Parametric maximum likelihood inference for copula models with dependently left-truncated data

Takeshi Emura
Graduate Institute of Statistics
National Central University, TAIWAN

Joint work with
Chi-Hung Pan

December 13, 2015

Left-truncation:

- (L, X): a pair of random variables
- L: left-truncation time
- X: failure time

If $L \le X$, the sample is available

If $L > X$, nothing is available !

What we observe:

$\{(L_j, X_j); j = 1, 2, \ldots, n\}$ subject to $L_j \le X_j$

Target of Estimation:

$S_X(x) = P(X > x)$, $E(X) = \int x dF_X(x)$, etc.

Examples of left-truncation

- Channing house data for elderly residents (Hyde, 1980)

  $L =$ Age at entry, $X =$ **Age at death**

- Car brake pads data (Kalbfleisch and Lawless,1992)

  $L =$ kilometers driven at study start

  $X =$ **Kilometers driven at failure**

- Unemployment data for women in Spain (De Uña-álvarez, 2004)

  $L =$ Time at inquiry, $X =$ **Time to finding a job**

- Twin-City study for dementia in France (Rondeau et al. 2015)

  $L =$ Age at entry, $X =$ **Age at diagnosis of dementia**

All studies target on the distribution of X

Estimation with left-truncated survival data

-Lynden-Bell (1971) : Product-limit estimator

$$\hat{S}_X(x) = \prod_{u \le x} \left\{ 1 - \frac{\sum_j \mathbf{I}(L_j \le u,\, X_j = u)}{\sum_j \mathbf{I}(L_j \le u \le X_j)} \right\}$$

Asymptotic properties studied well

(Woodroofe 1985; Wang et al. 1986; Zhou & Yip, 1999)

-Kalbfleisch and Lawless (1992): Parametric MLE

$$\hat{S}_X(x) = 1 - \Phi[\{\log(x) - \hat{\mu}\} / \hat{\sigma}],$$

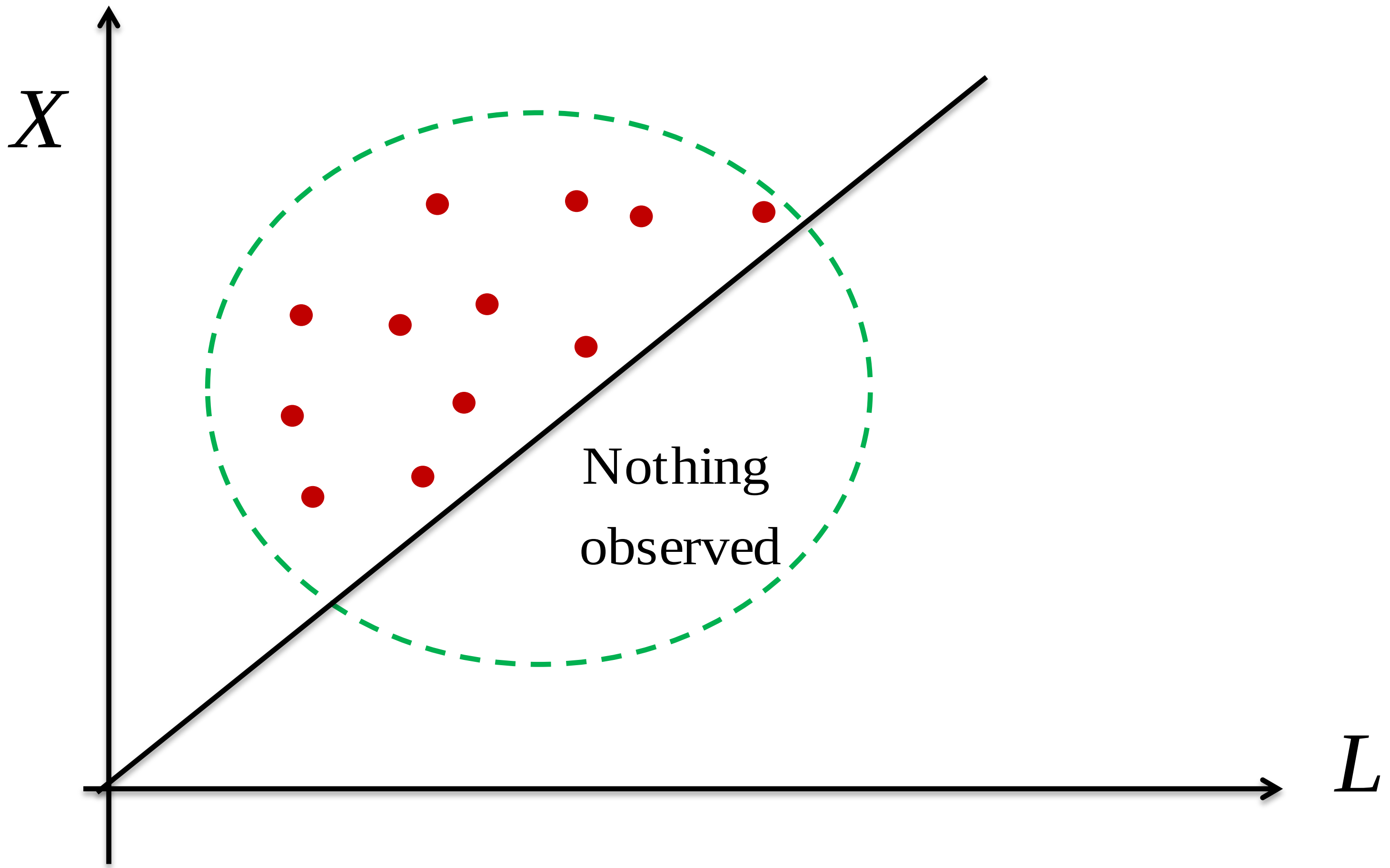
$$(\hat{\mu}, \hat{\sigma}) = \arg\max_{(\mu,\sigma)} \prod_{i=1}^{n} \left[\frac{f_X(X_i; \mu, \sigma)}{P(L_i \le X; \mu, \sigma)} \right]$$

These estimators are consistent under $L \perp X$:

, **independent truncation assumption**

- Independence assumption $L \perp X$ is testable by left-truncated data (Tsai, 1990)

$$\{(L_j, X_j);\ j = 1, 2, \ldots, n\} \text{ subject to } L_j \le X_j$$



Quasi-independence assumption (Tsai, 1990)

$$H_0: \ \Pr(X = x, L = l \mid X \le Y) \propto dF_X(x) dF_L(l)$$

Available test statistics:

1. Chen et al. (1996) - test with the conditional Pearson-correlation
2. Tsai (1990); Martin & Betensky (2005)
   - test with the conditional Kendall's tau
3. Emura & Wang (2010)
   - weighted-logrank test ; - score test under copulas
4. De Uña-Álvarez (2012); Rodríguez Girondo & de Uña-Álvarez (2012)
   - test with the Markov condition
5. Strzalkowska-Kominiak & Stute (2013)
   - test with Kendall's tau or Spearman's rho

- Quasi-independence is rejected in real examples
- Results not robust against dependent truncation (Bakoyannis & Touloumi, 2015)

Methods for dependent truncation

1. **Semiparametric model (marginal unspecified)**

   Chaieb et al.(2006), Emura et al. (2011), Emura and Murotani (2015)

   - Estimation under Archimedean copulas

   Beaudoin & Lakhal-Chaieb (2008),

   - Goodness-of-fit for Archimedean copulas

   Strzalkowska-Kominiak & Stute (2013)

   - Estimation under general copulas

   Emura & Wang (2012)

   - Estimation & model selection under general copulas

2. **Parametric model**

   Emura and Konno (2012a) - Bivariate normal model

   Emura and Konno (2012b) - Bivariate Poisson, Bernoulli-Poisson

Parametric models are still very limited

$\boldsymbol{\theta} = (\alpha, \boldsymbol{\theta}_L, \boldsymbol{\theta}_X) \in \Theta$

Proposed parametric models

- Joint distribution:

$$P_{\boldsymbol{\theta}}(L \le l, X \le x) = C_\alpha[F_L(l; \boldsymbol{\theta}_L), F_X(x; \boldsymbol{\theta}_X)]$$

- Copula:

$$C_\alpha : [0,1]^2 \mapsto [0,1],$$

$$\alpha \in R: \text{ dependence parameter}$$

- Arbitrary continuous parametric margins

$$F_L(l; \boldsymbol{\theta}_L) = P_{\boldsymbol{\theta}_L}(L \le l), \quad F_X(x; \boldsymbol{\theta}_X) = P_{\boldsymbol{\theta}_X}(X \le x)$$

- Unknown parameters:

$$\boldsymbol{\theta} = (\alpha, \boldsymbol{\theta}_L, \boldsymbol{\theta}_X) \in \Theta$$

Likelihood under dependent left-truncation

- Joint distribution:

$$P_{\boldsymbol{\theta}}(L \le l, X \le x) = C_{\alpha}[F_L(l; \boldsymbol{\theta}_L), F_X(x; \boldsymbol{\theta}_X)]$$

- Joint density:

$$f_{L,X}(l, x; \boldsymbol{\theta}) = C_{\alpha}^{[1,1]}[F_L(l; \boldsymbol{\theta}_L), F_X(x; \boldsymbol{\theta}_X)] f_L(l; \boldsymbol{\theta}_L) f_X(x; \boldsymbol{\theta}_X)$$

$C_{\alpha}^{[1,1]} = \partial^2 C_{\alpha} / \partial u \partial v$: copula density

$f_L = dF_L / dl$ and $f_X = dF_X / dx$: marginal densities

- Likelihood under left-truncation criterion: $L_j \le X_j$

$$L_n(\boldsymbol{\theta}) = \prod_{j=1}^{n} \frac{f_{L,X}(L_j, X_j; \boldsymbol{\theta})}{P_{\boldsymbol{\theta}}(L \le X)}$$

Need the form of $c(\boldsymbol{\theta}) = P_{\boldsymbol{\theta}}(L \le X) = \iint_{l \le x} f_{L,X}(l, x; \boldsymbol{\theta}) dxdl$

Forms of $c(\boldsymbol{\theta}) = \Pr(L \le X) = \iint_{l \le x} f_{L,X}(l, x; \boldsymbol{\theta})\,dx\,dl$

- Marshal and Olkin (1967) bivarate exponential

$$\Pr(L^O > l,\, X^O > x) = \exp\{-\lambda_L l - \lambda_X x - \lambda_{LX}\max(l, x)\} \qquad (l \ge 0,\, x \ge 0)$$

$$\Rightarrow \quad c(\boldsymbol{\theta}) = (\lambda_X + \lambda_{LX})(\lambda_L + \lambda_X + \lambda_{LX})^{-1}$$

- Bivariate normal, bivariate *t*

$$c(\boldsymbol{\theta}) = \Psi\left(\frac{\mu_X - \mu_L}{\sqrt{\sigma_X^2 + \sigma_L^2 - 2\sigma_{LX}}}; v \right)$$

- Bivariate Poisson

$$c(\boldsymbol{\theta}) = e^{-\lambda_L - \lambda_X} \sum_{l=0}^{\infty} \frac{\lambda_L^l}{l!} S_V(l)$$

- FGM copula with Burr III margin

$$c(\boldsymbol{\theta}) = \frac{\gamma}{\gamma + \beta} + \alpha \frac{\gamma\beta(\gamma - \beta)}{(\gamma + \beta)(2\gamma + \beta)(\gamma + 2\beta)}$$

- Bernoulli-Poisson

$$c(\boldsymbol{\theta}) = 1 - p e^{-\lambda_X}$$

Ref: Emura and Konno (2012a, b): Domma and Giordano (2013)

- However, a general formula of $c(\boldsymbol{\theta})$ seems unavailable

$$c(\boldsymbol{\theta}) = P_{\boldsymbol{\theta}}(L \le X) = \iint_{l \le x} f_{L,X}(l, x; \boldsymbol{\theta})\,dl\,dx$$

$$= \iint_{l \le x} C_{\alpha}^{[1,1]}[F_L(l; \boldsymbol{\theta}_L), F_X(x; \boldsymbol{\theta}_X)]\, f_L(l; \boldsymbol{\theta}_L)\, f_X(x; \boldsymbol{\theta}_X)\,dl\,dx$$

- Define *h-function* (Schepsmeier and StÖber, 2014)

$$h_{\alpha}(u, v) \equiv \partial C_{\alpha}(u, v) / dv$$

Theorem 1 (new result in this work)

The inclusion probability can be simplify as follows:

$$c(\boldsymbol{\theta}) = \Pr(L \le X) = \int_0^1 H(u; \boldsymbol{\theta})\,du,$$

where $H(u; \boldsymbol{\theta}) \equiv h_{\alpha}[F_L\{F_X^{-1}(u; \boldsymbol{\theta}_X); \boldsymbol{\theta}_L\}, u].$

Proposed likelihood estimation

- The log-likelihood function:

$$\ell_n(\boldsymbol{\theta}) = -n\log c(\boldsymbol{\theta}) + \sum_j \log f_L(L_j;\boldsymbol{\theta}_L) + \sum_j \log f_X(X_j;\boldsymbol{\theta}_X) + \sum_j \log C_\alpha^{[1,1]}[F_L(L_j;\boldsymbol{\theta}_L), F_X(X_j;\boldsymbol{\theta}_X)],$$

where $c(\boldsymbol{\theta}) = \Pr(L \le X) = \int_0^1 H(u;\boldsymbol{\theta})du$

- Maximum likelihood estimator (MLE):

$$\hat{\boldsymbol{\theta}} = (\hat{\alpha}, \hat{\boldsymbol{\theta}}_L, \hat{\boldsymbol{\theta}}_X) = \arg\max_{\boldsymbol{\theta}\in\Theta} \ell_n(\boldsymbol{\theta})$$

- The Newton-Raphson algorithm requires

$$\partial \ell_n(\boldsymbol{\theta}) / \partial\boldsymbol{\theta} \quad , \qquad \partial^2 \ell_n(\boldsymbol{\theta}) / \partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T$$

Score functions and Hessian matrix

Lemma 1 [p.301 Khuri (2003)]

If $H: R^{p+1} \rightarrow R$ *and* $\partial H / \partial \theta_i$ *are continuous in a rectangle, then*

$$\frac{\partial}{\partial \theta_i} \int_0^1 H(u; \boldsymbol{\theta})\, du = \int_0^1 \frac{\partial H(u; \boldsymbol{\theta})}{\partial \theta_i} du, \qquad i = 1, \ldots, p.$$

If Lemma 1 holds,

Score functions and Hessian matrix are calculated with

$$\begin{bmatrix} c_{\alpha}(\boldsymbol{\theta}) \\ c_{\boldsymbol{\theta}_L}(\boldsymbol{\theta}) \\ c_{\boldsymbol{\theta}_X}(\boldsymbol{\theta}) \end{bmatrix} \equiv \begin{bmatrix} \partial c(\boldsymbol{\theta}) / \partial \alpha \\ \partial c(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}_L \\ \partial c(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}_X \end{bmatrix} = \int_0^1 \begin{bmatrix} \partial H(u; \boldsymbol{\theta}) / \partial \alpha \\ \partial H(u; \boldsymbol{\theta}) / \partial \boldsymbol{\theta}_L \\ \partial H(u; \boldsymbol{\theta}) / \partial \boldsymbol{\theta}_X \end{bmatrix} du$$

$$\begin{bmatrix} c_{\alpha\alpha}(\boldsymbol{\theta}) & c_{\alpha\boldsymbol{\theta}_L}^{\mathrm{T}}(\boldsymbol{\theta}) & c_{\alpha\boldsymbol{\theta}_X}^{\mathrm{T}}(\boldsymbol{\theta}) \\ c_{\alpha\boldsymbol{\theta}_L}(\boldsymbol{\theta}) & c_{\boldsymbol{\theta}_L\boldsymbol{\theta}_L}(\boldsymbol{\theta}) & c_{\boldsymbol{\theta}_L\boldsymbol{\theta}_X}^{\mathrm{T}}(\boldsymbol{\theta}) \\ c_{\alpha\boldsymbol{\theta}_X}(\boldsymbol{\theta}) & c_{\boldsymbol{\theta}_L\boldsymbol{\theta}_X}(\boldsymbol{\theta}) & c_{\boldsymbol{\theta}_X\boldsymbol{\theta}_X}(\boldsymbol{\theta}) \end{bmatrix} \equiv \int_0^1 \begin{bmatrix} \partial^2 H(u; \boldsymbol{\theta}) / \partial \alpha^2 & \partial^2 H(u; \boldsymbol{\theta}) / \partial \alpha \partial \boldsymbol{\theta}_L^{\mathrm{T}} & \partial^2 H(u; \boldsymbol{\theta}) / \partial \alpha \partial \boldsymbol{\theta}_X^{\mathrm{T}} \\ \partial^2 H(u; \boldsymbol{\theta}) / \partial \alpha \partial \boldsymbol{\theta}_L & \partial^2 H(u; \boldsymbol{\theta}) / \partial \boldsymbol{\theta}_L \partial \boldsymbol{\theta}_L^{\mathrm{T}} & \partial^2 H(u; \boldsymbol{\theta}) / \partial \boldsymbol{\theta}_L \partial \boldsymbol{\theta}_X^{\mathrm{T}} \\ \partial^2 H(u; \boldsymbol{\theta}) / \partial \alpha \partial \boldsymbol{\theta}_X & \partial^2 H(u; \boldsymbol{\theta}) / \partial \boldsymbol{\theta}_X \partial \boldsymbol{\theta}_L^{\mathrm{T}} & \partial^2 H(u; \boldsymbol{\theta}) / \partial \boldsymbol{\theta}_X \partial \boldsymbol{\theta}_X^{\mathrm{T}} \end{bmatrix} du$$

Example:
Weibull models with the Clayton copula

- Clayton copula: $C_\alpha(u_1, u_2) = (u_1^{-\alpha} + u_2^{-\alpha} - 1)^{-1/\alpha}, \qquad \alpha \geq 0$

  ➔ h-function: $h_\alpha(u_1, u_2) = u_2^{-\alpha-1}(u_1^{-\alpha} + u_2^{-\alpha} - 1)^{-1/\alpha-1}$

- Weibull marginals: $F_L(l; \lambda_L, \nu_L) = 1 - \exp(-\lambda_L l^{\nu_L})$

  $F_X(x; \lambda_X, \nu_X) = 1 - \exp(-\lambda_X x^{\nu_X})$

- Unknown parameters: $\boldsymbol{\theta} = (\alpha, \lambda_L, \lambda_X, \nu_L, \nu_X) \in \Theta = (0, \infty)^5$

- Sample inclusion probability:

$$c(\boldsymbol{\theta}) = \Pr(L \leq X) = \int_0^1 H(u; \boldsymbol{\theta})\, du$$

$$H(u; \boldsymbol{\theta}) = u^{-\alpha-1} B(u, \boldsymbol{\theta})^{-1/\alpha-1}$$

$$B(u; \boldsymbol{\theta}) = (1 - \exp[-\lambda_L \{-\lambda_X^{-1} \log(1-u)\}^{\nu_L/\nu_X}])^{-\alpha} + u^{-\alpha} - 1$$

Example:
Weibull models with the Clayton copula

- Score function (for α):

$$\partial \ell_n(\boldsymbol{\theta})/\partial\alpha = -nc_\alpha(\boldsymbol{\theta})/c(\boldsymbol{\theta}) + \sum_j \partial \log\{\, f_{L,X}(L_j, X_j; \boldsymbol{\theta})\}/\partial\alpha$$

where

$$c_\alpha(\boldsymbol{\theta}) \equiv \frac{\partial c(\boldsymbol{\theta})}{\partial\alpha} = \int_0^1 \frac{\partial H(u;\boldsymbol{\theta})}{\partial\alpha}\, du$$

$$H(u;\boldsymbol{\theta}) = u^{-\alpha-1} B(u,\boldsymbol{\theta})^{-1/\alpha-1}$$

$$B(u;\boldsymbol{\theta}) = (1-\exp[-\lambda_L\{-\lambda_X^{-1}\log(1-u)\}^{\nu_L/\nu_X}])^{-\alpha} + u^{-\alpha} - 1$$

$$\frac{\partial \log f_{L,X}(l,x;\boldsymbol{\theta})}{\partial\alpha} = \frac{1}{1+\alpha} - \log\{1-\exp(-\lambda_L l^{\nu_L})\} - \log\{1-\exp(-\lambda_X x^{\nu_X})\}$$
$$+ \frac{\log A(l,x;\boldsymbol{\theta})}{\alpha^2} - \left(\frac{1}{\alpha}+2\right)\frac{A_\alpha(l,x;\boldsymbol{\theta})}{A(l,x;\boldsymbol{\theta})},$$

$$A_\alpha(l,x;\boldsymbol{\theta}) \equiv \frac{\partial A(l,x;\boldsymbol{\theta})}{\partial\alpha} = -\frac{\log\{1-\exp(-\lambda_L l^{\nu_L})\}}{\{1-\exp(-\lambda_L l^{\nu_L})\}^\alpha} - \frac{\log\{1-\exp(-\lambda_X x^{\nu_X})\}}{\{1-\exp(-\lambda_X x^{\nu_X})\}^\alpha},$$

Example:
Weibull models with the Clayton copula

Score functions and Hessian matrix are explicitly written (up to 1-dimensional integration) since

$$\begin{bmatrix} c_{\alpha}(\boldsymbol{\theta}) \\ c_{\lambda_L}(\boldsymbol{\theta}) \\ c_{\lambda_X}(\boldsymbol{\theta}) \\ c_{\nu_L}(\boldsymbol{\theta}) \\ c_{\nu_X}(\boldsymbol{\theta}) \end{bmatrix} \equiv \begin{bmatrix} \partial c(\boldsymbol{\theta})/\partial \alpha \\ \partial c(\boldsymbol{\theta})/\partial \lambda_L \\ \partial c(\boldsymbol{\theta})/\partial \lambda_X \\ \partial c(\boldsymbol{\theta})/\partial \nu_L \\ \partial c(\boldsymbol{\theta})/\partial \nu_X \end{bmatrix} = \int_0^1 \begin{bmatrix} \partial H(u;\boldsymbol{\theta})/\partial \alpha \\ \partial H(u;\boldsymbol{\theta})/\partial \lambda_L \\ \partial H(u;\boldsymbol{\theta})/\partial \lambda_X \\ \partial H(u;\boldsymbol{\theta})/\partial \nu_L \\ \partial H(u;\boldsymbol{\theta})/\partial \nu_X \end{bmatrix} du$$

$$\partial H(u;\boldsymbol{\theta})/\partial \lambda_L = (-1/\alpha - 1)u^{-\alpha-1}B(u;\boldsymbol{\theta})^{-1/\alpha-2}B_{\lambda_L}(u;\boldsymbol{\theta}),$$

$$\partial H(u;\boldsymbol{\theta})/\partial \lambda_X = (-1/\alpha - 1)u^{-\alpha-1}B(u;\boldsymbol{\theta})^{-1/\alpha-2}B_{\lambda_X}(u;\boldsymbol{\theta}),$$

$$\partial H(u;\boldsymbol{\theta})/\partial \nu_L = (-1/\alpha - 1)u^{-\alpha-1}B(u;\boldsymbol{\theta})^{-1/\alpha-2}B_{\nu_L}(u;\boldsymbol{\theta}),$$

$$\partial H(u;\boldsymbol{\theta})/\partial \nu_X = (-1/\alpha - 1)u^{-\alpha-1}B(u;\boldsymbol{\theta})^{-1/\alpha-2}B_{\nu_X}(u;\boldsymbol{\theta}),$$

$\max\{|\alpha^{(k+1)}-\alpha^{(k)}|, |\lambda_L^{(k+1)}-\lambda_L^{(k)}|, |\lambda_X^{(k+1)}-\lambda_X^{(k)}|, |\nu_L^{(k+1)}-\nu_L^{(k)}|, |\nu_X^{(k+1)}-\nu_X^{(k)}|\} < \varepsilon$

Newton-Raphson algorithm

Step 1: Set the initial values

$$\alpha^{(0)} = 2\hat{\tau}/(1-\hat{\tau}),\ \lambda_L^{(0)} = 1/\bar{L},\ \lambda_X^{(0)} = 1/\bar{X},\ \nu_L^{(0)} = 1 \text{ and } \nu_X^{(0)} = 1$$

Step 2: Repeat

$$\begin{bmatrix} \alpha^{(k+1)} \\ \lambda_L^{(k+1)} \\ \lambda_X^{(k+1)} \\ \nu_L^{(k+1)} \\ \nu_X^{(k+1)} \end{bmatrix} = \begin{bmatrix} \alpha^{(k)} \\ \lambda_L^{(k)} \\ \lambda_X^{(k)} \\ \nu_L^{(k)} \\ \nu_X^{(k)} \end{bmatrix} - \begin{bmatrix} \frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \alpha^2} & \frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \lambda_L \partial \alpha} & \frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \lambda_X \partial \alpha} & \frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \nu_L \partial \alpha} & \frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \nu_X \partial \alpha} \\ \frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \lambda_L \partial \alpha} & \frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \lambda_L^2} & \frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \lambda_L \partial \lambda_X} & \frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \lambda_L \partial \nu_L} & \frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \lambda_L \partial \nu_X} \\ \frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \lambda_X \partial \alpha} & \frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \lambda_L \partial \lambda_X} & \frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \lambda_X^2} & \frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \nu_L \partial \lambda_X} & \frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \nu_X \partial \lambda_X} \\ \frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \nu_L \partial \alpha} & \frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \nu_L \partial \lambda_L} & \frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \nu_L \partial \lambda_X} & \frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \nu_L^2} & \frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \nu_L \partial \nu_X} \\ \frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \nu_X \partial \alpha} & \frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \nu_X \partial \lambda_L} & \frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \nu_X \partial \lambda_X} & \frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \nu_L \partial \nu_X} & \frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \nu_X^2} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial \ell_n(\boldsymbol{\theta})}{\partial \alpha} \\ \frac{\partial \ell_n(\boldsymbol{\theta})}{\partial \lambda_L} \\ \frac{\partial \ell_n(\boldsymbol{\theta})}{\partial \lambda_X} \\ \frac{\partial \ell_n(\boldsymbol{\theta})}{\partial \nu_L} \\ \frac{\partial \ell_n(\boldsymbol{\theta})}{\partial \nu_X} \end{bmatrix} \Bigg|_{\substack{\alpha=\alpha^{(k)} \\ \lambda_L=\lambda_L^{(k)},\ \lambda_X=\lambda_X^{(k)} \\ \nu_L=\nu_L^{(k)},\ \nu_X=\nu_X^{(k)}}}$$

If $\|\boldsymbol{\theta}^{(k+1)} - \boldsymbol{\theta}^{(k)}\| < \varepsilon$, (e.g., $\varepsilon = 10^{-4}$)

then, $\boldsymbol{\theta}^{(k+1)} = (\alpha^{(k+1)}, \lambda_L^{(k+1)}, \lambda_X^{(k+1)}, \nu_L^{(k+1)}, \nu_X^{(k+1)})$ is the MLE

Stabilizing the Newton-Raphson algorithm

Step 3: If the algorithm **diverges**

$$\max\{\,|\alpha^{(k+1)}-\alpha^{(k)}|,|\lambda_L^{(k+1)}-\lambda_L^{(k)}|,|\lambda_X^{(k+1)}-\lambda_X^{(k)}|,|\nu_L^{(k+1)}-\nu_L^{(k)}|,|\nu_X^{(k+1)}-\nu_X^{(k)}|\,\}>2$$

then, return to **Step 1** by replacing the initial values

$$(\,\alpha^{(0)},\lambda_L{}^{(0)},\lambda_X{}^{(0)},\nu_L{}^{(0)},\nu_X{}^{(0)}\,)$$

by "**randomized**" initial values

$$\{\,\alpha^{(0)}\times\exp(u_1),$$

$$\lambda_L^{(0)}\times\exp(u_2),\lambda_X^{(0)}\times\exp(u_3),\nu_L^{(0)}\times\exp(u_4),\nu_X^{(0)}\times\exp(u_5)\,\}$$

$$u_i\sim U(-r_i,r_i),\quad i=1,\ldots,5,\quad r_i>0:\ \text{radius}$$

Termed **"Randomized Newton-Raphson"** (Hu and Emura, 2015)

Asymptotic inference

- Target : $g(\boldsymbol{\theta})$
- Standard error (SE)

Last step of
the Newton-Raphson

$$SE\{g(\hat{\boldsymbol{\theta}})\} = \sqrt{\left\{\frac{\partial}{\partial\boldsymbol{\theta}} g(\hat{\boldsymbol{\theta}})\right\}^{\mathrm{T}} \times \left\{-\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}}} \ell_n(\hat{\boldsymbol{\theta}})\right\}^{-1} \times \frac{\partial}{\partial\boldsymbol{\theta}} g(\hat{\boldsymbol{\theta}})}$$

- $(1-\alpha)100\%$ confidence interval

$$[\, g(\hat{\boldsymbol{\theta}}) - Z_{\alpha/2} \cdot SE\{g(\hat{\boldsymbol{\theta}})\},\quad g(\hat{\boldsymbol{\theta}}) + Z_{\alpha/2} \cdot SE\{g(\hat{\boldsymbol{\theta}})\}\,] .$$

- Example: Mean failure time

$$g(\boldsymbol{\theta}) = E(X) = \int x f_X(x; \boldsymbol{\theta}_X)\,dx$$

Simulation setting

- Weibull model with the Clayton copula

$$\Pr(L \le l, X \le x)$$

$$= [\{1 - \exp(-\lambda_L l^{\nu_L})\}^{-\alpha} + \{1 - \exp(-\lambda_X x^{\nu_X})\}^{-\alpha} - 1]^{-1/\alpha}$$

where $\alpha = 2$ s.t. $(\tau = 0.5)$

- Parameters $(\lambda_L, \lambda_X, \nu_L, \nu_X)$ chosen to be

$c(\boldsymbol{\theta}) > 0.5$, $c(\boldsymbol{\theta}) = 0.5$, or $c(\boldsymbol{\theta}) < 0.5$

- Target: $(\alpha, \lambda_L, \lambda_X, \nu_L, \nu_X)$

$$\mu_X = E(X) = \Gamma(1 + 1/\nu_X)/(\lambda_X^{1/\nu_X})$$

- Generatè $\{(L_j, X_j);\ j = 1, 2, \ldots, n\}$ subject to $L_j \le X_j$

Simulation results under Weibull models with the Clayton copula based on 1000 replications

True: $\alpha = 2$ $(\tau = 0.5)$

$c(\boldsymbol{\theta})$	n	$E(\hat{\alpha})$	$E(\hat{\lambda}_L)$	$E(\hat{\lambda}_X)$	$E(\hat{\nu}_L)$	$E(\hat{\nu}_X)$	AI
0.804	100	2.065	2.040	1.031	1.007	1.003	6.8
	200	2.037	2.017	1.009	1.003	1.004	6.1
	300	2.029	2.012	1.008	1.001	1.002	6.1
0.500	100	2.167	1.023	1.130	0.992	0.979	65.2
	200	2.086	1.014	1.052	0.993	0.990	8.7
	300	2.057	1.009	1.032	0.996	0.995	8.2
0.387	100	2.064	0.971	1.169	2.123	1.025	69.9
	200	2.020	0.986	1.050	2.044	1.018	54.2
	300	2.003	0.986	1.019	2.036	1.019	50.0

AI = The average number of Newton-Raphson iterations until convergence

$\lambda_X = \nu_X = 1$ (True)

$v_0 - u_0$

Simulation results under Weibull models with the Clayton copula based on 1000 replications

$c(\boldsymbol{\theta})$	n	$SD(\hat{\lambda}_X)$	$E\{SE(\hat{\lambda}_X)\}$	95%Cov	$SD(\hat{\nu}_X)$	$E\{SE(\hat{\nu}_X)\}$	95%Cov
0.804	100	0.172	0.137	0.946	0.092	0.089	0.945
	200	0.091	0.088	0.940	0.060	0.060	0.947
	300	0.073	0.071	0.945	0.048	0.048	0.951
0.500	100	0.495	0.324	0.916	0.155	0.139	0.913
	200	0.246	0.203	0.923	0.106	0.100	0.927
	300	0.176	0.157	0.927	0.085	0.081	0.944
0.387	100	0.790	0.564	0.921	0.269	0.244	0.915
	200	0.373	0.341	0.949	0.172	0.174	0.934
	300	0.295	0.262	0.959	0.143	0.141	0.941

95%Cov = Coverage rates for the 95% confidence intervals.

$v_0 - u_0$

Simulation results under Weibull models with the Clayton copula based on 1000 replications

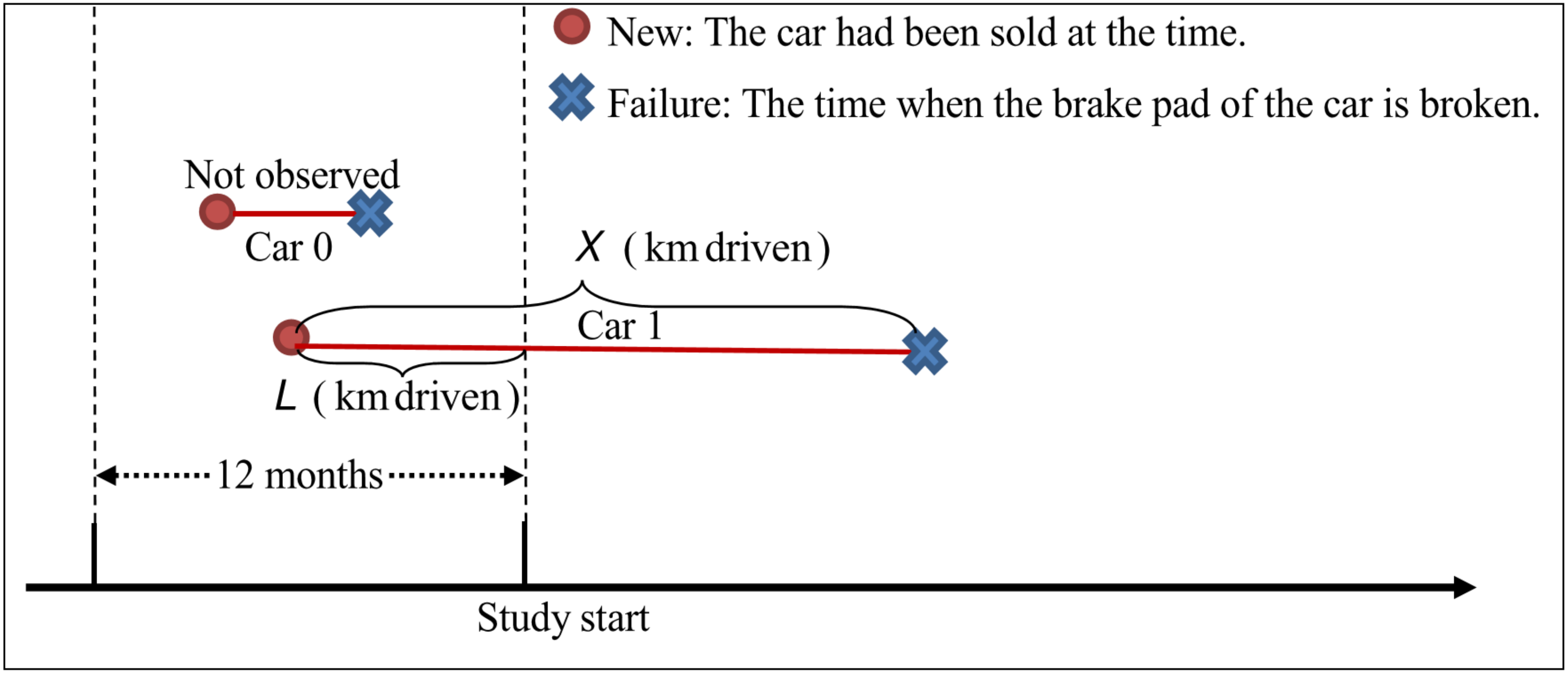
$c(\boldsymbol{\theta})$	n	$SD(\hat{\alpha})$	$E\{SE(\hat{\alpha})\}$	95%Cov	$SD(\hat{\mu}_X)$	$E\{SE(\hat{\mu}_X)\}$	95%Cov
0.804	100	0.428	0.448	0.954	0.117	0.112	0.947
	200	0.305	0.312	0.951	0.077	0.076	0.948
	300	0.250	0.254	0.949	0.062	0.062	0.960
0.500	100	0.576	0.552	0.944	0.232	0.198	0.916
	200	0.348	0.351	0.946	0.157	0.145	0.928
	300	0.268	0.279	0.948	0.123	0.118	0.941
0.387	100	1.009	0.894	0.921	0.344	0.308	0.919
	200	0.607	0.585	0.940	0.232	0.230	0.952
	300	0.474	0.456	0.934	0.190	0.187	0.958

95%Cov = Coverage rates for the 95% confidence intervals.

$$\mu_X = E(X) = \Gamma(1 + 1/\nu_X)/(\lambda_X^{1/\nu_X})$$

Car brake pads data (Kalbfleisch and Lawless,1992 JQT)

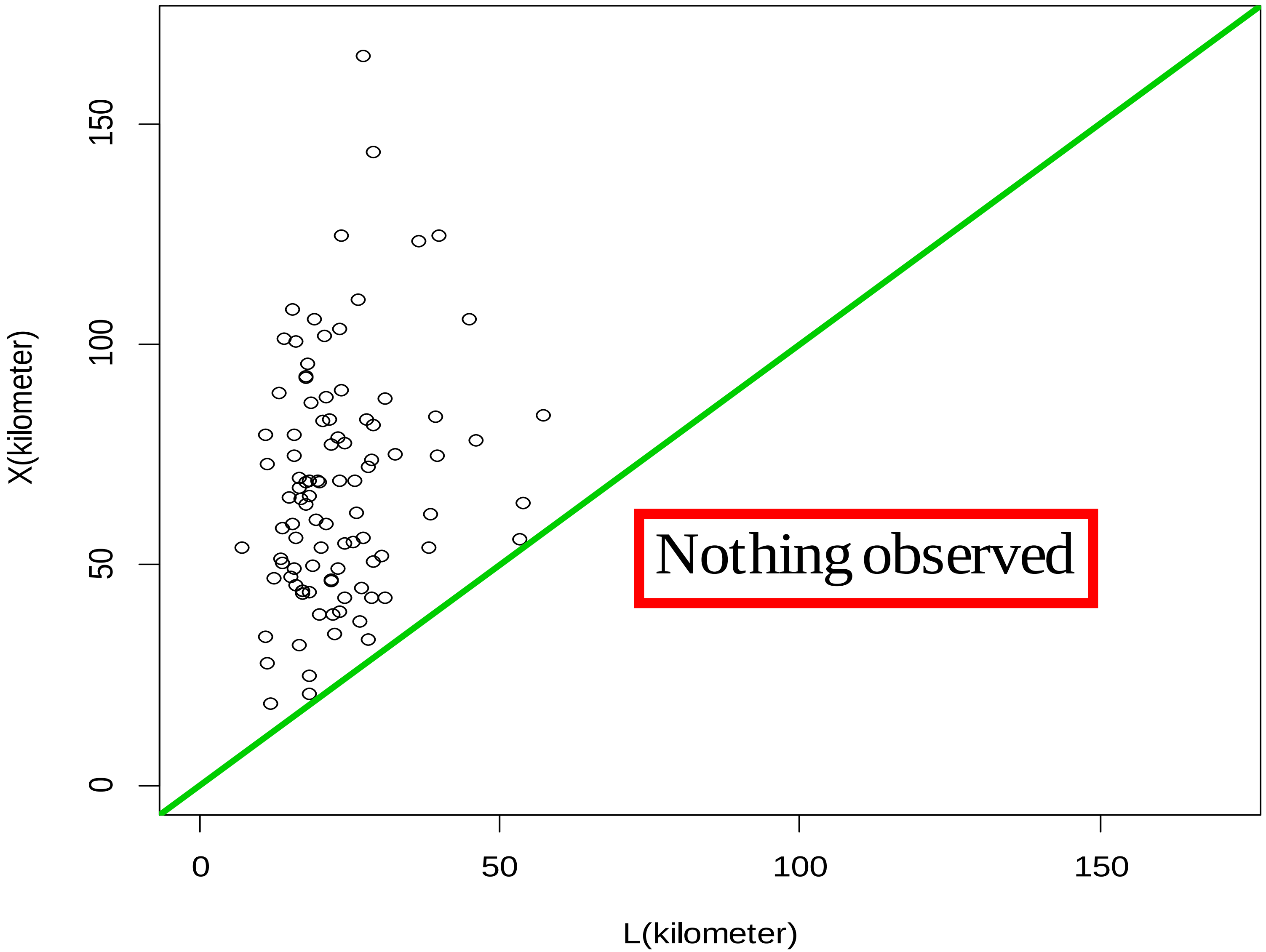
- X : The number of kilometers driven until failure
- L : The number of kilometers driven until the study start



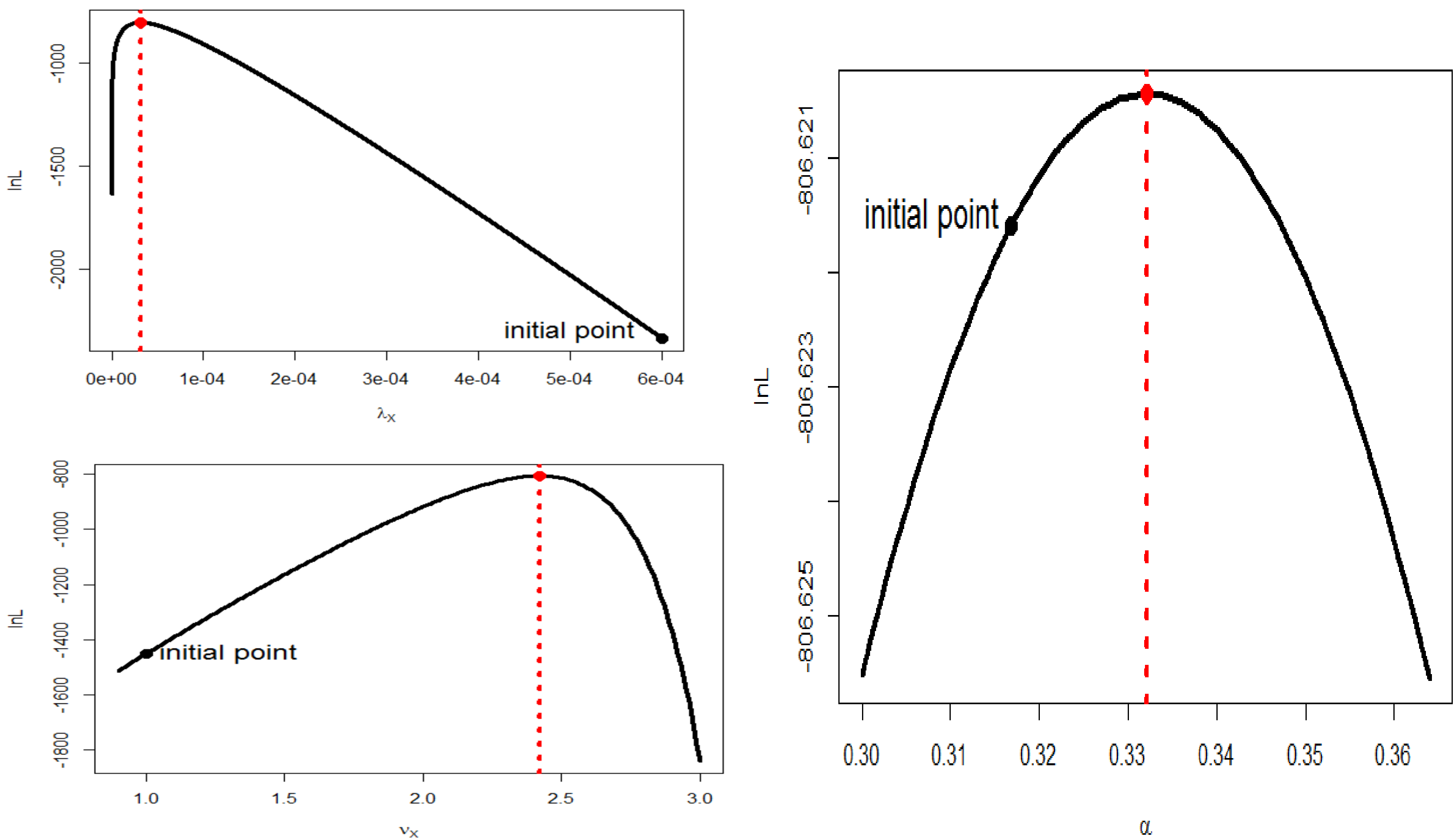
Observations : $\{ (L_j, X_j); j = 1, 2, \ldots, n \}$ subject to $L_j \le X_j$

Scatter plot of the brake pads data

$\{(L_j, X_j);\ j = 1, 2, \ldots, n\}, n = 98,$ subject to $L_j \le X_j$



The Newton-Raphson identifies the maximum of the likelihood



Estimation and model selection

Table 4 The MLE for parameters, maximized value of the log-likelihood function, AIC and BIC.

Model	$\hat{E}(X)$	$\hat{\alpha}$	$\hat{\lambda}_L$	$\hat{\lambda}_X$	$\hat{\nu}_L$	$\hat{\nu}_X$	$\log L$	AIC	BIC
M_1	48.31	0.924	0.0364	0.0207	1 (fixed)	1 (fixed)	-874.31	1754.61	1762.36
	(5.49)	(0.663)	(0.0081)	(0.0024)	-	-	-	-	-
M_2	64.82	0.332	2.89×10^{-4}	3.09×10^{-5}	2.493	2.419	-806.62	1623.24	1636.17
	(3.25)	(0.358)	(1.86×10^{-4})	(3.15×10^{-5})	(0.183)	(0.222)	-	-	-

M_1: The Clayton copula with exponential margins.

M_2: The Clayton copula with Weibull margins.

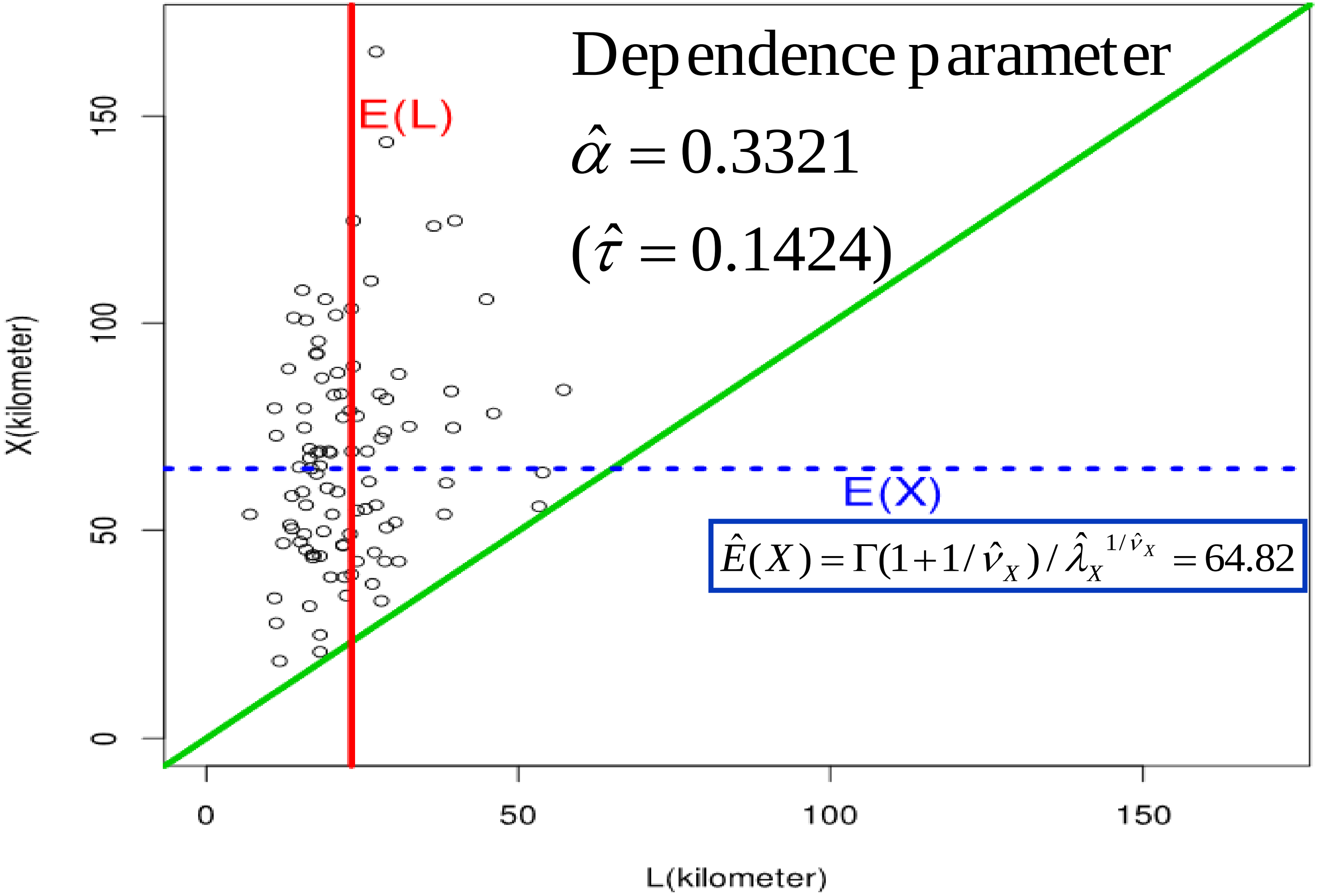
$(\cdot)$: The standard error (SE) of each parameters.

Smaller

- M_2 (Weibull marginal) shows better fit than M_1 (Exponential marginal)

Parameter estimates under the Weibull model

$$\hat{E}(L) = \Gamma(1+1/\hat{\nu}_L)/\hat{\lambda}_L^{1/\hat{\nu}_L} = 23.33$$



Dependence parameter

$$\hat{\alpha} = 0.3321$$

$$(\hat{\tau} = 0.1424)$$

$$\hat{E}(X) = \Gamma(1+1/\hat{\nu}_X)/\hat{\lambda}_X^{1/\hat{\nu}_X} = 64.82$$

Goodness-of-fit of the Weibull model

- Goodness-of fit test

$$H_0 : F_{\boldsymbol{\theta}}(l, x) = C_{\alpha}[\, F_L(l; \boldsymbol{\theta}_L), F_X(x; \boldsymbol{\theta}_X)\,] \qquad \exists\, \alpha, \boldsymbol{\theta}_L, \boldsymbol{\theta}_X$$

- A parametric Bootstrap test similar to Emura and Konno (2010a, b)

Step1. Compute the statistic CM (Cramér-von-Mises type statistics).

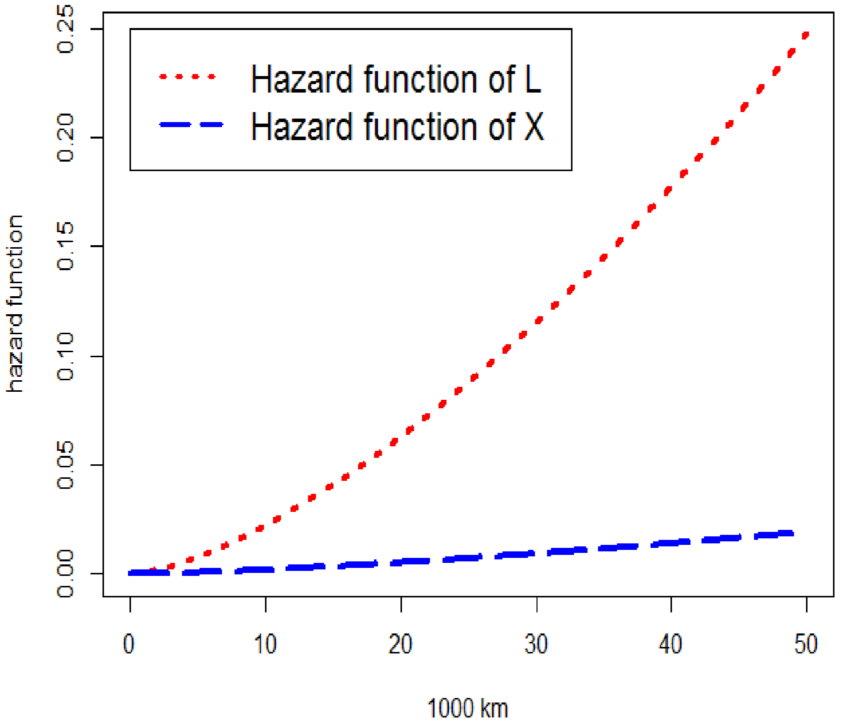
Step2. $(L_k^{(b)}, X_k^{(b)}) \sim C_{\hat{\alpha}}[\, F_L(l; \hat{\boldsymbol{\theta}}_L), F_X(x; \hat{\boldsymbol{\theta}}_X)\,] / c(\hat{\boldsymbol{\theta}})$, for $b = 1, 2, ..., B$, $k = 1, 2, ..., n$.

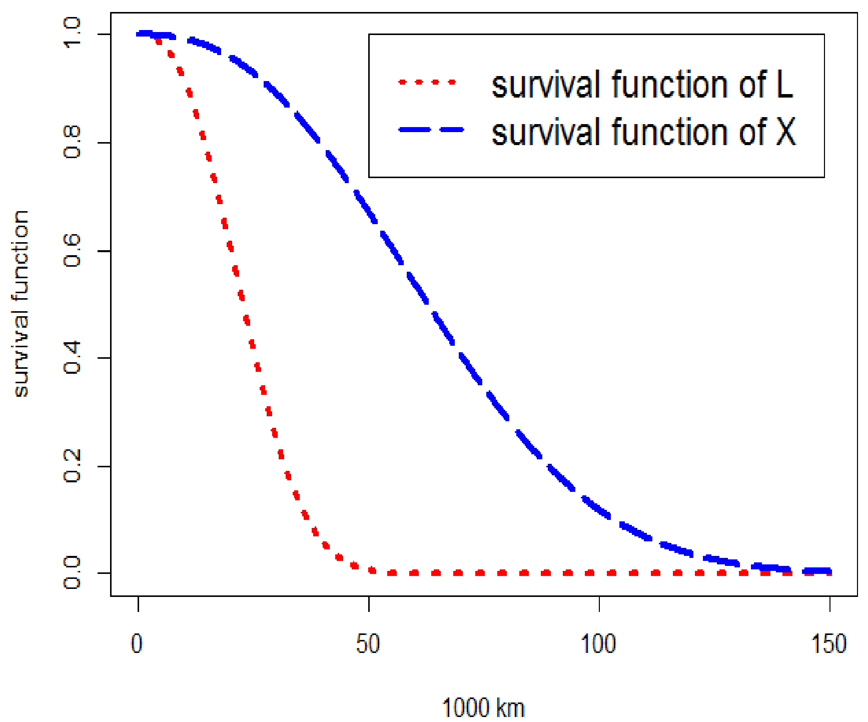
Step3. $CM^{*(b)} = \sum_k \{\, \hat{F}_e^{(b)}(L_k^{(b)}, X_k^{(b)}) - C_{\hat{\alpha}^{(b)}}[\, F_L(L_k^{(b)}; \hat{\boldsymbol{\theta}}_L^{(b)}), F_X(X_k^{(b)}; \hat{\boldsymbol{\theta}}_X^{(b)})\,] / c(\hat{\boldsymbol{\theta}}^{(b)})\,\}^2$

Step4. Approximate p-values is $\sum_b (CM^{*(b)} \geq CM) / B$.

- Approximate p-value is 0.069. (1000 bootstrap replications)
  - Does not reject H_0.

Estimated survival functions for marginal distributions: Weibull models





Estimates: $\hat{\lambda}_L = 2.89 \times 10^{-4}$, $\hat{\nu}_L = 2.4927$; $\hat{\lambda}_X = 3.09 \times 10^{-5}$, and $\hat{\nu}_X = 2.4193$.

Conclusion

- Copula-based parametric models for dependent truncation, extending existing models of
  - Emura and Konno (2012a): Bivariate normal , Bivariate *t*
  - Emura and Konno (2012b): Bivariate Poisson, Bernoulli-Poisson
- Key technical tools: novel expressions

$$c(\boldsymbol{\theta}) = \Pr(L \le X) = \int_0^1 H(u;\boldsymbol{\theta})du \qquad c_\alpha(\boldsymbol{\theta}) \equiv \partial c(\boldsymbol{\theta})/\partial\alpha = \int_0^1 \frac{\partial H(u;\boldsymbol{\theta})}{\partial\alpha} du$$

➔ allows an explicit (up to 1-dim. integral) Newton-Raphson

- Good finite sample properties (convergence speed, consistency, coverage probability) are confirmed by simulations

Future works

- Performance of the goodness-of-fit test must be examined by simulations (Type I error, power, etc.)