over to our case. First, approaches based on distance tests exploit the covariance structure, and are thus expected to achieve better power properties relative to approaches, such as ones based on $t$-statistics, that do not account for it. In a set of Monte Carlo experiments, they find that, indeed, distance tests are worth the extra amount of computational work. Second, it is possible that nonrejection of the null of dominance, here concordance, by distance tests occurs along with the nonrejection of the null of non-dominance, here non-concordance, by intersection-union tests. This is due to the highly conservative nature of the latter, and will typically occur in our setting if $\tilde{Z}_k$ is close enough to zero for a number of coordinates. This empirical feature has already been observed on tests for positive quadrant dependence (PQD) in DENUIT and SCAILLET (2001).

7 An empirical illustration: US Losses and ALAE’s

7.1 Presentation of the data

Often insurance processes involve correlated pairs of variables. A fine example is the LOSS and allocated adjustment expenses (ALAE, in short) on a single claim. ALAE’s are type of insurance company expenses that are specifically attributable to the settlement of individual claims such as lawyers’ fees and claims investigation expenses. The joint modelling in parametric settings of those two variables is examined by FREES and VALDEZ (1998) who choose the Pareto distribution to model the margins and select Gumbel and Frank’s copulas. Both models express PQD by their estimated parameter values. KLUGMAN and PARSA (1999) opt for the Inverse Paralogistic for LOSS and for the Inverse Burr for ALAE’s and use Frank’s copula to model the dependence between them. DENUIT and SCAILLET (2001)
test the existence of PQD for LOSS and ALAE using a nonparametric approach and find that, as both previous models suggest, significant positive quadrant dependence exists.

The database we have considered consist in $T = 1,466$ uncensored observed values of the random vector $(LOSS, ALAE)$. The estimated values for Pearson’s $r$, Kendall’s $\tau$ and Spearman’s $\rho$ are 0.381, 0.307 and 0.444, respectively; all of them are significantly positive at a 1% level. Summary statistics for $(LOSS, ALAE)$ are provided in Table 7.1.

<table>
<thead>
<tr>
<th></th>
<th>LOSS</th>
<th>ALAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>37,109.58</td>
<td>12,017.47</td>
</tr>
<tr>
<td>Std Dev.</td>
<td>92,512.80</td>
<td>26,712.35</td>
</tr>
<tr>
<td>Skew.</td>
<td>10.95</td>
<td>10.07</td>
</tr>
<tr>
<td>Kurt.</td>
<td>209.62</td>
<td>152.39</td>
</tr>
<tr>
<td>Min</td>
<td>10.00</td>
<td>15.00</td>
</tr>
<tr>
<td>Max</td>
<td>2,173,595.00</td>
<td>501,863.00</td>
</tr>
<tr>
<td>1st Quart.</td>
<td>3,750.00</td>
<td>2,318.25</td>
</tr>
<tr>
<td>Median</td>
<td>11,048.50</td>
<td>5,420.50</td>
</tr>
<tr>
<td>3rd Quart.</td>
<td>32,000.00</td>
<td>12,292.00</td>
</tr>
</tbody>
</table>

Table 7.1: Summary statistics for variables LOSS and ALAE.

Because some very high values of the variables are contained in the data set, we will work on a logarithmic scale to represent the data. This will not affect testing for concordance ordering since this order enjoys a functional invariance property (cf. Section 3.2). Figure 7.1 shows the kernel estimator of the bivariate pdf of the couple $(\log(LOSS), \log(ALAE))$, together with its contour plot. This estimation relies on a product of Gaussian kernels and bandwidth values selected by the standard rule of thumb (Scott (1992)). The graphs obviously suggest strong positive dependence between both variables.

### 7.2 Inference under parametric specification

First, the parametric framework suggested by Frees and Valdez (1998) is studied. It relies on a Gumbel copula

$$C(u_1, u_2; \theta) = \exp \left\{ - \left[ (-\ln(u_1))^\theta + (-\ln(u_2))^\theta \right]^{1/\theta} \right\},$$

for the dependence structure and Pareto distributions

$$F_i(x) = 1 - \left( 1 + \xi_i \frac{x}{\gamma_i} \right)^{-1/\xi_i}, \quad i = 1, 2,$$

for the marginal behaviours.

Estimated values for the parameter $\nu = (\xi_1, \gamma_1, \xi_2, \gamma_2, \theta)$' are shown in Table 7.2.

In the testing procedures we opt for inference based on probability levels, and take 81 points built on the grid $\{0.1, 0.2, \ldots, 0.9\} \times \{0.1, 0.2, \ldots, 0.9\}$. Since 105 of the 162 components of the vector $\hat{Z}_2$ are negative, with 30 among them less than -0.1, a concordance test is applied. See Figure 7.2 for a representation.
Figure 7.1: Kernel estimation of the bivariate pdf for \((\log(\text{LOSS}), \log(\text{ALAE}))\).
Figure 7.2: Estimated copulas: Gumbel (top), empirical (mid) and difference (bottom).
Table 7.2: Estimated parameter values of the bivariate distribution of (LOSS, ALAE).

<table>
<thead>
<tr>
<th></th>
<th>Distribution</th>
<th>( \hat{\xi} )</th>
<th>( \hat{\beta} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOSS</td>
<td>Pareto</td>
<td>0.760</td>
<td>12,816.9</td>
</tr>
<tr>
<td>ALAE</td>
<td>Pareto</td>
<td>0.425</td>
<td>6,756.5</td>
</tr>
<tr>
<td>Copula</td>
<td>Gumbel</td>
<td>( \theta = 1.425 )</td>
<td></td>
</tr>
</tbody>
</table>

To compute the solution \( \tilde{Z}_2 \) of the quadratic minimisation problem (6.1), a local minimiser for nonlinear functions subject to boundary constraints is used (specifically, \texttt{nlminb} in Splus). Since \( \tilde{Z}_2 \) represents in a way the closest point to \( \bar{Z}_2 \) under the null, we take the vector \( \max(0, \bar{Z}_2) \) as starting point for the numerical optimisation routine. This initial value satisfies the boundary restrictions. The minimum value of the function is then found to be \( \hat{\xi}_2 = 2.096 \).

According to the bounds given in Kodde and Palm (1986), the null hypothesis of a greater concordance of the fitted distribution cannot be rejected at any level lower than 5%. This indicates that the amount of positive dependence expressed by the parametric framework is at least as large as that suggested by the data. This is particularly appealing to actuaries since it ensures that most actuarial quantities computed in the Gumbel-Pareto model will not be underestimated.

It can also be of interest to test the concordance behaviour only in the upper tails. We consider a 81 grid formed by the percentiles in \( \{0.91, 0.92, \ldots, 0.99\} \times \{0.91, 0.92, \ldots, 0.99\} \). See Figure 7.3 for a representation. In this case only 10 components of \( \tilde{Z}_2 \) are found to be negative and the minimum value of the function is \( \hat{\xi}_2 = 0.000035 \). Thus, again the null hypothesis cannot be rejected.

### 7.3 Inference under semiparametric specification

In this section we wish to test the same type of hypothesis than in the previous subsection but using the semiparametric approach. We thus drop the Pareto modelling of the marginals and leave them unspecified. Note however the similarity between the parametric and empirical estimations of both margins in Figure 7.4. This explains why the estimated value \( \tilde{\theta} = 1.415 \) of the Gumbel copula is not much affected.

Again we perform inference based on probability levels and we use the same grid \( \{0.1, 0.2, \ldots, 0.9\} \times \{0.1, 0.2, \ldots, 0.9\} \). Figure 7.5 displays the differences between the semiparametric and empirical estimations and the semiparametric and parametric estimations of the copula on our data. Note again the small difference between the semiparametric and parametric estimations. We thus expect to get the same conclusion under the semiparametric framework as under the parametric one.

The minimum value of the function is now found to be \( \hat{\xi}_4 = 14.161 \). Since this value does not allow us to get a conclusion about its significance using the bounds of Kodde and Palm (1986), we need to rely on the simple Monte Carlo technique described in Section 6. A \( p \)-value equal to 0.98 has been obtained which clearly yields to not reject the null hypothesis of concordance.

Besides the results about the concordance behaviour in the upper tails are equivalent to the ones from the parametric approach using the same grid. Differences between the
Figure 7.3: Estimated copulas in the upper tails: Gumbel (top), empirical (mid) and difference (bottom).
empirical and the semiparametric estimations (left) and between the semiparametric and the parametric estimations (right) are shown in Figure 7.6. Only 12 components of $\hat{Z}_4$ are negative and the minimum value of the function is $\hat{\xi}_4 = 0.000036$, which does not allow to reject the null hypothesis.

8 Concluding remarks

In this paper we have analysed simple distributional free inference for concordance ordering. The testing procedures have proven to be empirically relevant to the analysis of dependencies among US insurance claim data. In particular they suggest that the Gumbel copula reflects the dependence structure in the data safely enough. This should reassure actuaries in their use of this copula when computing an insurance premium.
Figure 7.5: Differences between estimated copulas: semiparametric - empirical (left) and semiparametric - parametric (right).

Figure 7.6: Differences between estimated copulas in the upper tails: semiparametric - empirical (left) and semiparametric - parametric (right) in the tails.
APPENDICES

A Proof of Proposition 3.1

Let us recall that given two r.v’s $Z_1$ and $Z_2$, $Z_1 \preceq_{cv} Z_2$ holds if, and only if, $EZ_1 = EZ_2$ and $E(x - Z_1)_+ \geq E(x - Z_1)_+$ is valid for all $x \in \mathbb{R}$. Now, note that

$$\int_{-\infty}^{x} P[Y_1 + Y_2 \leq t] dt = \left[ tP[Y_1 + Y_2 \leq t] \right]_{-\infty}^{x} - \int_{-\infty}^{x} tdP[Y_1 + Y_2 \leq t] = E(x - Y_1 - Y_2)_+.$$  

So, we want to show that the inequality $E(x - Y_1 - Y_2)_+ \geq E(x - X_1 - X_2)_+$ holds for any real constant $x$ when $X \preceq_c Y$. Now, let us express $E(x - Y_1 - Y_2)_+$ in terms of the joint cdf of $Y$. Note that

$$\int_{-\infty}^{x} \mathbb{I}[y_1 \leq t, y_2 \leq x-t] dt = \int_{-\infty}^{x} \mathbb{I}[y_1 \leq t \leq x-y_2] dt = (x - y_1 - y_2)_+$$

whence it follows that

$$E(x - Y_1 - Y_2)_+ = \int_{-\infty}^{x} P[Y_1 \leq t, Y_2 \leq x-t] dt.$$  

Finally,

$$E(x - Y_1 - Y_2)_+ - E(x - X_1 - X_2)_+ = \int_{-\infty}^{x} \left\{ P[Y_1 \leq t, Y_2 \leq x-t] - P[X_1 \leq t, X_2 \leq x-t] \right\} dt$$

where the integrand $\ldots$ is non-negative provided $X \preceq_c Y$, which ends the proof.

B Asymptotic distributions

We first derive the asymptotic distribution of the parametric estimator $\hat{\nu} = (\hat{\beta}', \hat{\theta}')'$ and the semiparametric estimator $\hat{\theta}$ in a misspecified framework. For the well specified case the results can be found in Genest, Ghoudi and Rivest (1995) and Shih and Louis (1995). Then we proceed with the asymptotic distribution of the various difference vectors $\hat{D}_k$ and $\hat{\bar{D}}_k$, $k = 1, ..., 4$.

B.1 Asymptotic distribution of the parametric estimator

The asymptotic distribution of $\hat{\nu}$ immediately results from usual pseudo maximum likelihood theory (see e.g. White (1982), Gouriéroux, Monfort and Trognon (1984)). Indeed from a standard Taylor expansion of the first order condition of the maximum likelihood criterion and the law of large numbers, we get:

$$\sqrt{T} (\hat{\nu} - \nu_0) = J_{\nu_0}^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial}{\partial \nu} \log f(Y_t; \nu_0) + o_p(1).$$
with

\[ J_{\nu_0} = E_0 \left[ -\frac{\partial^2}{\partial \nu \partial \nu'} \log f(Y; \nu_0) \right], \]

where \( E_0 \) denotes expectation w.r.t. the true distribution \( F_0 \), and by application of the central limit theorem

\[ \sqrt{T} (\tilde{\nu} - \nu_0) \overset{d}{\rightarrow} N(0, J_{\nu_0}^{-1} I_{\nu_0} J_{\nu_0}^{-1}), \]

where

\[ I_{\nu_0} = E_0 \left[ \frac{\partial}{\partial \nu} \log f(Y; \nu_0) \frac{\partial}{\partial \nu'} \log f(Y; \nu_0) \right]. \]

When the parametric model is well specified, i.e. \( F(\cdot, \nu_0) = F_0(\cdot) \), we have \( J_{\nu_0} = I_{\nu_0} \).

### B.2 Asymptotic distribution of the semiparametric estimator

From a Taylor expansion of the first order condition of the maximum likelihood criterion and the law of large numbers, we get:

\[ \sqrt{T} (\tilde{\theta} - \theta_0^*) = J_{\theta_0^*}^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial}{\partial \theta} \log c(\hat{F}_1(Y_{1t}), ..., \hat{F}_n(Y_{nt}); \theta_0^*) + o_p(1), \] (B.1)

where

\[ J_{\theta_0^*} = E_0 \left[ -\frac{\partial^2}{\partial \theta \partial \theta'} \log c(F_{10}(Y_1), ..., F_{n0}(Y_n); \theta_0^*) \right]. \]

The random part of the right hand side in (B.1) can be written:

\[ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial}{\partial \theta} \log c(\hat{F}_1(Y_{1t}), ..., \hat{F}_n(Y_{nt}); \theta_0^*) \]

\[ = \sqrt{T} \int_{\mathbb{R}^n} \frac{\partial}{\partial \theta} \log c(\hat{F}_1(z_1), ..., \hat{F}_n(z_n); \theta_0^*) d\hat{F}(z_1, ..., z_n), \]

and decomposed into three terms:

\[ \sqrt{T} \int_{\mathbb{R}^n} \frac{\partial}{\partial \theta} \log c(\hat{F}_1(z_1), ..., \hat{F}_n(z_n); \theta_0^*) d\hat{F}(z_1, ..., z_n) \]

\[ = \sqrt{T} \int_{\mathbb{R}^n} \frac{\partial}{\partial \theta} \log c(\hat{F}_1(z_1), ..., \hat{F}_n(z_n); \theta_0^*) dF_0(z_1, ..., z_n) \]

\[ + \sqrt{T} \int_{\mathbb{R}^n} \left[ \frac{\partial}{\partial \theta} \log c(\hat{F}_1(z_1), ..., \hat{F}_n(z_n); \theta_0^*) - \frac{\partial}{\partial \theta} \log c(F_{10}(z_1), ..., F_{n0}(z_n); \theta_0^*) \right] \]

\[ \times d[\hat{F}(z_1, ..., z_n) - F_0(z_1, ..., z_n)] \]

\[ + \sqrt{T} \int_{\mathbb{R}^n} \frac{\partial}{\partial \theta} \log c(F_{10}(z_1), ..., F_{n0}(z_n); \theta_0^*) d[\hat{F}(z_1, ..., z_n) - F_0(z_1, ..., z_n)]. \]
The second term converges to zero, and by the central limit theorem the third term converges to $N(0, I_{\theta_0}^*)$, where

$$I_{\theta_5} = E_0 \left[ \frac{\partial}{\partial \theta} \log c(F_1(Y_1), ..., F_n(Y_n); \theta_0^*) \frac{\partial}{\partial \theta} \log c(F_1(Y_1), ..., F_n(Y_n); \theta_0^*) \right].$$

Now a Taylor expansion of the first term (see Serfling (1980) for expansion of von Mises differentiable statistical functions) leads to

$$\sqrt{T} \int_{\mathbb{R}^n} \frac{\partial}{\partial \theta} \log c(\hat{F}(z_1), ..., \hat{F}_n(z_n); \theta_0^*) dF_0(z_1, ..., z_n)$$

$$= \sqrt{T} \int_{\mathbb{R}^n} \frac{\partial}{\partial \theta} \log c(F_{10}(z_1), ..., F_{n0}(z_n); \theta_0^*) dF_0(z_1, ..., z_n)$$

$$+ \sqrt{T} \sum_{j=1}^n \int_{\mathbb{R}^n} \frac{\partial^2}{\partial \theta \partial u_j} \log c(F_{10}(z_1), ..., F_{n0}(z_n); \theta_0^*) dF_0(z_1, ..., z_n)$$

$$\times (\hat{F}_j(z_j) - F_{j0}(z_j)) + o_p(1)$$

$$= 0 + \sqrt{T} \sum_{j=1}^n \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \Psi[w_j \leq z_j] \frac{\partial^2}{\partial \theta \partial u_j} \log c(F_{10}(z_1), ..., F_{n0}(z_n); \theta_0^*) dF_0(z_1, ..., z_n)$$

$$\times d[\hat{F}_j(w_j) - F_{j0}(w_j)] + o_p(1).$$

Hence by the central limit theorem the first term converges to $N(0, M_{\theta_0}^*)$, where

$$M_{\theta_0}^* = \text{Var}_0 \left[ \sum_{j=1}^n \int_{\mathbb{R}^n} \Psi[Y_j \leq z_j] \frac{\partial^2}{\partial \theta \partial u_j} \log c(F_{10}(z_1), ..., F_{n0}(z_n); \theta_0^*) dF_0(z_1, ..., z_n) \right].$$

Since the conditional expectation of $\frac{\partial}{\partial \theta} \log c(F_{10}(Y_1), ..., F_{n0}(Y_n); \theta_0^*)$ w.r.t. any $Y_j$ is null, the first term and the third term are uncorrelated, and we finally get:

$$\sqrt{T}(\hat{\theta} - \theta_0^*) \Rightarrow N(0, J_{\theta_0}^{-1}(I_{\theta_5}^* + M_{\theta_0}^*).J_{\theta_0}^{-1}).$$

When the parametric model is well specified, i.e. $F(\cdot; \theta_0^*) = F_0(\cdot)$, we have $J_{\theta_5} = I_{\theta_5^*}$.

### B.3 Asymptotic distribution of the difference vectors

#### B.3.1 Asymptotic distribution of $\hat{D}_1$ and $\hat{D}_1$

A Taylor expansion of $C(F_1(y_{i1}; \hat{\beta}_1), ..., F_n(y_{in}; \hat{\beta}_n); \hat{\theta})$ around $\nu_0 = (\beta_0^*, \theta_0')$ gives:

$$C(F_1(y_{i1}; \hat{\beta}_1), ..., F_n(y_{in}; \hat{\beta}_n); \hat{\theta})$$

$$= C(F_1(y_{i1}; \beta_{10}), ..., F_n(y_{in}; \beta_{n0}); \theta_0) + \frac{\partial}{\partial \nu} C(F_1(y_{i1}; \hat{\beta}_{10}), ..., F_n(y_{in}; \hat{\beta}_{n0}); \hat{\theta}_0)(\nu - \nu_0) + o_p(T^{-1/2}),$$
where the overline stands for some mean values.

Hence we deduce from Subsection A:

\[
\sqrt{T}(\hat{D}_1^i - D_1^i) = \nabla^i_{\nu_0} C^J J_{\nu_0}^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial}{\partial \nu} \log f(Y_t; \nu_0) + \sqrt{T}(\hat{F}(y_i) - F(y_i)) + o_p(1),
\]

where

\[
\nabla^i_{\nu_0} C = \frac{\partial}{\partial \nu} C(F_1(y_{i1}; \beta_{i10}), ..., F_n(y_{in}; \beta_{i0}); \theta_0).
\]

An application of the central limit theorem delivers:

\[
\sqrt{T}(\hat{D}_1^i - D_1^i) \implies N(0, B^i_{\nu_0} \text{Cov}_0 [S^i_{\nu_0}, S^i_{\nu_0}] B^i_{\nu_0}),
\]

where

\[
B^i_{\nu_0} = (\nabla^i_{\nu_0} C^J J_{\nu_0}^{-1}, 1)',
\]

and

\[
S^i_{\nu_0} = \left(\frac{\partial}{\partial \nu} \log f(Y; \nu_0), \mathbb{I}[Y \leq y_i]\right)',
\]

while

\[
\lim_{T \to \infty} T \text{Cov}_0 \left[\hat{D}_1^k, \hat{D}_1^l\right] = B^{k}_{\nu_0} \text{Cov}_0 \left[S^{k}_{\nu_0}, S^{l}_{\nu_0}\right] B^{l}_{\nu_0}, \quad k, l = 1, ..., d.
\]

Similarly we may deduce for the survival quantity:

\[
\sqrt{T}(\hat{\hat{D}}_1^i - D_1^i) \implies N(0, \hat{B}^i_{\nu_0} \text{Cov}_0 [\hat{S}^i_{\nu_0}, \hat{S}^i_{\nu_0}] \hat{B}^i_{\nu_0}),
\]

where

\[
\hat{B}^i_{\nu_0} = (\nabla^i_{\nu_0} \hat{C}^J J_{\nu_0}^{-1}, 1)',
\]

and

\[
\hat{S}^i_{\nu_0} = \left(\frac{\partial}{\partial \nu} \log f(Y; \nu_0), \mathbb{I}[Y > y_i]\right)',
\]

while

\[
\lim_{T \to \infty} T \text{Cov}_0 \left[\hat{\hat{D}}_1^k, \hat{\hat{D}}_1^l\right] = \hat{B}^{k}_{\nu_0} \text{Cov}_0 \left[\hat{S}^{k}_{\nu_0}, \hat{S}^{l}_{\nu_0}\right] \hat{B}^{l}_{\nu_0}, \quad k, l = 1, ..., d.
\]

Finally we also have:

\[
\lim_{T \to \infty} T \text{Cov}_0 \left[\hat{D}_1^k, \hat{D}_1^l\right] = B^{k}_{\nu_0} \text{Cov}_0 \left[S^{k}_{\nu_0}, S^{l}_{\nu_0}\right] B^{l}_{\nu_0}, \quad k, l = 1, ..., d.
\]

**B.3.2 Asymptotic distribution of \( \hat{D}_2 \) and \( \hat{\hat{D}}_2 \)**

A Taylor expansion of \( C(u_i; \hat{\theta}) \) around \( \theta_0 \) gives:

\[
C(u_i; \hat{\theta}) = C(u_i; \theta_0) + \frac{\partial}{\partial \theta} C(u_i; \bar{\theta}_0)'(\hat{\theta} - \theta_0) + o_p(T^{-1/2}),
\]

where \( \bar{\theta}_0 \) lies between \( \hat{\theta} \) and \( \theta_0 \).
We get using Subsection A:

\[ \sqrt{T}(C(u_i; \hat{\theta}) - C(u_i; \theta_0)) = \nabla_{\theta_0}^i C' J_{\theta_0}^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial}{\partial \theta} \log f(Y_t; \nu_0) + o_p(1), \]

where

\[ \nabla_{\theta_0}^i C = \frac{\partial}{\partial \theta} C(u_i; \theta_0), \quad J_{\theta_0} = E_0 \left[ -\frac{\partial^2}{\partial \theta \partial \theta'} \log f(Y; \nu_0) \right]. \]

Furthermore let \( M = \{ [\cdot \leq x_1] \ldots [\cdot \leq x_n] : x_j \in \mathbb{R}, j = 1, \ldots, n \}. \) Since \( M \) satisfies Pollard’s entropy condition for some finite constant taken as envelope, the sequence

\[ \left\{ \hat{F}(x) = T^{-1} \sum_{t=1}^{T} \prod_{j=1}^{n} \mathbb{I}[Y_{jt} \leq x_j] : T \geq 1 \right\} \]

is stochastically differentiable at \( \zeta_i \) with random derivative \((d \times 1)\)-vector \( \hat{D} \hat{F}(\zeta_i) \) (see e.g. POLLARD (1985), ANDREWS (1994,1999) for definition, use and check of stochastic differentiability). It means that we have the approximation:

\[ \hat{F}(\zeta_i) = \hat{F}^{\prime}(\zeta_i) + D \hat{F}(\zeta_i) \prime (\zeta_i - \hat{\zeta}_i) + o_p(T^{-1/2}), \]

where \( \hat{\zeta}_i \) is a mean value located between \( \zeta_i \) and \( \zeta_i \).

Similarly we get the approximations:

\[ \hat{F}_i(\zeta_{ij}) = \hat{F}_i(\zeta_{ij}) + D \hat{F}_i(\zeta_{ij}) \prime (\zeta_{ij} - \hat{\zeta}_{ij}) + o_p(T^{-1/2}). \]

Combining these approximations and using \( F_{jo}(\zeta_{ij}) = u_{ij} = \hat{F}_j(\zeta_{ij}) \) leads to

\[ \hat{F}(\zeta_i) = \hat{F}(\zeta_i) - D \hat{F}(\zeta_i) \prime \text{diag } S_i + o_p(T^{-1/2}), \]

where \( S_i \) is the stack of \((\hat{F}_j(\zeta_{ij}) - u_{ij}) / D \hat{F}_j(\zeta_{ij}), j = 1, \ldots, n, \) and \( \text{diag } S_i \) is the diagonal matrix built from this stack.

Hence we get:

\[ \sqrt{T}(\hat{D}_i - D_i^i) \]

\[ = \nabla_{\theta_0}^i C' J_{\theta_0}^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial}{\partial \theta} \log f(Y_t; \nu_0) - \sqrt{T}(\hat{F}(\zeta_i) - F_0(\zeta_i)) + D \hat{F}(\zeta_i) \prime \sqrt{T} \text{diag } S_i + o_p(1). \]

An application of the central limit theorem delivers:

\[ \sqrt{T}(\hat{D}_i - D_i^i) \Rightarrow N(0, B_{\theta_0}^i \text{Cov}_0 [S_{\theta_0}^i, S_{\theta_0}^i] B_{\theta_0}^i), \]

where

\[ B_{\theta_0}^i = \left( \nabla_{\theta_0}^i C' J_{\theta_0}^{-1}, -1, \frac{\partial F_0(\zeta_i)}{\partial x_1}, \ldots, \frac{\partial F_0(\zeta_i)}{\partial x_n} \right)' \]

25
and
\[ S_{\theta_0}^i = \left( \frac{\partial}{\partial \theta} \log f(Y; \nu_0), \mathbb{I}[Y \leq \zeta_i], \mathbb{I}[Y_1 \leq \zeta_{i1}], \ldots, \mathbb{I}[Y_n \leq \zeta_{in}] \right), \]

while
\[
\lim_{T \to \infty} T \text{Cov}_0 \left[ \hat{D}_2^k, \hat{D}_2^l \right] = B_{\theta_0}^{k\ell} \text{Cov}_0 \left[ S_{\theta_0}^i, S_{\theta_0}^l \right] B_{\theta_0}^l, \quad k, l = 1, \ldots, d.
\]

We also get:
\[
\sqrt{T}(\hat{D}_2 - D_2) \rightarrow N(0, B_{\theta_0}^{ij} \text{Cov}_0 \left[ S_{\theta_0}^i, S_{\theta_0}^j \right] B_{\theta_0}^j),
\]

where
\[
\hat{B}_{\theta_0}^i = \left( \nabla_{\theta_0}^i C', J_{\theta_0}^{-1}, -1, \frac{\partial f_0(\zeta_1)}{\partial x_1}, \ldots, \frac{\partial f_0(\zeta_n)}{\partial x_n} \right)'
\]

and
\[
\hat{S}_{\theta_0}^i = \left( \frac{\partial}{\partial \theta} \log f(Y; \nu_0), \mathbb{I}[Y > \zeta_i], \mathbb{I}[Y_1 \leq \zeta_{i1}], \ldots, \mathbb{I}[Y_n \leq \zeta_{in}] \right)'.
\]

while
\[
\lim_{T \to \infty} T \text{Cov}_0 \left[ \hat{D}_2^k, \hat{D}_2^l \right] = \hat{B}_{\theta_0}^{k\ell} \text{Cov}_0 \left[ \hat{S}_{\theta_0}^i, \hat{S}_{\theta_0}^j \right] \hat{B}_{\theta_0}^l, \quad k, l = 1, \ldots, d,
\]

and
\[
\lim_{T \to \infty} T \text{Cov}_0 \left[ \hat{D}_2^k, \hat{D}_2^l \right] = B_{\theta_0}^{k\ell} \text{Cov}_0 \left[ S_{\theta_0}^i, S_{\theta_0}^j \right] B_{\theta_0}^l, \quad k, l = 1, \ldots, d.
\]

\[ \text{B.3.3 \ Asymptotic distribution of } \hat{D}_3 \text{ and } \hat{D}_3 \]

Using a Taylor expansion of \( C(\hat{F}_1(y_{i1}), \ldots, \hat{F}_n(y_{in}); \hat{\theta}_0) \) around \( \theta_0^* \) and \( F_{j0}(y_{ij}), j = 1, \ldots, n, \) we get:
\[
\sqrt{T}(\hat{D}_3 - D_3) = \nabla_{\theta_0}^i C' \sqrt{T}(\hat{\theta}_0 - \theta_0^*) + \sum_{j=1}^{n} \nabla_{u_j}^i C' \sqrt{T}(\hat{F}_j(y_{ij}) - F_{j0}(y_{ij})) - \sqrt{T}(\hat{F}(y_i) - F_0(y_i)) + o_p(1),
\]

where
\[
\nabla_{\theta_0}^i C = \frac{\partial}{\partial \theta} C(F_{10}(y_{i1}), \ldots, F_{n0}(y_{in}); \theta_0^*), \quad \nabla_{u_j}^i C = \frac{\partial}{\partial u_j} C(F_{10}(y_{i1}), \ldots, F_{n0}(y_{in}); \theta_0^*).
\]

From Subsection B and the central limit theorem, we deduce
\[
\sqrt{T}(\hat{D}_3 - D_3) \rightarrow N(0, B_{\theta_0}^{ii} \text{Cov}_0 \left[ S_{\theta_0}^i, S_{\theta_0}^j \right] B_{\theta_0}^l),
\]

where
\[
B_{\theta_0}^{ii} = \left( \nabla_{\theta_0}^i C', J_{\theta_0}^{-1}, \nabla_{u_1}^i C, \ldots, \nabla_{u_n}^i C, -1 \right)'
\]

and
\[
S_{\theta_0}^i = \left( U_{\theta_0}^i, \mathbb{I}[Y_1 \leq y_{i1}], \ldots, \mathbb{I}[Y_n \leq y_{in}], \mathbb{I}[Y \leq y_i] \right)'
\]

26
and
\[ U_{\theta_0} = \frac{\partial}{\partial \theta} \log c(F_{10}(Y_1), \ldots, F_{n0}(Y_n); \theta_0^*) + \sum_{j=1}^{n} \int_{\mathbb{R}^n} \mathbb{I}[Y_j \leq z_j] \frac{\partial^2}{\partial \theta \partial u_j} \log c(F_{10}(z_1), \ldots, F_{n0}(z_n); \theta_0^*) dF_0(z_1, \ldots, z_n). \]

We also have:
\[ \lim_{T \to \infty} T \text{Cov}_0 \begin{bmatrix} \hat{D}_3^k, \hat{D}_3^l \end{bmatrix} = B_{\theta_0}^{ik} \text{Cov}_0 \begin{bmatrix} \bar{S}_{\theta_0}^k, \bar{S}_{\theta_0}^l \end{bmatrix} \bar{B}_{\theta_0}^l, \quad k, l = 1, \ldots, d. \]

Besides,
\[
\sqrt{T}(\hat{D}_3 - \bar{D}_3) \Rightarrow N(0, \bar{B}_{\theta_0}^i \text{Cov}_0 \begin{bmatrix} \bar{S}_{\theta_0}^i, \bar{S}_{\theta_0}^i \end{bmatrix} \bar{B}_{\theta_0}^i),
\]
where
\[
\bar{B}_{\theta_0}^i = \left( \nabla_{\theta_0^i} C' J_{\theta_0^i}^{-1}, \nabla_{u_1} C, \ldots, \nabla_{u_n} C, -1 \right)',
\]
and
\[
\bar{S}_{\theta_0}^i = \left( U_{\theta_0}^i, \mathbb{I}[Y_1 \leq y_{i1}], \ldots, \mathbb{I}[Y_n \leq y_{in}], \mathbb{I}[Y > y_i] \right)'.
\]

while
\[
\lim_{T \to \infty} T \text{Cov}_0 \begin{bmatrix} \hat{D}_3, \hat{D}_3 \end{bmatrix} = B_{0}^{ik} \text{Cov}_0 \begin{bmatrix} S_{\theta_0}^k, S_{\theta_0}^l \end{bmatrix} B_{\theta_0}^l, \quad k, l = 1, \ldots, d,
\]
and
\[
\lim_{T \to \infty} T \text{Cov}_0 \begin{bmatrix} \hat{D}_3, \hat{D}_3 \end{bmatrix} = B_{\theta_0}^{ik} \text{Cov}_0 \begin{bmatrix} S_{\theta_0}^k, S_{\theta_0}^l \end{bmatrix} B_{\theta_0}^l, \quad k, l = 1, \ldots, d.
\]

### B.3.4 Asymptotic distribution of \( \hat{D}_4 \) and \( \hat{D}_4 \)

The only difference between \( \hat{D}_2 \) and \( \hat{D}_4 \) lies in the replacement of the parametric estimator \( \hat{\theta} \) by the semiparametric estimator \( \hat{\theta} \). Hence the asymptotic distribution of \( \hat{D}_4 \) is obtained after substituting \( \nabla_{\theta_0^i} C' J_{\theta_0^i}^{-1} \) for \( \nabla_{\theta_0^i} C' J_{\theta_0^i}^{-1} \), and \( U_{\theta_0} \) for \( \frac{\partial}{\partial \theta} \log f(Y; \theta_0) \) in the asymptotic results for \( \hat{D}_2 \).

Similarly in order to derive the asymptotic normality of \( \hat{D}_4 \), we only have to replace \( \nabla_{\theta_0^i} C' J_{\theta_0^i}^{-1} \) with \( \nabla_{\theta_0^i} C' J_{\theta_0^i}^{-1} \), and \( \frac{\partial}{\partial \theta} \log f(Y; \theta_0) \) with \( U_{\theta_0} \) in the asymptotic results for \( \hat{D}_2 \).

### Acknowledgements

The authors would like to thank Professors Frees and Valdez for kindly providing the Loss-ALAE data, which were collected by the US Insurance Services Office (ISO). Ana Cebrián thanks the Université catholique de Louvain, Louvain-la-Neuve, Belgium, for financial support through an FSR research grant. Michel Denuit acknowledges financial support of the Belgian Government under “Projet d’Actions de Recherche Concertées” (No. 98/03-217).
Olivier Scaillet receives support by the Swiss National Science Foundation through the National Center of Competence: Financial Valuation and Risk Management. Part of this research was done when he was visiting THEMA and IRES.

REFERENCES:


