Conditional and partial copulas and measures of associations

Irène Gijbels[1], Noël Veraverbeke[2], Marek Omelka[3]

[1]Department of Statistics,
Katholieke Universiteit Leuven

[2]Center for Statistics, Hasselt University

[3]Department of Probability and Statistics,
Charles University in Prague

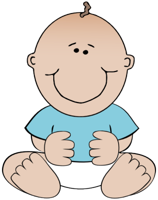
Robust, 2012

Outline

Life expectancies of males and females

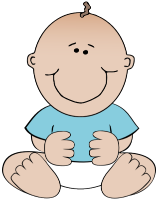
CIA World Factbook contains various demographic and economic characteristics of countries.

We are interested in the relationship of the **life expectancies** of males (Y_1) and females (Y_2) across countries.



Life expectancies of males and females

CIA World Factbook contains various demographic and economic characteristics of countries.



We are interested in the relationship of the **life expectancies** of males (Y_1) and females (Y_2) across countries.

Question: How does this relationship change when GDP per capita (X) is taken into consideration?

Measures of dependence (concordance)

Pearson correlation coefficient – $\rho(Y_1, Y_2) = \frac{\operatorname{cov}(Y_1, Y_2)}{\sqrt{\operatorname{var}(Y_1)}\sqrt{\operatorname{var}(Y_2)}}$.

Measures of dependence (concordance)

Pearson correlation coefficient – $\rho(Y_1, Y_2) = \frac{\text{cov}\,(Y_1,Y_2)}{\sqrt{\text{var}\,(Y_1)}\sqrt{\text{var}\,(Y_2)}}$.

Kendall's tau –

$$\tau(Y_1, Y_2) = 2\,\mathbb{P}\big[(Y_1 - Y_1')(Y_2 - Y_2') > 0\big] - 1,$$

where (Y_1', Y_2') is an independent copy of (Y_1, Y_2).

Measures of dependence (concordance)

Pearson correlation coefficient – $\rho(Y_1, Y_2) = \frac{\mathrm{cov}\,(Y_1, Y_2)}{\sqrt{\mathrm{var}\,(Y_1)}\sqrt{\mathrm{var}\,(Y_2)}}$.

Kendall's tau –

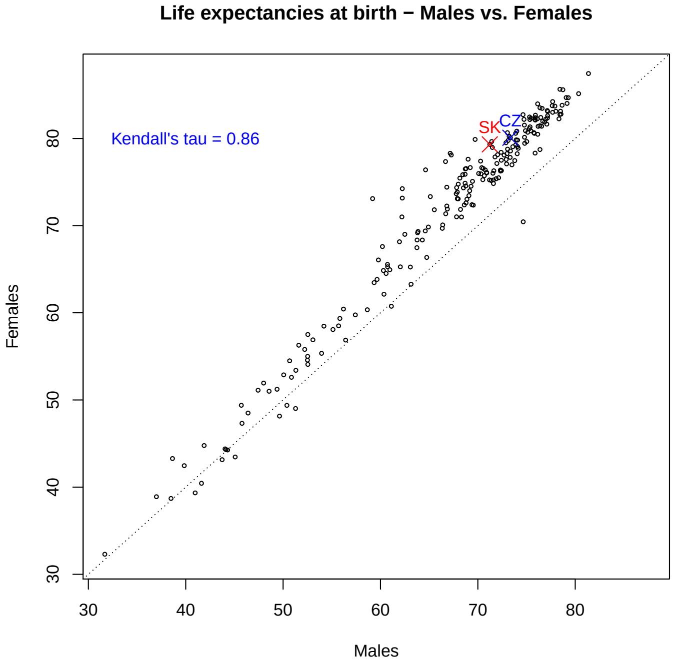
$$\tau(Y_1, Y_2) = 2\,\mathbb{P}\big[(Y_1 - Y_1')(Y_2 - Y_2') > 0\big] - 1,$$

where (Y_1', Y_2') is an independent copy of (Y_1, Y_2).

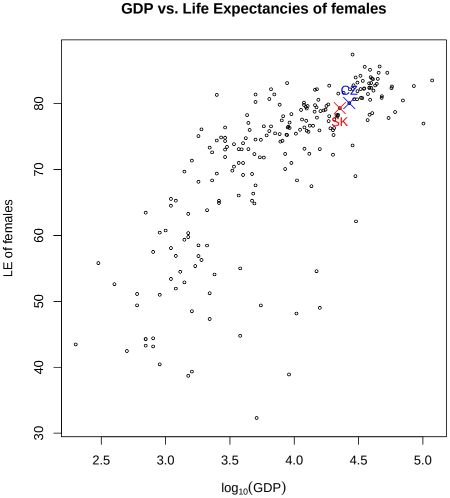
Other measures - Spearman's rho,
Blomqvist beta

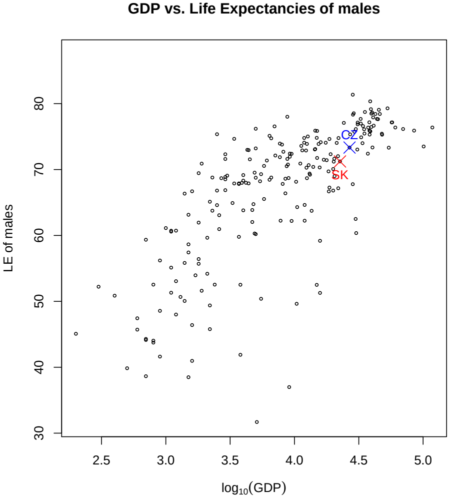
$$\beta = \mathbb{P}\left[\left(Y_1 - F_1^{-1}(0.5)\right)\left(Y_2 - F_2^{-1}(0.5)\right) > 0\right] - \mathbb{P}\left[\left(Y_1 - F_1^{-1}(0.5)\right)\left(Y_2 - F_2^{-1}(0.5)\right) < 0\right],$$

. . .


Life expectancies at birth − Males vs. Females
Kendall's tau = 0.86
SK
CZ
Females
Males
30
40
50
60
70
80

Life expectancies vs. GDP





Partial vs. conditional correlation coefficients

Partial correlation coefficient measures correlation of Y_1 and Y_2 after removal of any part of the variation due to the influence of X (*Cramér, Mathematical Methods of Statistics, 1946*).

Partial vs. conditional correlation coefficients

Partial correlation coefficient measures correlation of Y_1 and Y_2 after removal of any part of the variation due to the influence of X (*Cramér, Mathematical Methods of Statistics, 1946*).

Pearson partial correlation coefficient
Estimation: First compute 'the residuals'

$$\hat{\varepsilon}_{1i} = Y_{1i} - \hat{a}_1 - \hat{b}_1 X_i, \qquad \hat{\varepsilon}_{2i} = Y_{2i} - \hat{a}_2 - \hat{b}_2 X_i$$

where $(\hat{a}_1, \hat{b}_1)$ are the LS estimates of the parameters of the linear model $Y_1 = a_1 + b_1 X + \varepsilon_1$ and similarly for $(\hat{a}_2, \hat{b}_2)$.

Partial vs. conditional correlation coefficients

Partial correlation coefficient measures correlation of Y_1 and Y_2 after removal of any part of the variation due to the influence of X (*Cramér, Mathematical Methods of Statistics, 1946*).

Pearson partial correlation coefficient
Estimation: First compute 'the residuals'

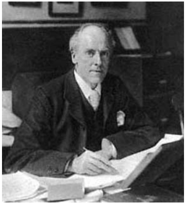
$$\hat{\varepsilon}_{1i} = Y_{1i} - \hat{a}_1 - \hat{b}_1 X_i, \qquad \hat{\varepsilon}_{2i} = Y_{2i} - \hat{a}_2 - \hat{b}_2 X_i$$

where $(\hat{a}_1, \hat{b}_1)$ are the LS estimates of the parameters of the linear model $Y_1 = a_1 + b_1 X + \varepsilon_1$ and similarly for $(\hat{a}_2, \hat{b}_2)$.

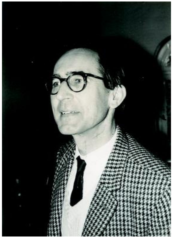
Then

$$\hat{\rho}\big((Y_1, Y_2) \,|\, X\big) := \hat{\rho}(\hat{\varepsilon}_1, \hat{\varepsilon}_2).$$

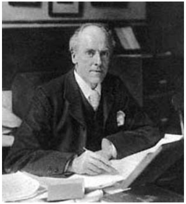
Population version of Pearson partial correlation coefficient may be also computed as



$$\rho((Y_1, Y_2)|X) = \frac{\rho(Y_1, Y_2) - \rho(Y_1, X)\,\rho(Y_2, X)}{\sqrt{1-\rho^2(Y_1, X)}\,\sqrt{1-\rho^2(Y_2, X)}}.$$

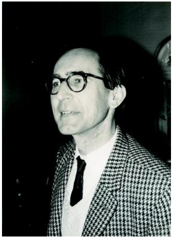


Population version of Pearson partial correlation coefficient may be also computed as



$$\rho((Y_1, Y_2)|X) = \frac{\rho(Y_1, Y_2) - \rho(Y_1, X)\,\rho(Y_2, X)}{\sqrt{1-\rho^2(Y_1, X)}\,\sqrt{1-\rho^2(Y_2, X)}}.$$

Kendall's partial tau



$$\tau((Y_1, Y_2)|X) = \frac{\tau(Y_1, Y_2) - \tau(Y_1, X)\,\tau(Y_2, X)}{\sqrt{1-\tau^2(Y_1, X)}\,\sqrt{1-\tau^2(Y_2, X)}}.$$

Kendall, M. G. (1942). Partial rank correlation. *Biometrika*, 32:277–283.

Conditional correlation coefficients

Partial correlation coefficients are measures of *'global dependence'* when adjusted for X.

Conditional correlation coefficients

Partial correlation coefficients are measures of *'global dependence'* when adjusted for X.

Conditional correlation coefficient measures correlation of Y_1 and Y_2 when $X = x$.

Conditional correlation coefficients

Partial correlation coefficients are measures of *'global dependence'* when adjusted for X.

Conditional correlation coefficient measures correlation of Y_1 and Y_2 when $X = x$.

If the conditional dependence structure of (Y_1, Y_2) given $X = x$ does not change with x, then the concepts of partial and conditional dependence coincide.

Estimating conditional measures of dependence

We observe independent triples $\{(Y_{1i}, Y_{2i}, X_i)\}_{i=1}^{n}$.

Estimating conditional measures of dependence

We observe independent triples $\{(Y_{1i}, Y_{2i}, X_i)\}_{i=1}^n$.

Rough: Take only the pairs (Y_{1i}, Y_{2i}) with X_i *'close'* to a fixed x.

Estimating conditional measures of dependence

We observe independent triples $\{(Y_{1i}, Y_{2i}, X_i)\}_{i=1}^n$.

Rough: Take only the pairs (Y_{1i}, Y_{2i}) with X_i *'close'* to a fixed x.

Refinement: The pair (Y_{1i}, Y_{2i}) is given a weight

$$w_{ni}(x, h_n) = \frac{K\left(\frac{X_i - x}{h_n}\right)}{\sum_{j=1}^n K\left(\frac{X_j - x}{h_n}\right)}, \qquad i = 1, \ldots, n,$$

where K is a given kernel, e.g. the Epanechnikov kernel

$$K(x) = \tfrac{3}{4}\,(1 - x^2)\,\mathbb{I}\{|x| \leq 1\},$$

and h_n is the width of a 'smoothing window'.

Conditional Kendall's tau

(Standard) **Kendall's tau**

$$\hat{\tau}(Y_1, Y_2) = \frac{4}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \mathbb{I}\{Y_{1i} < Y_{1j}, Y_{2i} < Y_{2j}\} - 1.$$

Conditional Kendall's tau

(Standard) **Kendall's tau**

$$\hat{\tau}(Y_1, Y_2) = \frac{4}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} \mathbb{I}\{Y_{1i} < Y_{1j}, Y_{2i} < Y_{2j}\} - 1.$$

Conditional Kendall's tau: $\hat{\tau}(x) := \hat{\tau}((Y_1, Y_2)|X = x)$

$$\hat{\tau}(x) = \frac{4}{A_n(x,h)} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ni}(x, h_n)\, w_{nj}(x, h_n)\, \mathbb{I}\{Y_{1i} < Y_{1j}, Y_{2i} < Y_{2j}\} - 1,$$

where $A_n(x,h) = 1 - \sum_{i=1}^{n} w_{ni}^2(x, h_n)$.

Which quantity are we estimating and where do copulas come in?

What is a copula?

Suppose we observe a bivariate vector (Y_1, Y_2) coming from a joint distribution function H with marginal distribution functions F_1 and F_2.

What is a copula?

Suppose we observe a bivariate vector (Y_1, Y_2) coming from a joint distribution function H with marginal distribution functions F_1 and F_2.

According to *Sklar's theorem* (see e.g. Nelsen (2006)) there exists a bivariate function C such that

$$P(Y_1 \leq y_1, Y_2 \leq y_2) = H(y_1, y_2) = C(F_1(y_1), F_2(y_2)).$$

What is a copula?

Suppose we observe a bivariate vector (Y_1, Y_2) coming from a joint distribution function H with marginal distribution functions F_1 and F_2.

According to *Sklar's theorem* (see e.g. Nelsen (2006)) there exists a bivariate function C such that

$$P(Y_1 \leq y_1, Y_2 \leq y_2) = H(y_1, y_2) = C(F_1(y_1), F_2(y_2)).$$

The function C is called a **copula** and it completely describes the dependence of (Y_1, Y_2). It is in itself a joint cumulative distribution function on $[0, 1]^2$ with uniform marginals.

Conditional copula

Suppose that in fact we observe a three-dimensional vector (Y_1, Y_2, X) and we are interested in the *conditional dependence structure* of (Y_1, Y_2) for a given value of $X = x$.

Conditional copula

Suppose that in fact we observe a three-dimensional vector (Y_1, Y_2, X) and we are interested in the *conditional dependence structure* of (Y_1, Y_2) for a given value of $X = x$.

Let

$$H_x(y_1, y_2) = \mathbb{P}(Y_1 \leq y_1, Y_2 \leq y_2 \,|\, X = x)$$

be the *conditional joint distribution function* with F_{1x} and F_{2x} being the corresponding marginals of H_x.

Conditional copula

Suppose that in fact we observe a three-dimensional vector (Y_1, Y_2, X) and we are interested in the *conditional dependence structure* of (Y_1, Y_2) for a given value of $X = x$.

Let

$$H_x(y_1, y_2) = \mathbb{P}(Y_1 \leq y_1, Y_2 \leq y_2 \,|\, X = x)$$

be the *conditional joint distribution function* with F_{1x} and F_{2x} being the corresponding marginals of H_x.

Then according to Sklar's theorem there exists a function C_x such that

$$H_x(y_1, y_2) = C_x(F_{1x}(y_1), F_{2x}(y_2)).$$

Conditional copula

Suppose that in fact we observe a three-dimensional vector (Y_1, Y_2, X) and we are interested in the *conditional dependence structure* of (Y_1, Y_2) for a given value of $X = x$.

Let

$$H_x(y_1, y_2) = \mathbb{P}(Y_1 \leq y_1, Y_2 \leq y_2 \,|\, X = x)$$

be the *conditional joint distribution function* with F_{1x} and F_{2x} being the corresponding marginals of H_x.

Then according to Sklar's theorem there exists a function C_x such that

$$H_x(y_1, y_2) = C_x(F_{1x}(y_1), F_{2x}(y_2)).$$

The function C_x is called the **conditional copula function** (Patton, 2006).

Patton, J. A. (2006). Modeling asymmetric exchange rate dependence. *International Economic Review*, 47(2):527–556.

Partial copula

A partial copula describes dependence structure of Y_1 and Y_2 after removing of any part of the variation due to the influence of X.

Partial copula

A partial copula describes dependence structure of Y_1 and Y_2 after removing of any part of the variation due to the influence of X.

Put

$$U_1 = F_{1X}(Y_1), \quad U_2 = F_{2X}(Y_2).$$

Partial copula

A partial copula describes dependence structure of Y_1 and Y_2 after removing of any part of the variation due to the influence of X.

Put

$$U_1 = F_{1X}(Y_1), \quad U_2 = F_{2X}(Y_2).$$

We suggest that the **partial copula function** is called the copula corresponding to (U_1, U_2), i.e.

$$\bar{\mathrm{C}}(u_1, u_2) \quad = \quad \mathbb{P}(U_1 \leq u_1, U_2 \leq u_2)$$

Partial copula

A partial copula describes dependence structure of Y_1 and Y_2 after removing of any part of the variation due to the influence of X.

Put

$$U_1 = F_{1X}(Y_1), \quad U_2 = F_{2X}(Y_2).$$

We suggest that the **partial copula function** is called the copula corresponding to (U_1, U_2), i.e.

$$\begin{aligned}\bar{\mathrm{C}}(u_1, u_2) &= \mathbb{P}(U_1 \le u_1, U_2 \le u_2) \\ &= \int \mathbb{P}(U_1 \le u_1,\, U_2 \le u_2 \,|\, X = x) f_X(x)\, dx \\ &= \int C_x(u_1, u_2) f_X(x)\, dx.\end{aligned}$$

That is why some researchers call $\bar{\mathrm{C}}$ an *average copula*.

Simplified pair-copula construction

Recall that

$$\bar{\mathrm{C}}(u_1, u_2) = \int C_x(u_1, u_2) f_X(x)\, dx.$$

Simplified pair-copula construction

Recall that

$$\bar{\mathrm{C}}(u_1, u_2) = \int C_x(u_1, u_2) f_X(x)\, dx.$$

Often, it is assumed that C_x does not change with x, which implies

$$\bar{\mathrm{C}} = C_x, \qquad \forall x \in \mathrm{supp}(X).$$

Acar, E., Genest, C., and Nešlehová, J. (2012). Beyond simplified pair-copula constructions. *J. Multivariate Anal.* Available online.

Hobæk Haff, I., Aas, K., and Frigessi, A. (2010). On the simplified pair-copula construction–simply useful or too simplistic? *J. Multivariate Anal.*, 101(5):1296–1310.

Different versions of Kendall's tau

(Standard) **Kendall's tau** of (Y_1, Y_2)

$$\tau = 4 \iint_{[0,1]^2} C(u_1, u_2)\, dC(u_1, u_2) - 1,$$

where C is the copula associated with (Y_1, Y_2).

Different versions of Kendall's tau

(Standard) **Kendall's tau** of (Y_1, Y_2)

$$\tau = 4 \iint_{[0,1]^2} C(u_1, u_2)\, dC(u_1, u_2) - 1,$$

where C is the copula associated with (Y_1, Y_2).

Conditional Kendall's tau of (Y_1, Y_2) given $X = x$

$$\tau(x) = 4 \iint_{[0,1]^2} C_x(u_1, u_2)\, dC_x(u_1, u_2) - 1,$$

where C_x is the conditional copula associated with (Y_1, Y_2) given $X = x$.

Different versions of Kendall's tau

Conditional Kendall's tau of (Y_1, Y_2) given $X = x$

$$\tau(x) = 4 \iint_{[0,1]^2} C_x(u_1, u_2)\, dC_x(u_1, u_2) - 1,$$

where C_x is the conditional copula associated with (Y_1, Y_2) given $X = x$.

Partial Kendall's tau of $(U_1, U_2) = (F_{1X}(Y_1), F_{2X}(Y_2))$

$$\bar{\tau} = 4 \iint_{[0,1]^2} \bar{\mathrm{C}}(u_1, u_2)\, d\bar{\mathrm{C}}(u_1, u_2) - 1,$$

where $\bar{\mathrm{C}}$ is the copula associated with (U_1, U_2).

Conditional copula estimation – overview

- **Parametric approach** – the conditional copula C_x belongs to a given parametric family whose parameters depend on the covariate through a known parametric function – Patton (2006), time-series literature, ...

Acar, E. F., Craiu, R. V., and Yao, F. (2011). Dependence calibration in conditional copulas: a nonparametric approach. *Biometrics*, 67:445–453.

Abegaz, F., Gijbels, I., and Veraverbeke, N. (2012). Semiparametric estimation of conditional copulas. *J. Multivariate Anal.*, 110:43–73.

Conditional copula estimation – overview

- **Parametric approach** – the conditional copula C_x belongs to a given parametric family whose parameters depend on the covariate through a known parametric function – Patton (2006), time-series literature, ...
- **Semiparametric approach** – the conditional copula C_x belongs to a given parametric family whose parameters depend on the covariate in an unspecified way.

Acar, E. F., Craiu, R. V., and Yao, F. (2011). Dependence calibration in conditional copulas: a nonparametric approach. *Biometrics*, 67:445–453.

Abegaz, F., Gijbels, I., and Veraverbeke, N. (2012). Semiparametric estimation of conditional copulas. *J. Multivariate Anal.*, 110:43–73.

Conditional copula estimation – overview

- **Parametric approach** – the conditional copula C_x belongs to a given parametric family whose parameters depend on the covariate through a known parametric function – Patton (2006), time-series literature, ...
- **Semiparametric approach** – the conditional copula C_x belongs to a given parametric family whose parameters depend on the covariate in an unspecified way.
  - Margins depends on the covariate in a known parametric way – Acar et al. (2011).

Acar, E. F., Craiu, R. V., and Yao, F. (2011). Dependence calibration in conditional copulas: a nonparametric approach. *Biometrics*, 67:445–453.

Abegaz, F., Gijbels, I., and Veraverbeke, N. (2012). Semiparametric estimation of conditional copulas. *J. Multivariate Anal.*, 110:43–73.

Conditional copula estimation – overview

- **Parametric approach** – the conditional copula C_x belongs to a given parametric family whose parameters depend on the covariate through a known parametric function – Patton (2006), time-series literature, ...
- **Semiparametric approach** – the conditional copula C_x belongs to a given parametric family whose parameters depend on the covariate in an unspecified way.
  - Margins depends on the covariate in a known parametric way – Acar et al. (2011).
  - Margins depends on the covariate in an unknown way – Abegaz et al. (2012).

Acar, E. F., Craiu, R. V., and Yao, F. (2011). Dependence calibration in conditional copulas: a nonparametric approach. *Biometrics*, 67:445–453.

Abegaz, F., Gijbels, I., and Veraverbeke, N. (2012). Semiparametric estimation of conditional copulas. *J. Multivariate Anal.*, 110:43–73.

Conditional copula estimation – overview

- **Parametric approach** – the conditional copula C_x belongs to a given parametric family whose parameters depend on the covariate through a known parametric function – Patton (2006), time-series literature, ...
- **Semiparametric approach** – the conditional copula C_x belongs to a given parametric family whose parameters depend on the covariate in an unspecified way.
  - Margins depends on the covariate in a known parametric way – Acar et al. (2011).
  - Margins depends on the covariate in an unknown way – Abegaz et al. (2012).
- **Nonparametric approach** – OGV.

Acar, E. F., Craiu, R. V., and Yao, F. (2011). Dependence calibration in conditional copulas: a nonparametric approach. *Biometrics*, 67:445–453.

Abegaz, F., Gijbels, I., and Veraverbeke, N. (2012). Semiparametric estimation of conditional copulas. *J. Multivariate Anal.*, 110:43–73.

Conditional copula estimation

Recall that

$$H_x(u_1, u_2) = C_x\left(F_{1x}(u_1), F_{2x}(u_2)\right).$$

where H_x is the conditional joint distribution of (Y_1, Y_2) given $X = x$ and F_{1x}, F_{2x} are the corresponding marginals.

Conditional copula estimation

Recall that

$$H_x(u_1, u_2) = C_x\left(F_{1x}(u_1), F_{2x}(u_2)\right).$$

where H_x is the conditional joint distribution of (Y_1, Y_2) given $X = x$ and F_{1x}, F_{2x} are the corresponding marginals.

The **conditional copula** C_x may be expressed as

$$C_x(u_1, u_2) = H_x\left(F_{1x}^{-1}(u_1), F_{2x}^{-1}(u_2)\right).$$

Empirical conditional copula

A straightforward estimator of the conditional copula is

$$\widehat{C}_x(u_1, u_2) = \widehat{H}_x\left(\widehat{F}_{1x}^{-1}(u_1), \widehat{F}_{2x}^{-1}(u_2)\right),$$

Empirical conditional copula

A straightforward estimator of the conditional copula is

$$\widehat{C}_x(u_1, u_2) = \widehat{H}_x\left(\widehat{F}_{1x}^{-1}(u_1), \widehat{F}_{2x}^{-1}(u_2)\right),$$

where $\widehat{H}_x$ is the *estimate of the joint conditional distribution function* H_x given by

$$\widehat{H}_x(y_1, y_2) = \sum_{i=1}^{n} w_{ni}(x, h_n)\, \mathbb{I}\{Y_{1i} \leq y_1, Y_{2i} \leq y_2\}$$

and $\widehat{F}_{1x}$ and $\widehat{F}_{2x}$ are the corresponding marginals of $\widehat{H}_x$.

Empirical conditional copula

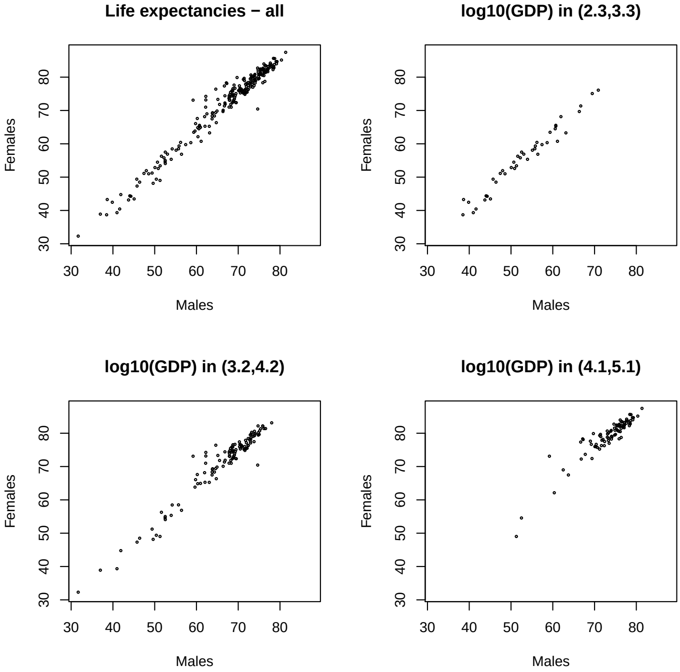
A straightforward estimator of the conditional copula is

$$\widehat{C}_x(u_1, u_2) = \widehat{H}_x\left(\widehat{F}_{1x}^{-1}(u_1), \widehat{F}_{2x}^{-1}(u_2)\right),$$

where $\widehat{H}_x$ is the *estimate of the joint conditional distribution function* H_x given by

$$\widehat{H}_x(y_1, y_2) = \sum_{i=1}^{n} w_{ni}(x, h_n)\, \mathbb{I}\{Y_{1i} \leq y_1, Y_{2i} \leq y_2\}$$

and $\widehat{F}_{1x}$ and $\widehat{F}_{2x}$ are the corresponding marginals of $\widehat{H}_x$.


Life expectancies − all
log10(GDP) in (2.3,3.3)
log10(GDP) in (3.2,4.2)
log10(GDP) in (4.1,5.1)
Females
Males
30
40
50
60
70
80

What might be a problem . . .

If X_i (GDP) is large (small), both (Y_{1i}, Y_{2i}) (life expectancies) are likely to be both large (small).

What might be a problem ...

If X_i (GDP) is large (small), both (Y_{1i}, Y_{2i}) (life expectancies) are likely to be both large (small).

$\Longrightarrow$ the effect of the of the covariate on the marginals may create an artificial (confounding) dependence.

What might be a problem ...

If X_i (GDP) is large (small), both (Y_{1i}, Y_{2i}) (life expectancies) are likely to be both large (small).

$\Longrightarrow$ the effect of the of the covariate on the marginals may create an artificial (confounding) dependence.

Conclusion: Before calculation of conditional measures of dependence, first remove the effect of the covariate on the marginal distributions.

A general transformation of the marginals

Recall that (F_{1x_i}, F_{2x_i}) are marginal distribution functions of (Y_{1i}, Y_{2i}) conditionally on $X_i = x_i$.

A general transformation of the marginals

Recall that (F_{1x_i}, F_{2x_i}) are marginal distribution functions of (Y_{1i}, Y_{2i}) conditionally on $X_i = x_i$.

If we knew (F_{1x_i}, F_{2x_i}), then by a (marginal probability integral) transformation

$$(U_{1i}, U_{2i}) = (F_{1x_i}(Y_{1i}), F_{2x_i}(Y_{2i}))\,,$$

A general transformation of the marginals

Recall that (F_{1x_i}, F_{2x_i}) are marginal distribution functions of (Y_{1i}, Y_{2i}) conditionally on $X_i = x_i$.

If we knew (F_{1x_i}, F_{2x_i}), then by a (marginal probability integral) transformation

$$(U_{1i}, U_{2i}) = (F_{1x_i}(Y_{1i}), F_{2x_i}(Y_{2i}))\,,$$

we get a vector (U_{1i}, U_{2i}) with the distribution given by the conditional copula C_{x_i}.

A general transformation of the marginals

Recall that (F_{1x_i}, F_{2x_i}) are marginal distribution functions of (Y_{1i}, Y_{2i}) conditionally on $X_i = x_i$.

If we knew (F_{1x_i}, F_{2x_i}), then by a (marginal probability integral) transformation

$$(U_{1i}, U_{2i}) = (F_{1x_i}(Y_{1i}), F_{2x_i}(Y_{2i})) \,,$$

we get a vector (U_{1i}, U_{2i}) with the distribution given by the conditional copula C_{x_i}.

Given the values of the covariate $x_1, \ldots, x_n$ we can construct pairs $(U_{11}, U_{21}), \ldots, (U_{1n}, U_{2n})$ such that :

A general transformation of the marginals

Recall that (F_{1x_i}, F_{2x_i}) are marginal distribution functions of (Y_{1i}, Y_{2i}) conditionally on $X_i = x_i$.

If we knew (F_{1x_i}, F_{2x_i}), then by a (marginal probability integral) transformation

$$(U_{1i}, U_{2i}) = (F_{1x_i}(Y_{1i}), F_{2x_i}(Y_{2i})) \,,$$

we get a vector (U_{1i}, U_{2i}) with the distribution given by the conditional copula C_{x_i}.

Given the values of the covariate $x_1, \ldots, x_n$ we can construct pairs $(U_{11}, U_{21}), \ldots, (U_{1n}, U_{2n})$ such that :

- the margins of (U_{1i}, U_{2i}) are uniform;

A general transformation of the marginals

Recall that (F_{1x_i}, F_{2x_i}) are marginal distribution functions of (Y_{1i}, Y_{2i}) conditionally on $X_i = x_i$.

If we knew (F_{1x_i}, F_{2x_i}), then by a (marginal probability integral) transformation

$$(U_{1i}, U_{2i}) = (F_{1x_i}(Y_{1i}), F_{2x_i}(Y_{2i})) \, ,$$

we get a vector (U_{1i}, U_{2i}) with the distribution given by the conditional copula C_{x_i}.

Given the values of the covariate $x_1, \ldots, x_n$ we can construct pairs $(U_{11}, U_{21}), \ldots, (U_{1n}, U_{2n})$ such that :

- the margins of (U_{1i}, U_{2i}) are uniform;
- the conditional copula is the same for original (Y_{1i}, Y_{2i}) and transformed (U_{1i}, U_{2i}).

A general transformation of the marginals - practice

Let us fix the values of the covariate $x_1, \ldots, x_n$.

A general transformation of the marginals - practice

Let us fix the values of the covariate $x_1, \ldots, x_n$.

In practice we need to estimate the unknown conditional marginal distribution functions (F_{1x_i}, F_{2x_i}).

A general transformation of the marginals - practice

Let us fix the values of the covariate $x_1, \ldots, x_n$.

In practice we need to estimate the unknown conditional marginal distribution functions (F_{1x_i}, F_{2x_i}).

To be fully 'nonparametric' we use nonparametric estimators

$$
\begin{aligned}
\widehat{F}_{1x_i g_1}(y_1) &= \sum_{j=1}^{n} w_{nj}(x_i, g_{1n}) \, \mathbb{I}\{Y_{1i} \leq y_1\}, \\
\widehat{F}_{2x_i g_2}(y_2) &= \sum_{j=1}^{n} w_{nj}(x_i, g_{2n}) \, \mathbb{I}\{Y_{2i} \leq y_2\},
\end{aligned}
$$

with g_{n1}, g_{n2} being sequences of bandwidths going to zero.

A general transformation of the marginals - practice

Let estimate the unobserved

$$(U_{1i}, U_{2i}) = (F_{1x_i}(Y_{1i}), F_{2x_i}(Y_{2i}))$$

A general transformation of the marginals - practice

Let estimate the unobserved

$$(U_{1i}, U_{2i}) = (F_{1x_i}(Y_{1i}), F_{2x_i}(Y_{2i}))$$

with

$$(\widetilde{U}_{1i}, \widetilde{U}_{2i}) = \left(\widehat{F}_{1x_i g_1}(Y_{1i}), \widehat{F}_{2x_i g_2}(Y_{2i})\right).$$

A general transformation of the marginals - practice

Let estimate the unobserved

$$(U_{1i}, U_{2i}) = (F_{1x_i}(Y_{1i}), F_{2x_i}(Y_{2i}))$$

with

$$(\widetilde{U}_{1i}, \widetilde{U}_{2i}) = \left(\widehat{F}_{1x_i g_1}(Y_{1i}), \widehat{F}_{2x_i g_2}(Y_{2i})\right).$$

Under appropriate regularity assumptions

$$\sup_{u_1,u_2} \left| \sqrt{n\,h_n} \left(\widetilde{C}_x(u_1,u_2) - \widetilde{C}_x^{(or)}(u_1,u_2) \right) \right| = o_P(1)$$

where $\widetilde{C}_x$ is the estimator of the conditional copula based on $(\widetilde{U}_{1i}, \widetilde{U}_{2i})$ and $\widetilde{C}_x^{(or)}$ is an ‘oracle estimator’ based on unobserved (U_{1i}, U_{2i}).

Partial copula estimation

Recall that

$$(U_{1i}, U_{2i}) = (F_{1x_i}(Y_{1i}), F_{2x_i}(Y_{2i}))\,.$$

Partial copula estimation

Recall that

$$(U_{1i}, U_{2i}) = (F_{1x_i}(Y_{1i}), F_{2x_i}(Y_{2i}))\,.$$

Then the 'oracle' estimator of $\bar{\mathrm{C}}$ would be an *empirical copula* estimator given by

$$C_n^{(or)}(u_1, u_2) = G_n\left(G_{1n}^{-1}(u_1), G_{2n}^{-1}(u_2)\right), \tag{1}$$

where

$$G_n(u_1, u_2) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\{U_{1i} \le u_1, U_{2i} \le u_2\}, \tag{2}$$

and G_{1n} and G_{2n} are its corresponding marginals.

Partial copula estimation

Let $(\widetilde{U}_{1i}, \widetilde{U}_{2i})$ stand for (Y_{1i}, Y_{2i}) adjusted for the effect of X_i, e.g.

$$(\widetilde{U}_{1i}, \widetilde{U}_{2i}) = \left(\widehat{F}_{1x_ig_1}(Y_{1i}), \widehat{F}_{2x_ig_2}(Y_{2i})\right).$$

Partial copula estimation

Let $(\widetilde{U}_{1i}, \widetilde{U}_{2i})$ stand for (Y_{1i}, Y_{2i}) adjusted for the effect of X_i, e.g.

$$(\widetilde{U}_{1i}, \widetilde{U}_{2i}) = \left(\widehat{F}_{1x_i g_1}(Y_{1i}), \widehat{F}_{2x_i g_2}(Y_{2i})\right).$$

Let $\widetilde{C}_n$ be the empirical copula estimator based on $(\widetilde{U}_{1i}, \widetilde{U}_{2i})$.

Partial copula estimation

Let $(\widetilde{U}_{1i}, \widetilde{U}_{2i})$ stand for (Y_{1i}, Y_{2i}) adjusted for the effect of X_i, e.g.

$$(\widetilde{U}_{1i}, \widetilde{U}_{2i}) = \left(\widehat{F}_{1x_i g_1}(Y_{1i}), \widehat{F}_{2x_i g_2}(Y_{2i})\right).$$

Let $\widetilde{C}_n$ be the empirical copula estimator based on $(\widetilde{U}_{1i}, \widetilde{U}_{2i})$.

Then we aim at the the following result:

$$\sup_{u_1,u_2} \left| \sqrt{n} \left(\widetilde{C}_n(u_1, u_2) - C_n^{(or)}(u_1, u_2) \right) \right| = o_P(1).$$

Life expectancies at birth

Let $(\widetilde{U}_{1i}, \widetilde{U}_{2i})$ stand for (Y_{1i}, Y_{2i}) adjusted for the effect of X_i.

Life expectancies at birth

Let $(\widetilde{U}_{1i}, \widetilde{U}_{2i})$ stand for (Y_{1i}, Y_{2i}) adjusted for the effect of X_i.

Partial Kendall's tau:

$$\hat{\hat{\tau}} = \frac{4}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbb{I}\{\widetilde{U}_{1i} < \widetilde{U}_{1j}, \widetilde{U}_{2i} < \widetilde{U}_{2j}\} - 1.$$

Life expectancies at birth

Let $(\widetilde{U}_{1i}, \widetilde{U}_{2i})$ stand for (Y_{1i}, Y_{2i}) adjusted for the effect of X_i.

Partial Kendall's tau:

$$\hat{\bar{\tau}} = \frac{4}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbb{I}\{\widetilde{U}_{1i} < \widetilde{U}_{1j}, \widetilde{U}_{2i} < \widetilde{U}_{2j}\} - 1.$$

Conditional Kendall's tau:

$$\hat{\tau}(x) = \frac{4}{A_n(x,h)} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ni}(x,h_n)\, w_{nj}(x,h_n)\, \mathbb{I}\{\widetilde{U}_{1i} < \widetilde{U}_{1j}, \widetilde{U}_{2i} < \widetilde{U}_{2j}\} - 1,$$

where $A_n(x,h) = 1 - \sum_{i=1}^{n} w_{ni}^2(x,h_n)$.

Life expectancies at birth

Let $(\widetilde{U}_{1i}, \widetilde{U}_{2i})$ stand for (Y_{1i}, Y_{2i}) adjusted for the effect of X_i.

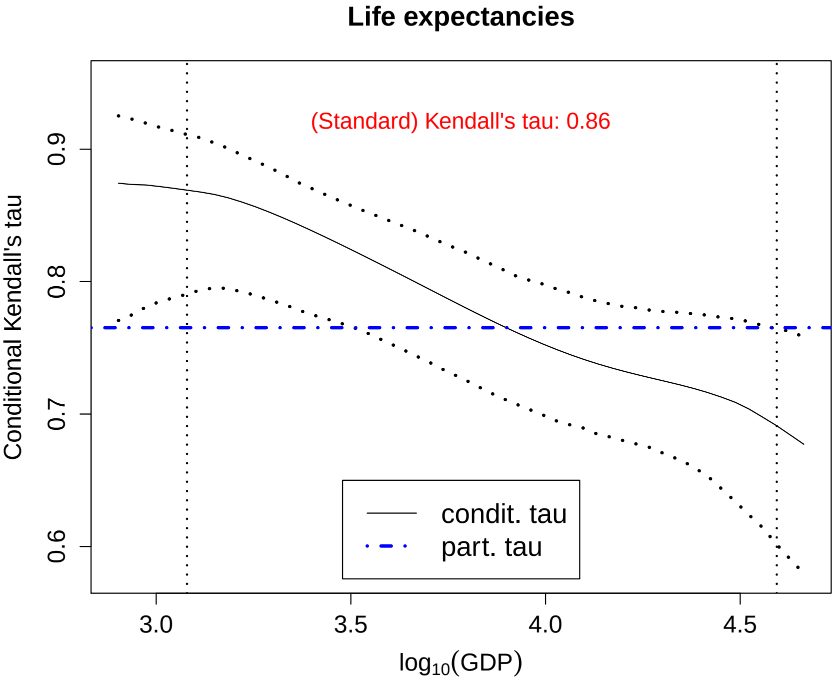
Partial Kendall's tau:

$$\hat{\bar{\tau}} = \frac{4}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbb{I}\{\widetilde{U}_{1i} < \widetilde{U}_{1j}, \widetilde{U}_{2i} < \widetilde{U}_{2j}\} - 1.$$

Conditional Kendall's tau:

$$\hat{\tau}(x) = \frac{4}{A_n(x,h)} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ni}(x,h_n)\, w_{nj}(x,h_n)\, \mathbb{I}\{\widetilde{U}_{1i} < \widetilde{U}_{1j}, \widetilde{U}_{2i} < \widetilde{U}_{2j}\} - 1,$$

where $A_n(x,h) = 1 - \sum_{i=1}^{n} w_{ni}^2(x,h_n)$.


Life expectancies
(Standard) Kendall's tau: 0.86
Conditional Kendall's tau
0.6
0.7
0.8
0.9
3.0
3.5
4.0
4.5
log10(GDP)
condit. tau
part. tau

Uranium dataset

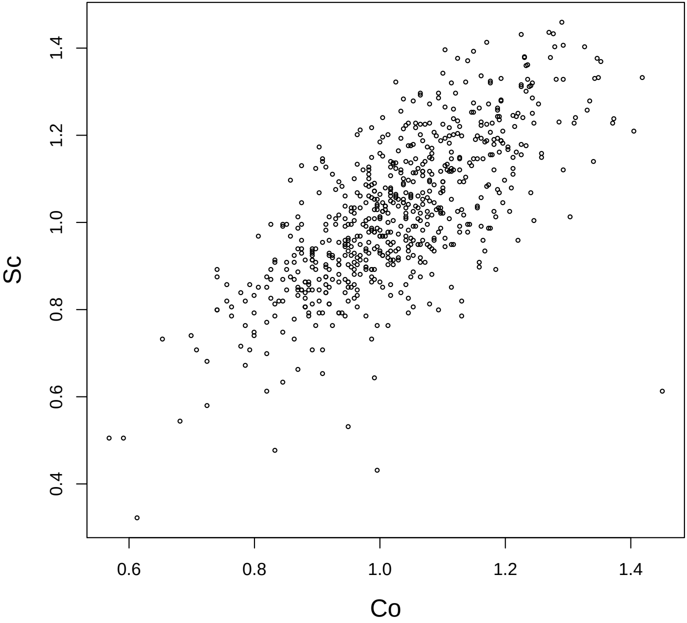
The dataset consists of the observed log-concentrations of seven chemicals in 655 water samples collected near Grand Junction, Colorado. The data can be found e.g. as a data set called `uranium` in the R-package `copula`.

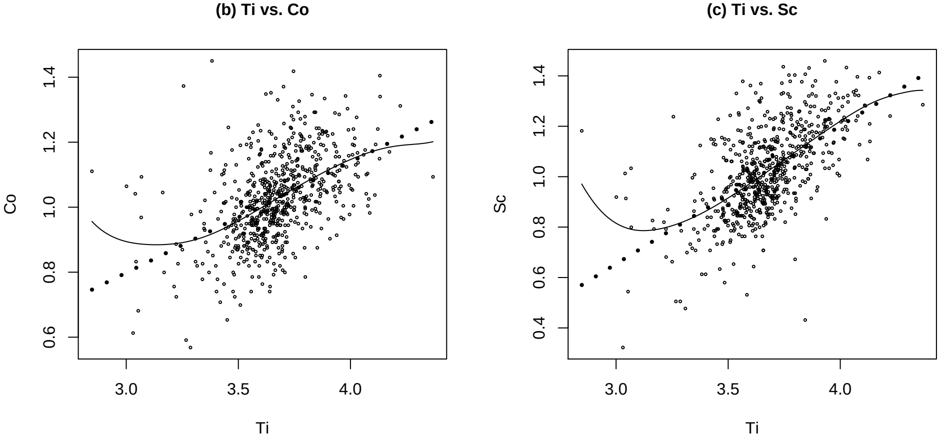
Uranium dataset

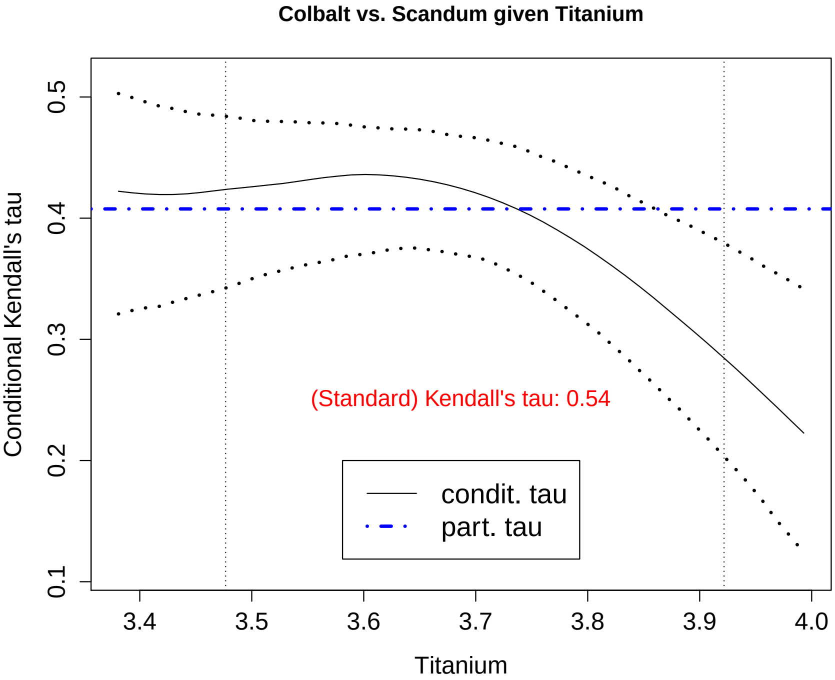
The dataset consists of the observed log-concentrations of seven chemicals in 655 water samples collected near Grand Junction, Colorado. The data can be found e.g. as a data set called `uranium` in the R-package `copula`.

Let concentrate on cobalt (Co), scandium (Sc) and titanium (Ti).

(a) Co vs. Sc




(b) Ti vs. Co
Co
Ti
0.6
0.8
1.0
1.2
1.4
3.0
3.5
4.0
(c) Ti vs. Sc
Sc
Ti
0.4
0.6
0.8
1.0
1.2
1.4
3.0
3.5
4.0


Colbalt vs. Scandum given Titanium
Conditional Kendall's tau
0.1
0.2
0.3
0.4
0.5
3.4
3.5
3.6
3.7
3.8
3.9
4.0
Titanium
(Standard) Kendall's tau: 0.54
condit. tau
part. tau

Results and conclusions

- Conditional/partial measures of correlation give us a deeper insight into the dependence structure of (Y_1, Y_2) when a covariate X is present.

Results and conclusions

- Conditional/partial measures of correlation give us a deeper insight into the dependence structure of (Y_1, Y_2) when a covariate X is present.
- With the help of conditional/partial copula one can express nonparametric conditional/partial measures of dependence (Kendall's tau, Spearman's rho, Blomqvist beta, $\ldots$).

Results and conclusions

- Conditional/partial measures of correlation give us a deeper insight into the dependence structure of (Y_1, Y_2) when a covariate X is present.
- With the help of conditional/partial copula one can express nonparametric conditional/partial measures of dependence (Kendall's tau, Spearman's rho, Blomqvist beta, $\ldots$).
- We have suggested a nonparametric estimator of a conditional copula and proved weak convergence of this estimator.

Results and conclusions

- Conditional/partial measures of correlation give us a deeper insight into the dependence structure of (Y_1, Y_2) when a covariate X is present.
- With the help of conditional/partial copula one can express nonparametric conditional/partial measures of dependence (Kendall's tau, Spearman's rho, Blomqvist beta, $\ldots$).
- We have suggested a nonparametric estimator of a conditional copula and proved weak convergence of this estimator.
- We suggested a bootstrap method for the estimator of the conditional copula.

Results and conclusions

- Conditional/partial measures of correlation give us a deeper insight into the dependence structure of (Y_1, Y_2) when a covariate X is present.
- With the help of conditional/partial copula one can express nonparametric conditional/partial measures of dependence (Kendall's tau, Spearman's rho, Blomqvist beta, ...).
- We have suggested a nonparametric estimator of a conditional copula and proved weak convergence of this estimator.
- We suggested a bootstrap method for the estimator of the conditional copula.
- We work at this moment on asymptotic properties of the nonparametric estimator of a partial copula.

A more complex covariate is possible

So far we consider X to be one-dimensional covariate, i.e. $X \in \mathbb{R}$.

A more complex covariate is possible

So far we consider X to be one-dimensional covariate, i.e. $X \in \mathbb{R}$.

Multivariate covariate – $\mathbf{X} \in \mathbb{R}^d$

$$w_{ni}(\mathbf{x}, \mathbb{H}_n) = \frac{K_{\mathbb{H}_n}(\mathbf{X}_i - \mathbf{x})}{\sum_{j=1}^{n} K_{\mathbb{H}_n}(\mathbf{X}_j - \mathbf{x})}, \quad i = 1, \ldots, n,$$

where K is a d-variate kernel, $\mathbb{H}_n$ is a bandwidth matrix with the determinant $|\mathbb{H}_n|$ and $K_{\mathbb{H}_n}(\mathbf{y}) = K(|\mathbb{H}_n|^{-1/2}\mathbf{y})$.

A more complex covariate is possible

So far we consider X to be one-dimensional covariate, i.e. $X \in \mathbb{R}$.

Multivariate covariate – $\mathbf{X} \in \mathbb{R}^d$

$$w_{ni}(\mathbf{x}, \mathbb{H}_n) = \frac{K_{\mathbb{H}_n}(\mathbf{X}_i - \mathbf{x})}{\sum_{j=1}^{n} K_{\mathbb{H}_n}(\mathbf{X}_j - \mathbf{x})}, \quad i = 1, \ldots, n,$$

where K is a d-variate kernel, $\mathbb{H}_n$ is a bandwidth matrix with the determinant $|\mathbb{H}_n|$ and $K_{\mathbb{H}_n}(\mathbf{y}) = K(|\mathbb{H}_n|^{-1/2}\mathbf{y})$.

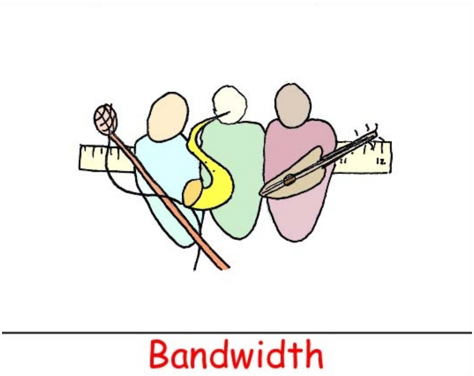
Functional covariate – $\mathcal{X} \in \mathcal{E}$

$$w_{ni}(\chi, h_n) = \frac{K\left(\frac{\|\mathcal{X}_i - \chi\|}{h_n}\right)}{\sum_{j=1}^{n} K\left(\frac{\|\mathcal{X}_j - \chi\|}{h_n}\right)}, \quad i = 1, \ldots, n,$$

where $\|\cdot\|$ stands for a norm on $\mathcal{E}$ and K is a given (univariate) kernel.

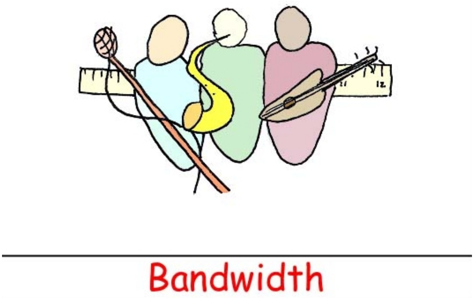
Further research

- The bandwidth selection problem. Plug-in bandwidths? Cross-validation procedure?



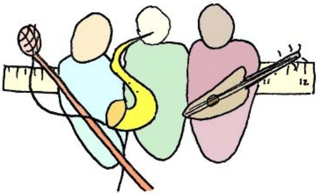
Further research

- The bandwidth selection problem. Plug-in bandwidths? Cross-validation procedure?
- Inference for the estimators of a conditional copula;

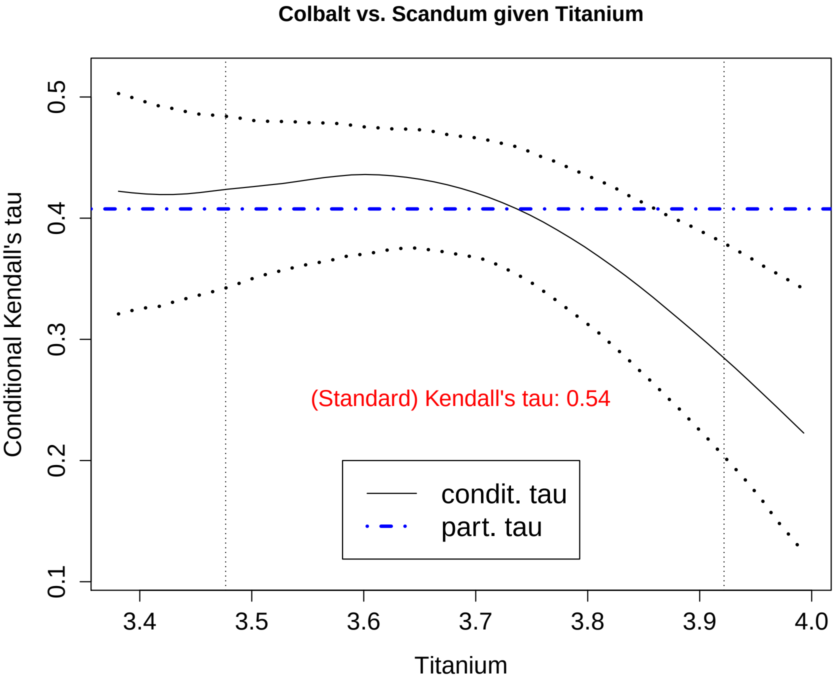


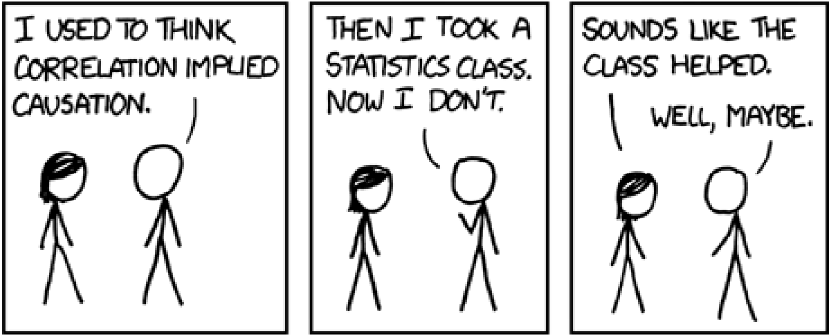
Further research

- The bandwidth selection problem. Plug-in bandwidths? Cross-validation procedure?
- Inference for the estimators of a conditional copula;
- Construction of diagnostic tests based on conditional copula estimation.



Bandwidth


Colbalt vs. Scandum given Titanium
Conditional Kendall's tau
0.1
0.2
0.3
0.4
0.5
3.4
3.5
3.6
3.7
3.8
3.9
4.0
Titanium
(Standard) Kendall's tau: 0.54
condit. tau
part. tau



Děkuji za pozornost!