Tail negative dependence and its applications for aggregate loss modeling

Lei Hua

Division of Statistics

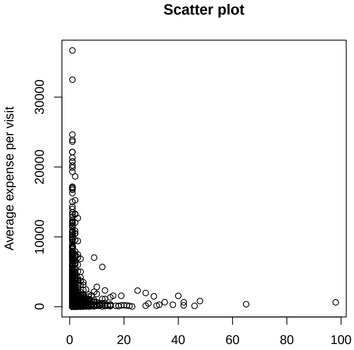
Northern Illinois University

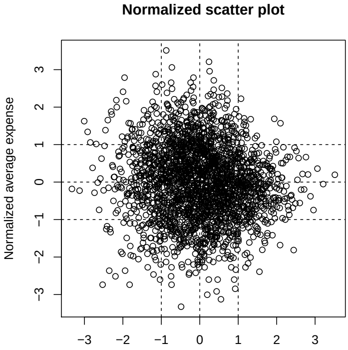
Oct 20, 2014, ISU

Aggregate loss modeling [3]

Modeling aggregate medical expenditures per year by considering:

- Loss frequency: number of outpatient department visits to physicians.
- Loss severity: average facility expenses per visit.
- Covariates: age, gender, race, income, insurance coverage, etc.





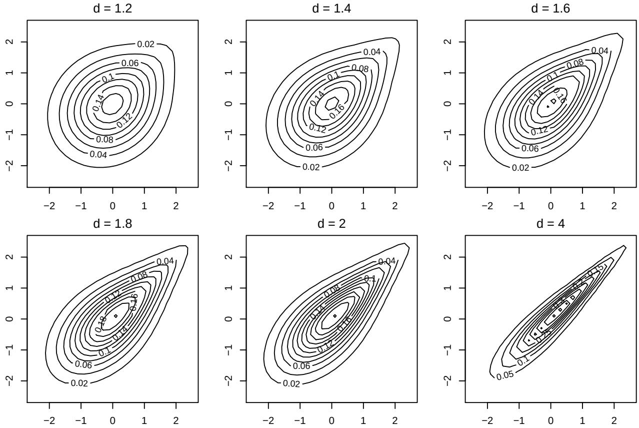
Aggregate loss modeling (cont'd)

- Zeros were removed to better illustrate dependence structures.
- A binary regression model, say, Logistic regression can be used first for loss frequency.
- Challenges:
  - regression on loss frequency and loss severity together while accounting for the dependence structure between them.
  - no suitable dependence models for the special dependence pattern: negative dependence in the upper tail, and independence in the lower tail.

Tail dependence - Gumbel copula

$$\overline{C}(1-u, 1-u) \sim {\color{red}\lambda} u, \quad u \to 0^+; \qquad 0 < \lambda \leq 1.$$

$$\lim_{u \to 1^-} \mathbb{P}[U_1 > u | U_2 > u] = \lim_{u \to 1^-} \mathbb{P}[U_2 > u | U_1 > u] = {\color{red}\lambda}.$$



From tail dependence to tail order

Tail dependence:

$$\overline{C}(1-u,1-u) \sim \lambda u, \quad u \to 0^+; \qquad 0 < \lambda \leq 1.$$

Tail order:

$$\overline{C}(1-u,1-u) \sim u^{\kappa}\ell(u), \quad u \to 0^+; \qquad 1 \leq \kappa.$$

Smaller tail order $\Rightarrow$ stronger dependence in the tails.

- $\kappa = 1$: usual tail dependence;
- $1 < \kappa < 2$: intermediate tail dependence;
- $\kappa = 2$: tail quadrant independence;
- $\kappa > 2$: tail negative dependence.

Upper and lower tails can be quantified separately.

Example - Gaussian copula

Bivariate Gaussian copula:

$$C_{\Phi_2}(u_1, u_2) = \Phi_2(\Phi^{-1}(u_1), \Phi^{-1}(u_2); \Sigma),$$

where $\Phi_2(\cdot; \Sigma)$ is the joint cumulative distribution function (cdf) of a standard bivariate Gaussian random vector with positive definite correlation matrix Σ.

$$\kappa = 2/(1+\rho);$$
$$\ell(u) = (-\log u)^{-\rho/(1+\rho)}.$$

Elliptical copula

let $\mathbf{X} := (X_1, X_2)$ be an elliptical random vector such that

$$\mathbf{X} \stackrel{d}{=} RA\mathbf{U}, \tag{1}$$

where $R \geq 0$ is independent of $\mathbf{U}$, $\mathbf{U}$ is uniformly distributed on the surface of the unit hypersphere $\{\mathbf{z} \in \mathbb{R}^2 | \mathbf{z}^{\mathrm{T}}\mathbf{z} = 1\}$,
$AA^{\mathrm{T}} = \Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, $\mathbf{X} \sim F_2$ and $X_1, X_2 \sim F$.

$$C(u_1, u_2) = F_2(F^{-1}(u_1), F^{-1}(u_2); \Sigma).$$

Univariate tail heaviness

- Need to quantify the tail heaviness of R.
- Maximum Domain of Attraction (MDA)
  (Fisher-Tippett theorem) Let (X_n) be a sequence of iid random variables. If there exist constants $c_n > 0$, $d_n \in \mathbb{R}$ and some non-degenerate H such that

$$\frac{\max_n\{X_1, \ldots, X_n\} - d_n}{c_n} \to_d H,$$

  then H is one of the following distributions: Fréchet (Φ_α), Gumbel (Λ) and Weibull (Ψ_α).

Elliptical copula (cont's)

Let C be the copula for an elliptical random vector $\mathbf{X} := (X_1, X_2)$ constructed as (1), and $b_\rho := \sqrt{2/(1+\rho)}$.

- If $R \in \mathrm{MDA}(\Lambda)$, then the upper and lower tail orders of C is

$$\kappa = \lim_{r\to\infty} \frac{\log(1 - F_R(b_\rho r))}{\log(1 - F_R(r))},$$

  provided that the limit exists.

- If $R \in \mathrm{MDA}(\Phi_\alpha)$, then $\kappa = 1$.

Elliptical copula (cont's)

- Example: Bivariate symmetric Kotz type copula
  The density generator

$$g(x) = Kx^{N-1} \exp\{-\beta x^{\xi}\}, \quad \beta, \xi, N > 0,$$

where K is a normalizing constant. The density function of R is $f_R(x) = 2\pi x g(x^2) = 2K\pi x^{2N-1} \exp\{-\beta x^{2\xi}\}$. Therefore, the tail order for the symmetric Kotz type copula is

$$\kappa = [2/(1+\varrho)]^{\xi}.$$

Gaussian copula belongs to this class with $N = 1, \beta = 1/2$ and $\xi = 1$, so its tail order is $2/(1+\varrho)$.

Extreme value copula

If a copula C satisfies $C(u_1^t, \dots, u_d^t) = C^t(u_1, \dots, u_d)$ for any $(u_1, \dots, u_d) \in [0,1]^d$ and $t > 0$, then we refer to C as an extreme value copula, and

$$C(u_1, \dots, u_d) = \exp\{-A(-\log u_1, \dots, -\log u_d)\}.$$

- Either upper tail dependence (eg: Gumbel copula, Galambos copula) or independence.
- Bivariate cases: $\kappa_U = 1$ with $\lambda_U = 2 - A(1,1)$, and $\kappa_L = A(1,1)$.

Archimedean copula - Laplace Transform

A LT-Archimedean copula

$$C_\psi(u_1, \ldots, u_d) = \psi(\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_d)) \tag{2}$$

has a mixture representation with LT ψ. That is,

$$C_\psi(u_1, \ldots, u_d) = \int_0^\infty \prod_{j=1}^d G^\eta(u_j) dF_Y(\eta),$$

where F_Y is the cdf of the resilience random variable Y, $G(u) = \exp\{-\psi^{-1}(u)\}$ $(0 \le u \le 1)$ are certain cumulative distribution functions, and $\psi(s) = \psi_Y(s) = \int_0^\infty e^{-s\eta} dF_Y(\eta)$ is the Laplace Transform (LT) of Y.

Archimedean copula - Laplace Transform (cont's)

- For a positive random variable Y with LT ψ, the order of the maximal non-negative moment is

$$M_Y = M_\psi = \sup\{m \geq 0 : \mathbb{E}(Y^m) < \infty\}.$$

- Let ψ be the LT of a positive random variable Y, under some regularity conditions with $1 \leq M_\psi \leq d$, the Archimedean copula C_ψ has upper tail order $\kappa_U = M_\psi$.
- One limitation of LT-Archimedean copula: there is only positive dependence.

A new 1-parameter copula family

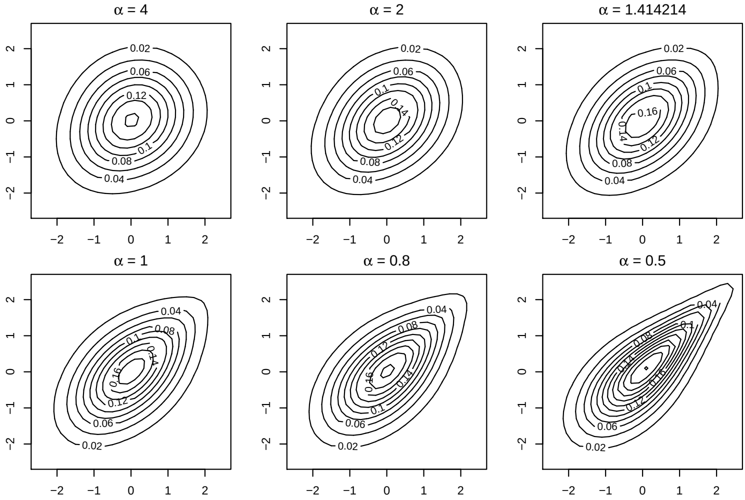
Archimedean family based on inverse Gamma LT (ACIG copula)

Let $Y = X^{-1}$ have the inverse Gamma (IΓ) distribution, where $X \sim \mathrm{Gamma}(\alpha, 1)$ for $\alpha > 0$. The LT of the inverse Gamma distribution:

$$\psi(s;\alpha) = \frac{2}{\Gamma(\alpha)} s^{\alpha/2} K_\alpha(2\sqrt{s}), \quad s \geq 0, \ \alpha > 0,$$

where K_α is the modified Bessel function of the second kind. This family is decreasing in concordance as α increases, with limits of the independence copula as $\alpha \to \infty$ and the comonotonic copula as $\alpha \to 0$. $\kappa_L = \sqrt{d}; \kappa_U = (d \wedge \alpha) \vee 1$.

Normalized contour plots of ACIG copula



Archimedean copula - Williamson d-Transform

Williamson d-transform of R is

$$\psi(s) = \int_s^\infty (1 - s/r)^{d-1} F_R(dr), \quad s \in [0, \infty).$$

The WT-Archimedean copula is the survival copula induced by the random vector

$$\mathbf{X} := (X_1, \ldots, X_d) \stackrel{d}{=} R \times (S_1, \ldots, S_d), \tag{3}$$

where R is a positive random variable and $(S_1, \ldots, S_d)$ is uniformly distributed on the simplex $\{\mathbf{x} \in \mathbb{R}_+^d : ||\mathbf{x}||_1 = 1\}$, and R and $(S_1, \ldots, S_d)$ are independent.

Tail negative dependence

- Tail negative dependence: tail order $\kappa > d$.
- For example, Gaussian copula with $\rho < 0$ has $\kappa = 2/(1+\rho) > 2$.
- Let a random vector $\mathbf{X} := (X_1, \ldots, X_d)$ be defined as (3). If $1/R \in \mathrm{MDA}(\Phi_\alpha)$, and $\mathbb{E}[1/R] < \infty$, then $\kappa_U = \alpha$.
- Let $\mathbf{X} := (X_1, \ldots, X_d)$ be defined as (3). If $1/R \in \mathrm{MDA}(\Lambda)$, then the upper tail order of the corresponding Archimedean copula is $\kappa_U = \infty$.

Tail negative dependence - IPS copula

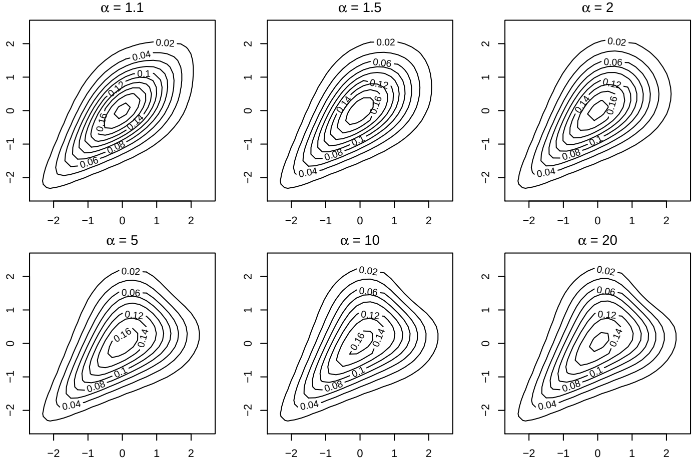
- Inverse-Pareto - Simplex copula, aka, IPS copula
  Let $X_i \stackrel{d}{=} RS_i, i = 1, 2$, (S_1, S_2) be uniformly distributed on $\{\mathbf{x} \geq \mathbf{0} : x_1 + x_2 = 1\}$, and $T := 1/R$ follow a Pareto distribution with cdf $F(x) = 1 - (1 + x)^{-\alpha}, x \geq 0$, $\alpha > 1$. Then the generator for the WT-Archimedean copula is

$$\psi(s) = \frac{s}{1-\alpha}\left[1 - (1 + 1/s)^{-\alpha+1}\right] + 1, \quad s \geq 0, \alpha > 1.$$

- $\kappa_U = \alpha > 1$ and $\kappa_L = 1$.

Tail negative dependence - IPS copula (cont's)

Normalized contour plots



Tail negative dependence - GGS copula

- Generalized-Gamma - Simplex mixture, aka, GGS copula
  Let $X_i \overset{d}{=} RS_i, i = 1, 2$, (S_1, S_2) be uniformly distributed on $\{\mathbf{x} \geq 0 : x_1 + x_2 = 1\}$, and $R^{1/\beta}$ follow a Gamma distribution with shape parameter α so that

$$F_R(x) = \frac{1}{\beta\Gamma(\alpha)} \int_0^x s^{\alpha/\beta-1} \exp\{-s^{1/\beta}\} ds \quad \alpha > 0, \beta > 0,$$

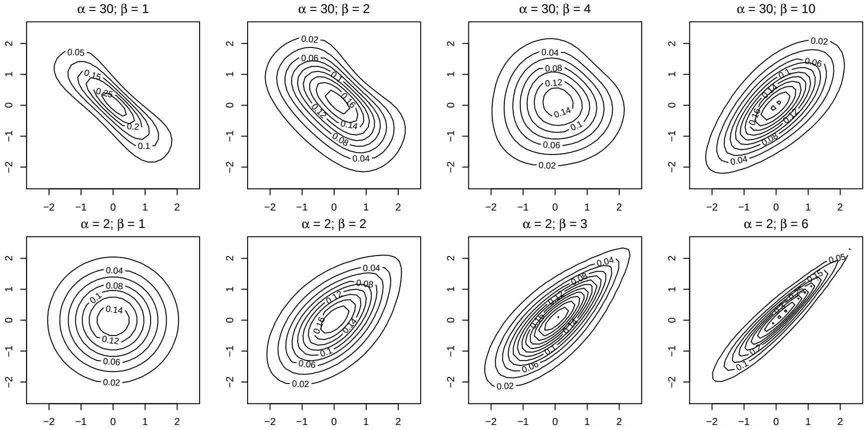
  and the Archimedean generator is

$$\psi(s) = \frac{1}{\Gamma(\alpha)} \left(\Gamma(\alpha, s^{1/\beta}) - s\Gamma(\alpha - \beta, s^{1/\beta}) \right),$$

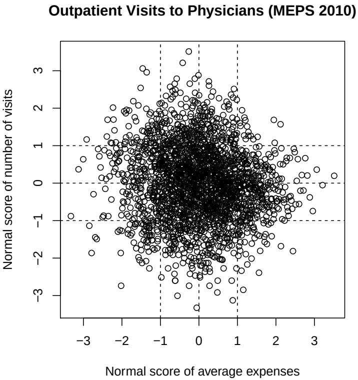
  where $\Gamma(\cdot, \cdot)$ is an upper incomplete gamma function.
- $\kappa_U = \max\{\alpha/\beta, 1\}$ and $\kappa_L = 2^{1/\beta}$.

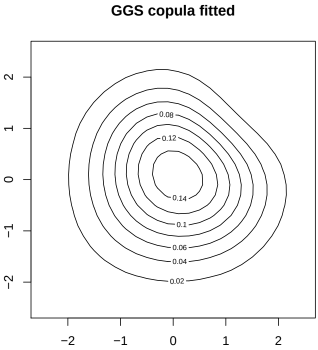
Tail negative dependence - GGS copula (cont's)

Normalized contour plots



Medical expenditure data with GGS copula





Applications in aggregate loss modeling

- Compound Poisson: $S = X_1 + \cdots + X_N$
- Tweedie regression: Poisson and Gamma mixtures
- Two-parts regression ([2]): Logistic regression + GLM for $X|N > 0$
- Mixed copula regression ([1]): copula for X and N

$$N|N > 0 : \text{Zero-truncated Poisson regression}$$
$$X|N > 0 : \text{Gamma regression}$$

Mixed copula regression with GGS copula

Dataset: Medical Expenditure Panel Survey, 2010
Responses: numbers of outpatient visits to physician (frequency)
average facility expenses (severity)
Covariates: ages, incomes, gender, education,
insurance coverage, races
Sample size: 2263

Model:

- frequency: Zipf $\sim$ age $+$ insurance coverages $+$ races
- severity: lognormal $\sim$ age $+$ insurance coverages $+$ races
- dependence: GGS copula, with homogeneous dependence parameters α, β: for both upper and lower tails.

Mixed copula regression with GGS copula (cont's)

	Min	1st quantile	Median	Mean	3rd quantile	Max
OPVEXP10	3	187	704	2373	2356	68370
Average Expense	3	132	406	1460	1573	36680
Age	0	29.5	50	46.44	64	85

Number of Visits	1	2	3	4	5	6	7	8	9	10
OPDRV10 (#obs)	1461	394	144	73	58	26	29	9	13	9
	11	12	13	14	15	16	17	18	19	20
	5	5	5	2	4	1	1	2	2	1
	21	22	23	25	28	29	31	32	33	35
	1	1	1	1	2	1	1	1	1	1
	38	40	42	46	48	65	98	-	-	-
	1	1	2	1	1	1	1	-	-	-

Insurance Coverage	Any Private (1)	Public Only (2)	Uninsured (3)
INSCOV10 (#obs)	1356	757	150

Race	Hispanic (1)	Black (2)	Asian (3)	Other (4)
RACETHNX (#obs)	433	440	91	1299

Marginal regression - Loss frequency

- Zipf's distribution can be written as

$$f_N(n|s,m) = \frac{n^{-s}}{\sum_{i=1}^{m} i^{-s}}, n = 1, 2, \dots, m, \quad s > 0,$$

where $s > 0, m \in \{1, 2, 3, \dots\}$ are the parameters of the Zipf's distribution, and m is the maximum value of N; for this dataset, we chose $m = 98$, the maximum number of visits in the dataset.

- Zipf's distribution has a power law, so the right tail of the distribution is heavier than the commmonly-used Poisson distribution.
- Zipf's distribution can be looked at as a discretized Pareto distribution, and the value of s determines the degree of tail heaviness.
- The covariates are introduced as follows

$$\ln(s_i) = \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\eta}, \quad i = 1, \dots, 2263,$$

where $\boldsymbol{\eta}$ is the regression coefficients.

Marginal regression - Loss severity

- Lognormal model can be written as

$$f_Y(y|\mu,\sigma) = \frac{1}{\sigma y\sqrt{2\pi}} \exp\left\{-\frac{(\ln y-\mu)^2}{2\sigma^2}\right\},$$

where μ is the location parameter and σ is the scale parameter σ.

- The covariates are introduced through the following equation

$$\mu_i = \mathbf{x}_i^{\mathrm{T}}\gamma, \quad i = 1, \ldots, 2263,$$

where γ is the corresponding regression coefficients

The joint model using copulas

- Joint density functions of continuous Y and discrete N:

  $$f_{Y,N}(y,n|\theta) = f_Y \left[D_1(F_Y(y), F_N(n)|\theta) - D_1(F_Y(y), F_N(n-1)|\theta) \right],$$

  where $D_1(u,v|\theta) := \frac{\partial C(u,v|\theta)}{\partial u}$.
- MLEs for the overall likelihood (margins and dependence) were obtained.

Mixed copula regression with GGS copula (cont's)

		Marginal	s.e.	GGS	s.e.
Frequency	Intercept	0.855	0.039	0.869	0.038
	age	-0.002	0.001	-0.002	0.001
	ins(2)	-0.114	0.029	-0.108	0.028
	ins(3)	-0.085	0.053	-0.118	0.052
	race(2)	0.044	0.041	0.026	0.040
	race(3)	0.138	0.075	0.133	0.072
	race(4)	0.155	0.036	0.133	0.035
Severity	Intercept	5.911	0.096	5.889	0.094
	age	0.005	0.001	0.005	0.001
	ins(2)	-0.707	0.070	-0.685	0.069
	ins(3)	-0.541	0.132	-0.447	0.130
	race(2)	0.115	0.104	0.168	0.101
	race(3)	0.025	0.177	0.064	0.172
	race(4)	0.387	0.089	0.412	0.087
	$\ln(\sigma)$	0.418	0.015	0.421	0.015
Dependence	$\ln(\alpha)$	-	-	5.476	0.011
	$\ln(\beta)$	-	-	2.430	0.022

Mixed copula regression with GGS copula (cont's)

Table: Aggregate loss comparisons, where
AIC $= -2 \times$ log likelihood $+ 2 \times$ number of parameters.

	GGS copula	Independence	Data
Aggregate Loss (USD)	$5,733,236$	$8,153,765$	$5,371,218$
AIC	$41,812$	$41,869$	–

Conclusions

1. Strength of dependence in the tails can be efficiently quantified by tail orders.
2. Tail negative dependence can be introduced by scale mixture models (eg: Elliptical and WT-Archimedean copulas), while WT-Archimedean copula can handle different dependence patterns in upper and lower tails, respectively, ranging from positive to negative dependence.
3. Modeling the dependence structure in the upper tail carefully is particularly important for aggregate loss modeling, when loss frequency and loss severity are not independent.

C. Czado, R. Kastenmeier, E. Brechmann, and A. Min.
A Mixed Copula Model for Insurance Claims and Claim Sizes.
Scand. Actuar. J., 2011.

E. W. Frees.
Regression Modeling with Actuarial and Financial Applications.
Cambridge University Press, 2010.

L. Hua.
Tail negative dependence and its applications for aggregate loss modeling.
under review, 2014.

Thank you!

Tail negative dependence and its applications for aggregate loss modeling

Lei Hua

Division of Statistics

Northern Illinois University

Oct 20, 2014, ISU