Web-based Supplementary Material for Dependence Calibration in Conditional Copulas: A Nonparametric Approach

Elif F. Acar, Radu V. Craiu, and Fang Yao

Web Appendix A: Technical Details

The score and hessian functions

The score vector $\nabla\mathcal{L}(\boldsymbol{\beta}, x)$ and hessian matrix $\nabla^2\mathcal{L}(\boldsymbol{\beta}, x)$ used in the Newton-Raphson algorithm are given by, denoting $\boldsymbol{x}_{i,x} = (1, X_i - x, \ldots, (X_i - x)^p)^T$,

$$\begin{aligned}\nabla\mathcal{L}(\boldsymbol{\beta}, x) &= \frac{\partial\mathcal{L}(\boldsymbol{\beta}, x, p, h)}{\partial\boldsymbol{\beta}} \\ &= \sum_i^n \frac{\partial}{\partial\boldsymbol{\beta}}\Big\{\ell(g^{-1}(\boldsymbol{x}_{i,x}^T\boldsymbol{\beta}), U_{1i}, U_{2i})\Big\} K_h(X_i - x) \\ &= \sum_i^n \ell'(g^{-1}(\boldsymbol{x}_{i,x}^T\boldsymbol{\beta}), U_{1i}, U_{2i})\ [(g^{-1})'(\boldsymbol{x}_{i,x}^T\boldsymbol{\beta})]\ \boldsymbol{x}_{i,x}\ K_h(X_i - x),\end{aligned} \tag{1}$$

$$\begin{aligned}\nabla^2\mathcal{L}(\boldsymbol{\beta}, x) &= \frac{\partial^2\mathcal{L}(\boldsymbol{\beta}, x, p, h)}{\partial\boldsymbol{\beta}^2} \\ &= \sum_i^n \frac{\partial}{\partial\boldsymbol{\beta}}\Big\{\ell'(g^{-1}(\boldsymbol{x}_{i,x}^T\boldsymbol{\beta}), U_{1i}, U_{2i})[(g^{-1})'(\boldsymbol{x}_{i,x}^T\boldsymbol{\beta})]\ \ \boldsymbol{x}_{i,x}\Big\} K_h(X_i - x) \\ &= \sum_{i=1}^n \Big\{\ell''(g^{-1}(\boldsymbol{x}_{i,x}^T\boldsymbol{\beta}), U_{1i}, U_{2i})\ \ [(g^{-1})'(\boldsymbol{x}_{i,x}^T\boldsymbol{\beta})]^2 \\ &\qquad + \ \ell'(g^{-1}(\boldsymbol{x}_{i,x}^T\boldsymbol{\beta}), U_{1i}, U_{2i})\ (g^{-1})''(\boldsymbol{x}_{i,x}^T\boldsymbol{\beta})\Big\}\ \boldsymbol{x}_{i,x}\ \boldsymbol{x}_{i,x}^T\ K_h(X_i - x).\end{aligned} \tag{2}$$

Derivations of Asymptotic Bias and Variance

Let $H = \text{diag}\{1, h, \ldots, h^p\}$, $W = \text{diag}\{K_h(X_1 - x), \ldots, K_h(X_n - x)\}$,

$$X_x = \begin{pmatrix} 1 & X_1 - x & \cdots & (X_1 - x)^p \\ 1 & X_2 - x & \cdots & (X_2 - x)^p \\ \vdots & \vdots & & \vdots \\ 1 & X_n - x & \cdots & (X_n - x)^p \end{pmatrix}.$$

Define $(p+1)\times(p+1)$ matrices $S_n = \sum_{i=1}^n \boldsymbol{x}_{i,x}\boldsymbol{x}_{i,x}^T\ K_h(X_i - x)$ and $S_n^* = \sum_{i=1}^n \boldsymbol{x}_{i,x}\boldsymbol{x}_{i,x}^T\ K_h^2(X_i - x)$, with entries $S_{n,j} = \sum_{i=1}^n (X_i - x)^j\ K_h(X_i - x)$ and $S_{n,j}^* = \sum_{i=1}^n (X_i - x)^j\ K_h^2(X_i - x)$. The proofs of Theorem 1 and Corollary 1 rely on the following Taylor expansion

$$0 = \nabla\mathcal{L}(\hat{\boldsymbol{\beta}}, x) \approx \nabla\mathcal{L}(\boldsymbol{\beta}, x) + \nabla^2\mathcal{L}(\boldsymbol{\beta}, x)\{\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\},$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^T$ is the vector of true local parameters and $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p)^T$ is the local likelihood estimator, yielding

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \approx -\left(\nabla^2\mathcal{L}(\boldsymbol{\beta}, x)\right)^{-1} \nabla\mathcal{L}(\boldsymbol{\beta}, x).$$

Hence, the leading terms in the asymptotic conditional bias and variance can be written as

$$E(\hat{\boldsymbol{\beta}} \mid \mathbb{X}) - \boldsymbol{\beta} \approx -E\left(\nabla^2\mathcal{L}(\boldsymbol{\beta}, x) \mid \mathbb{X}\right)^{-1}\ E\left(\nabla\mathcal{L}(\boldsymbol{\beta}, x) \mid \mathbb{X}\right), \tag{3}$$

$$Var(\hat{\boldsymbol{\beta}} \mid \mathbb{X}) \approx E\left(\nabla^2\mathcal{L}(\boldsymbol{\beta}, x) \mid \mathbb{X}\right)^{-1}\ Var\left(\nabla\mathcal{L}(\boldsymbol{\beta}, x) \mid \mathbb{X}\right)\ E\left(\nabla^2\mathcal{L}(\boldsymbol{\beta}, x) \mid \mathbb{X}\right)^{-1}. \tag{4}$$

In the following we approximate three terms appearing in (3) and (4). The expectation of the kernel weighted local score function is

$$E\left(\nabla\mathcal{L}(\boldsymbol{\beta}, x) \mid \mathbb{X}\right) = \sum_{i=1}^n E\{\ \ell'\ (g^{-1}(\boldsymbol{x}_{i,x}^T\boldsymbol{\beta}), U_{1i}, U_{2i}) \mid \mathbb{X}\}\ [(g^{-1})'(\boldsymbol{x}_{i,x}^T\boldsymbol{\beta})]\ \ \boldsymbol{x}_{i,x}\ K_h(X_i - x).$$

Since, for each $i = 1, 2, \ldots, n$, we have $(U_{1i}, U_{2i}) \mid X_i \sim C(u_1, u_2 \mid g^{-1}(\eta(X_i)))$, from the first order Bartlett's identity we have

$$0 = E\{\ \ell'\ (g^{-1}(\eta(X_i)), U_{1i}, U_{2i}) \mid \mathbb{X}\} = E\{\ \ell'\ (g^{-1}(\boldsymbol{x}_{i,x}^T\boldsymbol{\beta} + r_i), U_{1i}, U_{2i}) \mid \mathbb{X}\},$$

where r_i is the approximation error made by replacing the calibration function $\eta(X_i)$ locally

by a p^{th} degree polynomial,

$$r_i = \sum_{j=p+1}^{\infty} \beta_j (X_i - x)^j = \beta_{p+1}(X_i - x)^{p+1} + o_p\{(X_i - x)^{p+1}\}.$$

If g^{-1} is continuous and twice differentiable, then

$$\ell'\ (g^{-1}(\eta(X_i)), U_{1i}, U_{2i}) \ \approx\ \ell'\ (g^{-1}(\boldsymbol{x}_{i,x}^T\boldsymbol{\beta}), U_{1i}, U_{2i}) \ +\ [\ r_i\ (g^{-1})'(\boldsymbol{x}_{i,x}^T\boldsymbol{\beta})\ ]\ \ \ell''\ (g^{-1}(\boldsymbol{x}_{i,x}^T\boldsymbol{\beta}), U_{1i}, U_{2i}),$$

and, after taking conditional expectations,

$$E\{\ell'\ (g^{-1}(\boldsymbol{x}_{i,x}^T\boldsymbol{\beta}), U_{1i}, U_{2i}) \mid \mathbb{X}\} \approx\ -\ [\ r_i\ (g^{-1})'(\boldsymbol{x}_{i,x}^T\boldsymbol{\beta})\ ]\ \ E\{\ \ell''\ (g^{-1}(\boldsymbol{x}_{i,x}^T\boldsymbol{\beta}), U_{1i}, U_{2i}) \mid \mathbb{X}\}.$$

Hence, we obtain

$$E\left(\nabla\mathcal{L}(\boldsymbol{\beta}, x) \mid \mathbb{X}\right) \approx -E\{\ \ell''\ (g^{-1}(\eta(x)), U_1, U_2) \mid x\}\ [(g^{-1})'(\eta(x))]^2\ \ X_x^T\, W\boldsymbol{r},$$

where $\boldsymbol{r} = (r_1, r_2, \ldots, r_n)^T$.

Taking conditional expectation of the hessian matrix in (2) yields

$$E\left(\nabla^2\mathcal{L}(\boldsymbol{\beta}, x) \mid \mathbb{X}\right) \approx E\left(\ell''(g^{-1}(\eta(x)), U_1, U_2) \mid x\right)[(g^{-1})'(\eta(x))]^2\ S_n.$$

The variance of the weighted local score function is

$$Var\left(\nabla\mathcal{L}(\boldsymbol{\beta}, x) \mid \mathbb{X}\right) \approx [(g^{-1})'(\eta(x))]^2\ Var\{\ \ell'\ (g^{-1}(\eta(x)), U_1, U_2)\ \mid\ x\}\ S_n^*.$$

The second order Bartlett's identity implies

$$\sigma^2(x) = -E\left(\ell''(g^{-1}(\eta(x)), U_1, U_2) \mid x\right) = Var\{\ \ell'\ (g^{-1}(\eta(x)), U_1, U_2) \mid x\}.$$

Hence, we obtain

$$\begin{aligned} E(\hat{\boldsymbol{\beta}} \mid \mathbb{X}) - \boldsymbol{\beta}\ &\approx\ S_n^{-1} X_x^T\, W\mathbf{r}, \\ Var(\hat{\boldsymbol{\beta}} \mid \mathbb{X}) &\approx\ \frac{1}{[(g^{-1})'(\eta(x))]^2\ \sigma^2(x)}\ S_n^{-1} S_n^* S_n^{-1}. \end{aligned}$$

The approximations $S_n = nf(x)H\ S\ H\ \{1 + o_p(1)\}$ and $S_n^* = nh^{-1}f(x)HS^*H\{1 + o_p(1)\}$ are outlined in Fan and Gijbels (1996). The entries of $X_x^T\, W\boldsymbol{r}$ can be written as

$$
\begin{aligned}
k_j &= \sum_{i=1}^{n} r_i (X_i - x)^j K_h(X_i - x) \\
&= \sum_{i=1}^{n} \Big\{ \sum_{m=p+1}^{\infty} \beta_m (X_i - x)^m \Big\} (X_i - x)^j K_h(X_i - x) \\
&= \beta_{p+1}\, S_{n,j+p+1} \;+\; o_p\,\{nh^{j+p+1}\}.
\end{aligned}
$$

Letting $\boldsymbol{s}_n = (S_{n,p+1}, S_{n,p+2}, \ldots, S_{n,2p+1})^T$, we obtain $X_x^T\, W\boldsymbol{r} = \beta_{p+1}\, \boldsymbol{s}_n \;+\; o_p\,\{nh^{p+1}\}$. Hence, the bias and variance expressions become

$$
\begin{aligned}
E(\hat{\boldsymbol{\beta}} \mid \mathbb{X}) - \boldsymbol{\beta} \;&\approx\; H^{-1}S^{-1}\, \boldsymbol{s}_p\, \beta_{p+1}\, h^{p+1}\{1 + o_p(1)\}, \\
Var(\hat{\boldsymbol{\beta}} \mid \mathbb{X}) &\approx\; \frac{1}{nh\; f(x)\; [(g^{-1})'(\eta(x))]^2\; \sigma^2(x)}\; H^{-1}S^{-1}\; S^*\; S^{-1}H^{-1}\; \{1 + o_p(1)\}.
\end{aligned}
$$

The result of Theorem 1 is obtained by considering the first entry in the above expressions. Since

$$
\hat{\eta}(x) - \eta(x) = g(\hat{\theta}(x)) - g(\theta(x)) \approx g'(\theta(x))\; \{\hat{\theta}(x) - \theta(x)\},
$$

we reach the result of Corollary 1 by dividing the asymptotic bias and variance of the local calibration estimator by $g'(\theta(x))$ and $[g'(\theta(x))]^2$, respectively.

Web Appendix B: Simulations Study under the Frank Copula

This additional simulation study presents the results for the data $\{(U_{1i}, U_{2i} \mid X_i) : i = 1, 2, \ldots, n\}$ generated from the Frank copula under each of the following models:

(1) Linear calibration function : $\eta(X) = 25 - 4.2\, X$

$$
(U_1, U_2) \mid X \sim C(u_1, u_2 \mid\; \theta = 25 - 4.2\, X\;) \quad \text{where } X \sim \text{Uniform}(2, 5),
$$

(2) Nonlinear calibration function: $\eta(X) = 12 + 8\sin(0.4\, X^2)$

$$
(U_1, U_2) \mid X \sim C(u_1, u_2 \mid\; \theta = 12 + 8\sin(0.4\, X^2)\;) \quad \text{where } X \sim \text{Uniform}(2, 5).
$$

The true copula parameter varies from 4 to 16.6 in the linear calibration model, and from 4 to 20 in the nonlinear one. Under each calibration model, we conduct $m = 100$ experiments with sample size n = 200. The local linear ($p = 1$) and parametric linear

estimation are performed to estimate the copula parameter under the Clayton, Frank and Gumbel copulas, and the results are converted to the Kendall's tau scale. Web Table 1 summarizes the performance of the Kendall's tau estimates in terms of integrated mean square error, integrated square bias and integrated variance.

[Web Table 1 about here.]

We construct approximate 90% pointwise condence bands for the copula parameter under the correctly selected Frank family, using half of the optimum bandwidth. Web Figure 1 displays the results in the Kendalls tau scale, averaged over 100 Monte Carlo samples. We also present the Monte Carlo based condence bands obtained from 100 estimates of the Kendalls tau.

[Web Figure 1 about here.]

Our copula selection method successfully identifies the Frank family 95% and 97% of the times under the linear and nonlinear calibration model, respectively.

Web Appendix C: Framingham Heart Study

We analyze a subset of the Framingham Heart study to illustrate our proposed methodology. The data consist of 348 participants who had stroke incidence during the follow-up study and the clinical data was collected on each participant during three examination periods, approximately 6 years apart. Of interest is the dependence structure between the log-pulse pressures of the first two examination periods, $\log(PP_1)$ and $\log(PP_2)$. It is known that chronic malnutrition may occur after stroke, particularly in severe cases that result in eating difficulties. Conversely, some patients may gain weight as a side effect of medication. Therefore, as an indicator of health status, we consider the change in the body mass index (ΔBMI) as the covariate.

For the initial investigation, we categorize the change in BMI, and focus on the three

extreme cases, namely quite stable ($-0.2 \leqslant \Delta\text{BMI} \leqslant 0.2$), highly increased ($\Delta\text{BMI} > 2$) and highly decreased ($\Delta\text{BMI} < -2$). From Web Figure 2, when there is a significant change in BMI in either direction, the dependence between the pulse pressures seems weaker compared to the case when BMI remains stable.

[Web Figure 2 about here.]

The scatterplot and histograms of the log-pulse pressures are given in Web Figure 3(a), from which the marginals are seen to be well-modeled by linear regression models assuming normal errors. We consider $U_j = \Phi((\log(\text{PP}_j) - \hat{\mu}_j(\Delta\text{ BMI}))/\hat{\sigma}_j)$ to transform the response variables to uniform scale, where $\hat{\mu}_j(\Delta\text{ BMI}))$ are obtained from fitting linear models, $\hat{\sigma}_j$ are the estimated standard errors in the linear regression models and Φ is the c.d.f. of $N(0,1)$. Web Figure 3(b) gives the scatterplot and histograms of the transformed random variables U_j, $j = 1, 2$.

[Web Figure 3 about here.]

For comparison, the calibration function is estimated using both a parametric linear model and the proposed nonparametric model with local linear ($p = 1$) specification under the Clayton, Frank and Gumbel families. In the local linear estimation, the optimum bandwidths are chosen as 9.47, 6.72, and 4.25, for the Clayton, Frank and Gumbel copulas, respectively. We constructed approximate 90% confidence intervals under each family using half of these bandwidth values. Web Figure 4 displays the results converted to the Kendall's tau scale.

[Web Figure 4 about here.]

A natural interest in modeling the conditional dependence of log-pulse pressures at different periods is to predict the latter from the previous measurement. Therefore, in copula selection, we use the cross-validated prediction errors only for the second log-pulse pressure as the selection criterion which chooses the Frank copula family. Since, at a given change in BMI,

observing two high pulse pressures together or two low pulse pressures together would be equally likely, our choice of Frank copula model seems practically sensible. The results obtained using the nonparametric approach under the Frank copula indicates that the dependence between two log-pulse pressures is stronger when BMI remains stable, as reflected by a steady health status. However, the nonlinear pattern does not seem to be significant as the parametric linear fit in fact falls within the approximate 90% pointwise confidence bands, suggesting that the linear model may be already adequate.

Web Appendix D: Birth weight dependence in twins - Revisited

The results presented in Section 4 are based on parametrically estimated conditional marginals. Here, we demonstrate the use of bootstrapping the raw data to construct confidence bands for Kendall's tau. Specifically, for each bootstrap sample we use kernel based conditional distribution estimation (Hall et al., 1999; Li and Racine, 2008) to estimate the marginals and then transform the responses to the uniform scale, as shown in Web Figure 5. Subsequently, we obtain the local linear estimate of the calibration function under the selected Frank copula model and construct the 90% quantile-based bootstrap interval based on 200 resamples. For comparison, we also plot in Web Figure 6 the asymptotic 90% confidence interval described in Section 2.3.

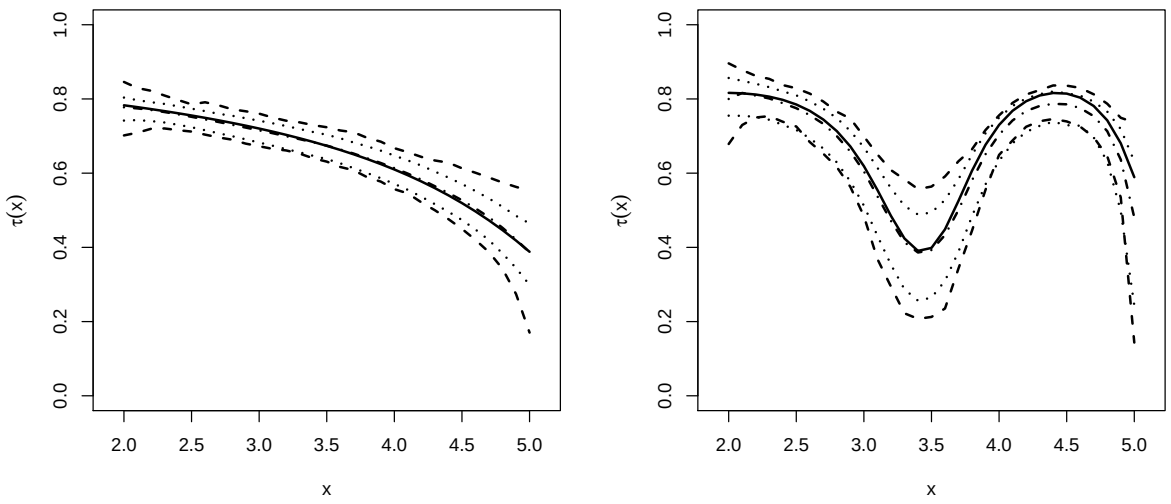
[Web Figure 5 about here.]

[Web Figure 6 about here.]

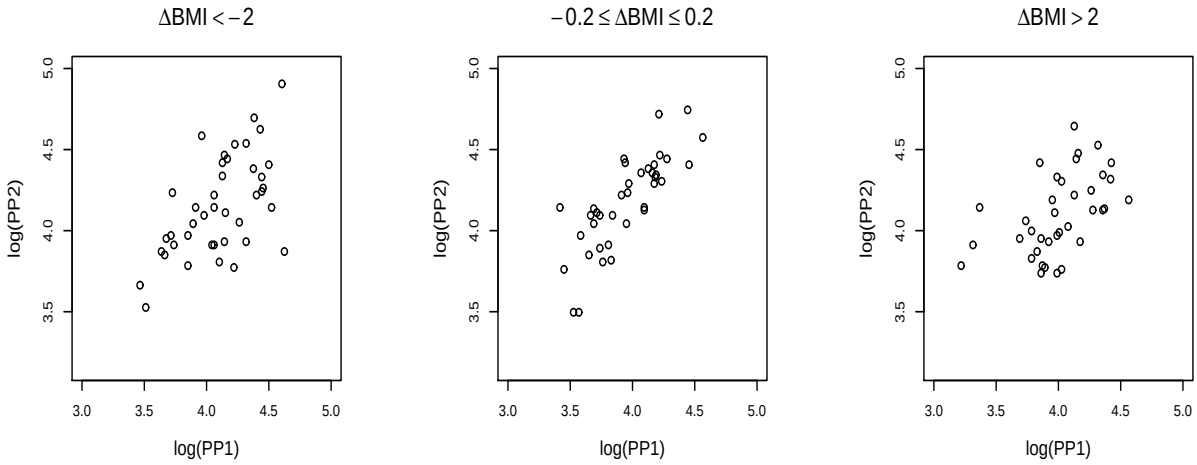
References

Hall, P., Wolff, R. C. L., and Yao, Q. (1999). Methods for estimating a conditional distribution function. *Journal of the American Statistical Association* **94,** 154–163.

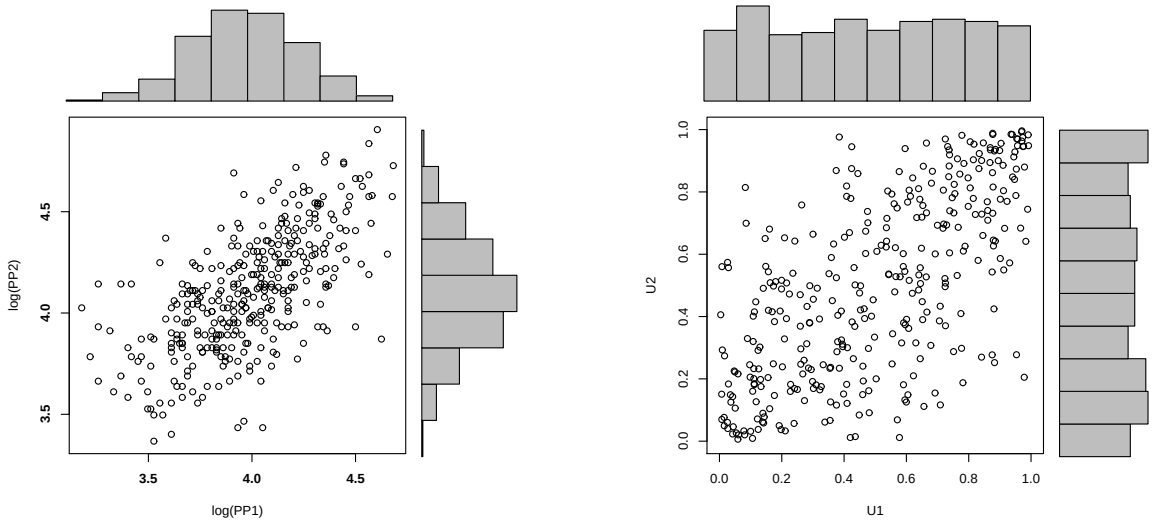
Li, Q. and Racine, J. S. (2008). Nonparametric estimation of conditional cdf and quantile functions with mixed categorical and continuous data. *Journal of Business & Economic Statistics* **26,** 423–434.



Web Figure 1. Simulation study: 90% confidence intervals for the Kendall's tau under the Frank copula with the linear (left panel) and quadratic (right panel) calibration models: truth (solid line), averaged local linear estimates (dashed line), approximate confidence intervals (dotted line), and Monte Carlo confidence intervals (dotdashed line)

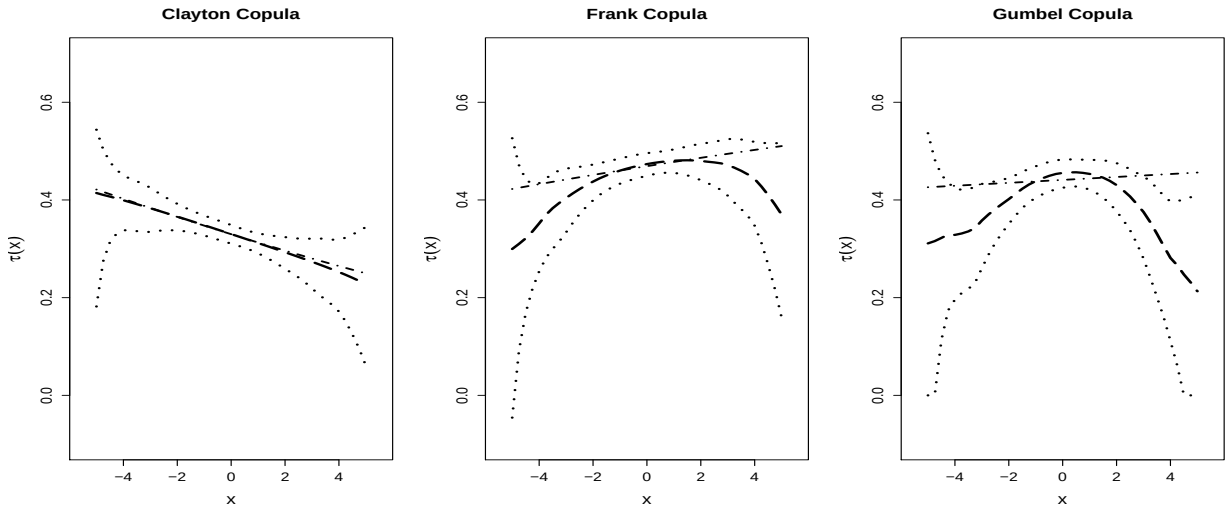


Web Figure 2. Framingham data: scatterplots of the log-pulse pressures at different covariate ranges.

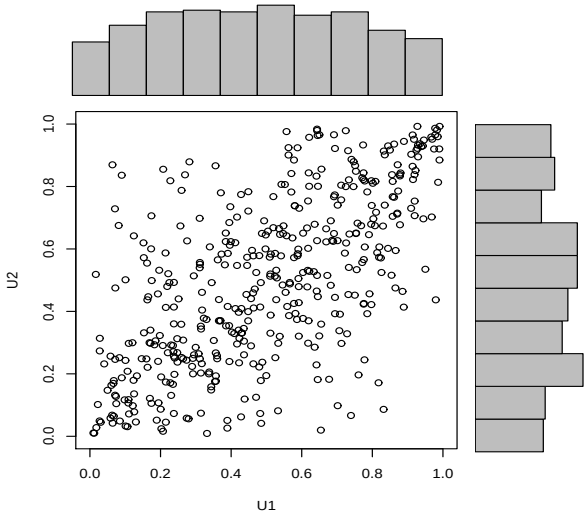


(a) (b)

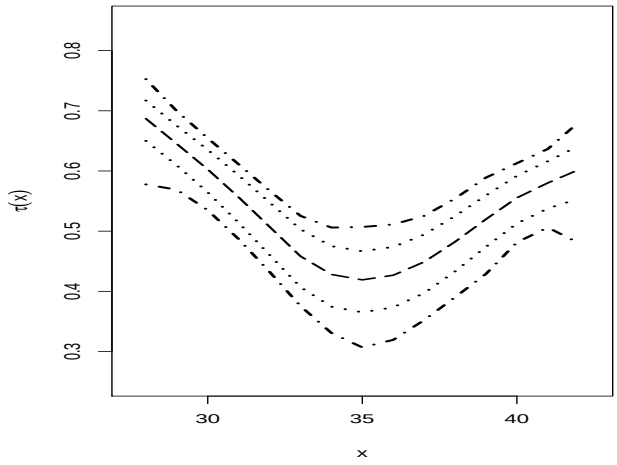
Web Figure 3. Framingham data: histograms and scatterplots of (a) the log-pulse pressures at period 1 and period 2; (b) the marginal distributions of the transformed variables (U_1, U_2).



Web Figure 4. Framingham data: the Kendalls tau estimates under three copula families: global linear estimates (dotdashed line), local linear estimates at the optimum bandwidth(longdashed line), and an approximate 90% pointwise condence bands (dotted lines).



Web Figure 5. Twin data: histograms and scatterplot of the nonparametric estimates of the conditional marginal distributions.



Web Figure 6. Twin data: the Kendall's tau estimates under the Frank copula: local linear estimates at the optimum bandwidth (longdashed line), an approximate 90% asymptotic confidence interval (dotted lines), a 90% bootstrap confidence interval (dotdashed line) based on 200 resamples.

Web Table 1

Integrated Squared Bias, Integrated Variance and Integrated Mean Square Error of the Kendall's tau estimator (multiplied by 100) and the averages of the selected bandwidths h^*.

	Linear Calibration Model						
	Parametric estimation			Local estimation			
	IBIAS^2	IVAR	IMSE	IBIAS^2	IVAR	IMSE	$\bar{h}^*$
Clayton	8.611	1.125	9.736	8.334	1.514	9.848	2.056
Frank	0.006	0.316	0.322	0.006	0.444	0.450	2.206
Gumbel	1.887	0.478	2.365	1.793	0.746	2.539	2.133
	Nonlinear Calibration Model						
	Parametric estimation			Local estimation			
	IBIAS^2	IVAR	IMSE	IBIAS^2	IVAR	IMSE	$\bar{h}^*$
Clayton	18.192	3.470	21.662	10.593	2.833	13.426	0.822
Frank	6.162	0.673	6.835	0.127	1.280	1.407	0.468
Gumbel	9.597	1.183	10.780	3.504	2.067	5.571	0.569