Supplementary Materials to "Nonparametric Estimation of Copula Regression Models with Discrete Outcomes"

A. Proofs

A.1 Proof of Consistency

Here are some simplified notations in the proof: $F_1(k_1) = F_1(k_1|X_1)$, $F_2(k_2) = F(k_2|X_2)$, $H(s,t) = H(s,t;\mathbf{X})$, $H(s,t;\theta) = H(s,t;\mathbf{X},\theta)$, and $\hat{C}(s,t) = \hat{C}(s,t;\beta)$, where β is the underlying parameter.

Proof of Lemma 2.1. Recall that $\epsilon_n \to 0$. For $v_1^k, k = 1,2,\ldots$ as in Section 2.3, taking minimum for the first n elements, $u_n = \min_{k=1,\ldots,n} v_1^k$ is a nonzero decreasing sequence. Therefore, an appropriate order of $a_n(s)$ can be chosen such that $u_{a_n(s)} > \epsilon_n$, i.e., $v_1^k > \epsilon_n$ for $k \le a_n(s)$. $\square$

To show the asymptotic properties of $\hat{C}(s,t)$ defined in (6), we analyze it by pieces. We first show the denominator is a consistent estimator of $f_{H(s,t)}(s,t)$. Then, we show the consistency of the numerator.

Denote the denominator as

$$\hat{f}_{H(s,t)}(s,t) = \frac{1}{n\epsilon_n^2}\sum_{i=1}^{n} K\left[{\epsilon_n}^{-1}(H_{i1}(s)-s), {\epsilon_n}^{-1}(H_{i2}(t)-t)\right].$$

Lemma A.1 shows the consistency of the denominator.

Lemma A.1. *Under Assumptions 2.2 and 2.1,*

$$\hat{f}_{H(s,t)}(s,t) \to_p f_{H(s,t)}(s,t). \tag{A.1}$$

Proof. Recall that K is a bounded on compact support. Without loss of generality, in the proof we assume $K(u,v) \leq 1$ with support $|(u,v)| \leq 1$.

Let $\hat{f}_{H(s,t)}(s,t) = \frac{1}{n}\sum_{i=1}^{n} T_{ni}$, where $T_{ni} = 1/\epsilon_n^2 K\left[{\epsilon_n}^{-1}(H_{i1}(s)-s), {\epsilon_n}^{-1}(H_{i2}(t)-t)\right]$. That is, $\hat{f}_{H(s,t)}(s,t)$ is the summation of a triangular array. We demonstrate the consistency of $\hat{f}_{H(s,t)}(s,t)$ using the weak law of large numbers (WLLN) for triangular arrays. It is sufficient to show

$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}T_{ni} \to f_{H(s,t)}(s,t), \tag{A.2}$$

$$\frac{1}{n}\mathrm{E}T_{ni}^2 \to 0. \tag{A.3}$$

First, to show (A.2), we divide the range of μ_1 and μ_2 into four pieces, i.e.

$$\begin{aligned}
\mathrm{E}T_{ni} =& \mathrm{E}\left[T_{ni}1\left(\mu_1 \leq M_1^{a_n(s)}, \mu_2 \leq M_2^{b_n(t)}\right)\right] + \mathrm{E}\left[T_{ni}1\left(\mu_1 \leq M_1^{a_n(s)}, \mu_2 > M_2^{b_n(t)}\right)\right] \\
&+ \mathrm{E}\left[T_{ni}1\left(\mu_1 > M_1^{a_n(s)}, \mu_2 \leq M_2^{b_n(t)}\right)\right] + \mathrm{E}\left[T_{ni}1\left(\mu_1 > M_1^{a_n(s)}, \mu_2 > M_2^{b_n(t)}\right)\right] \\
:=& T_1 + T_2 + T_3 + T_4.
\end{aligned} \tag{A.4}$$

We analyze the four pieces one by one.

Let f_{μ_1,μ_2} denote the joint density of (μ_1, μ_2). The first term of (A.4) equals

$$T_1 = \sum_{k_1=0}^{a_n(s)} \sum_{k_2=0}^{b_n(t)} T_1(k_1, k_2),$$

where

$$T_1(k_1,k_2) = \frac{1}{\epsilon_n^2}\int_{M_1^{k_1}}^{M_1^{k_1+1}}\int_{M_2^{k_2}}^{M_2^{k_2+1}} K\left[{\epsilon_n}^{-1}(F_1(k_1)-s), {\epsilon_n}^{-1}(F_2(k_2)-t)\right] f_{\mu_1,\mu_2}(\mu_1,\mu_2)\mathrm{d}\mu_1\mathrm{d}\mu_2. \tag{A.5}$$

Recall that K takes nonzero values only when

$$|(F_1(k_1)-s, F_2(k_2)-t)| \leq \epsilon_n. \tag{A.6}$$

For $k_1 \leq a_n(s), k_2 \leq b_n(t)$, a necessary condition for (A.6) to hold is that $M_1^{k_1} \leq \mu_1 < M_1^{k_1+1}, M_2^{k_2} \leq \mu_2 < M_2^{k_2+1}$. Therefore, we can write the limits in (A.5) as $(-\infty,\infty)$. That is,

$$\begin{aligned}
T_1(k_1,k_2) =& \frac{1}{\epsilon_n^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K\left[{\epsilon_n}^{-1}(F_1(k_1)-s), {\epsilon_n}^{-1}(F_2(k_2)-t)\right] f_{\mu_1,\mu_2}(\mu_1,\mu_2)\mathrm{d}\mu_1\mathrm{d}\mu_2\\
=& \frac{1}{\epsilon_n^2}\int K\left(\frac{a-s}{\epsilon_n},\frac{b-t}{\epsilon_n}\right) f_{F_1(k_1),F_2(k_2)}(a,b)\mathrm{d}a\mathrm{d}b\\
=& \int K(u,v) f_{F_1(k_1),F_2(k_2)}(s+u\epsilon_n, t+v\epsilon_n)\mathrm{d}u\mathrm{d}v,
\end{aligned}$$

where the last equation is derived by substitution. A Taylor series expansion up to first order for $f_{F_1(k_1),F_2(k_2)}$ yields

$$\begin{aligned}
&f_{F_1(k_1),F_2(k_2)}(s+u\epsilon_n, t+v\epsilon_n)\\
=&f_{F_1(k_1),F_2(k_2)}(s,t) + f_{F_1(k_1),F_2(k_2),1}(s,t)u\epsilon_n + f_{F_1(k_1),F_2(k_2),2}(s,t)v\epsilon_n +\\
&\frac{1}{2}f_{F_1(k_1),F_2(k_2),11}(s,t)u^2\epsilon_n^2 + \frac{1}{2}f_{F_1(k_1),F_2(k_2),22}(s,t)v^2\epsilon_n^2 + f_{F_1(k_1),F_2(k_2),12}(s,t)uv\epsilon_n^2 + o(\epsilon_n^2).
\end{aligned}$$

Since K is symmetric, $\int K(u,v)u\mathrm{d}u\mathrm{d}v = 0$, and $\int K(u,v)uv\mathrm{d}u\mathrm{d}v = 0$. Moreover, $\int K(u,v)\mathrm{d}u\mathrm{d}v = 1$. Denote $f_{H(s,t),n}(s,t) = \sum_{k_1=0}^{a_n(s)}\sum_{k_2=0}^{b_n(t)} f_{F_1(k_1),F_2(k_2)}(s,t)$, then we have

$$T_1 = f_{H(s,t),n}(s,t) + o(\epsilon_n),$$

where the residual term is $o(\epsilon_n)$ since $\sum_{k_1=0}^{a_n(s)}\sum_{k_2=0}^{b_n(t)} f_{F_1(k_1),F_2(k_2),jj}(s,t) < \infty, j = 1, 2$, where $f_{F_1(k_1),F_2(k_2),jj}$ is the second order derivatives of $f_{F_1(k_1),F_2(k_2)}$ with respect to jth component, by Assumption 2.2. As $a_n(s)$ and $b_n(t)$ go to infinity with n,

$$T_1 \to f_{H(s,t)}(s,t).$$

Then, we consider the second term T_2. As $|K(u,v)| \leq 1$,

$$T_2 \leq \frac{1}{\epsilon_n^2}\int_{-\infty}^{\infty}\int_{M_2^{b_n(t)}}^{\infty} f_{\mu_1,\mu_2}(\mu_1,\mu_2)\mathrm{d}\mu_1\mathrm{d}\mu_2 = \frac{1}{\epsilon_n^2}P\left(\mu_2 > M_2^{b_n(t)}\right). \tag{A.7}$$

By Assumption 2.1, $T_2 \to 0$. Similar arguments can be used to show that

$$T_3 \leq \frac{1}{\epsilon_n^2}P\left(\mu_1 > M_1^{a_n(s)}\right) \to 0,$$
$$T_4 \leq \frac{1}{\epsilon_n^2}P\left(\mu_1 > M_1^{a_n(s)}, \mu_2 > M_2^{b_n(s)}\right) \to 0.$$

So (A.2) follows immediately.

Finally, we show (A.3). Since K is bounded by 1,

$$T_{ni}^2 \leq \frac{1}{\epsilon_n^2}T_{ni}.$$

As T_{ni} is positive, combining with (A.2) and that $n\epsilon_n^2 \to \infty$, we have

$$\frac{1}{n}\mathrm{E}T_{ni}^2 \leq \frac{1}{n\epsilon_n^2}\mathrm{E}T_{ni} \to 0.$$

The lemma follows the WLLN for triangular arrays. □

Proof of Theorem 2.1. Given Lemma A.1, it is sufficient to show that the numerator of (6) is a consistent estimator of $f_{H(s,t)}(s,t)C(s,t)$. Similar to the denominator, the numerator of

(6), denoted as

$$\hat{N}(s,t) = \frac{1}{n\epsilon_n^2}\sum_{i=1}^{n} K\left[{\epsilon_n}^{-1}(H_{i1}(s)-s), {\epsilon_n}^{-1}(H_{i2}(t)-t)\right] Y_i(\beta) \tag{A.8}$$

is the summation of a triangular array, i.e.,

$$\hat{N}(s,t) = \frac{1}{n}\sum_{i=1}^{n} V_{ni},$$

where $V_{ni} = 1/\epsilon_n^2 K\left[{\epsilon_n}^{-1}(H_{i1}(s)-s), {\epsilon_n}^{-1}(H_{i2}(t)-t)\right] Y_i(\beta)$. It is sufficient to show

$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}V_{ni} \to f_{H(s,t)}(s,t)C(s,t), \tag{A.9}$$

$$\frac{1}{n}\mathrm{E}V_{ni}^2 \to 0. \tag{A.10}$$

We first show (A.9). Note that

$$\mathrm{E}\left[V_{ni}|H_{i1}(s),H_{i2}(t)\right] = K\left[{\epsilon_n}^{-1}(H_{i1}(s)-s), {\epsilon_n}^{-1}(H_{i2}(t)-t)\right] C\left[H_{i1}(s),H_{i2}(t)\right].$$

Similar ideas used to show (A.2) lead to the approximation

$$\mathrm{E}\left(V_{ni}\right) \approx \sum_{k_1=0}^{a_n(s)}\sum_{k_2=0}^{b_n(t)} \int K(u,v) f_{F_1(k_1),F_2(k_2)}(s+u\epsilon_n, t+v\epsilon_n)C(s+u\epsilon_n, t+v\epsilon_n)\mathrm{d}u\mathrm{d}v.$$

Taking the product of Taylor expansions of $f_{F_1(k_1),F_2(k_2)}$ and C at (s,t) yields

$$\mathrm{E}\left(V_{ni}\right) \approx \sum_{k_1=0}^{a_n(s)}\sum_{k_2=0}^{b_n(t)} C(s,t) f_{F_1(k_1),F_2(k_2)}(s,t) + o(\epsilon_n).$$

Note that some terms are eliminated due to the symmetry of K. When n approaches infinity, (A.9) follows immediately. In addition, since K and $Y_i(\beta)$ are bounded by 1,

$$\frac{1}{n}\mathrm{E}\left(V_{ni}^2\right) \leq \frac{1}{n\epsilon_n^2}\mathrm{E}\left(V_{ni}\right) \to 0.$$

So (A.10) holds and the stated result follows WLLN for triangular arrays. $\square$

A.2 Proof of Theorem 2.2

Proof of Theorem 2.2. Given consistency of the denominator in Section A.1, now we are in a position to show the weak convergence of the numerator. We check the bias and variance of the numerator and show they both converge to 0 at the appropriate rate.

Comparing $\hat{C}(s,t)$ and $C(s,t)$ yields

$$\hat{C}(s,t) = C(s,t) + \frac{\hat{m}_1(s,t)}{\hat{f}_{H(s,t)}(s,t)} + \frac{\hat{m}_2(s,t)}{\hat{f}_{H(s,t)}(s,t)},$$

where

$$\hat{m}_1(s,t) = \frac{1}{n\epsilon_n^2}\sum_{i=1}^{n} K\left[\epsilon_n^{-1}\left(H_{i1}(s)-s\right), {\epsilon_n}^{-1}(H_{i2}(t)-t)\right]\left[C\left[H_{i1}(s), H_{i2}(t)\right] - C(s,t)\right],$$
$$\hat{m}_2(s,t) = \frac{1}{n\epsilon_n^2}\sum_{i=1}^{n} K\left[\epsilon_n^{-1}\left(H_{i1}(s)-s\right), {\epsilon_n}^{-1}(H_{i2}(t)-t)\right]\left\{Y_i(\beta) - C\left[H_{i1}(s), H_{i2}(t)\right]\right\}.$$

Among them, $\hat{m}_1(s,t)$ contributes to the bias while $\hat{m}_2(s,t)$ contributes the variance of $\hat{C}(s,t)$.

Variance. Since $\mathrm{E}\left(Y_i(\beta)|\mu_{i1},\mu_{i2}\right) = C\left[H_{i1}(s), H_{i2}(t)\right]$, one has $\mathrm{E}(\hat{m}_2(s,t)|\mu_{i1},\mu_{i2}) = 0$, which leads to that $\mathrm{E}\left[\hat{m}_2(s,t)\right] = 0$. Thus, we focus on the variance of $\hat{m}_2(s,t)$. Note that

$$\mathrm{Var}\left[\hat{m}_2(s,t)\right] = \frac{1}{n\epsilon_n^4}\mathrm{E}\left[K\left[\epsilon_n^{-1}\left(H_{i1}(s)-s\right), {\epsilon_n}^{-1}(H_{i2}(t)-t)\right]\left\{Y_i(\beta) - C\left[H_{i1}(s), H_{i2}(t)\right]\right\}\right]^2 \tag{A.11}$$

To compute the variance of $\hat{m}_2(s,t)$, we condition on $H_i(s,t)$

$$\mathrm{E}\left(\{Y_i(\beta) - C\left[H_{i1}(s), H_{i2}(t)\right]\}^2 \left| H_{i1}(s) = a, H_{i2}(t) = b\right.\right) = C(a,b)\left[1 - C(a,b)\right] \coloneqq \sigma^2(a,b).$$

Since K and σ are bounded, arguments analogous to those use to prove (A.2) lead to the approximation

$$\begin{aligned}
&\mathrm{Var}\left[\hat{m}_2(s,t)\right] \\
\approx&\frac{1}{n\epsilon_n^4}\sum_{k_1=0}^{a_n(s)}\sum_{k_2=0}^{b_n(t)}\int_{M_1^{k_1}}^{M_1^{k_1+1}}\int_{M_2^{k_2}}^{M_2^{k_2+1}} K\left[{\epsilon_n}^{-1}(F_1(k_1)-s), {\epsilon_n}^{-1}(F_2(k_2)-t)\right]^2\sigma^2\left[F_1(k_1), F_2(k_2)\right]\cdot \\
&\qquad f_{\mu_1,\mu_2}(\mu_1,\mu_2)\mathrm{d}\mu_1\mathrm{d}\mu_2 + o\left(\frac{1}{n\epsilon_n^2}\right) \\
=&\frac{1}{n\epsilon_n^4}\sum_{k_1=0}^{a_n(s)}\sum_{k_2=0}^{b_n(t)}\int K\left(\frac{a-s}{\epsilon_n}, \frac{b-t}{\epsilon_n}\right)^2\sigma^2(a,b)f_{F_1(k_1),F_2(k_2)}(a,b)\mathrm{d}a\mathrm{d}b + o\left(\frac{1}{n\epsilon_n^2}\right) \\
=&\frac{1}{n\epsilon_n^2}\sum_{k_1=0}^{a_n(s)}\sum_{k_2=0}^{b_n(t)}\int K(u,v)^2\sigma^2(s+u\epsilon_n, t+v\epsilon_n)f_{F_1(k_1),F_2(k_2)}(s+u\epsilon_n, t+v\epsilon_n)\mathrm{d}u\mathrm{d}v + o\left(\frac{1}{n\epsilon_n^2}\right).
\end{aligned}$$

As C and $f_{F_1(k_1),F_2(k_2)}$ are twice continuously differentiable over V, $\sigma^2 f_{F_1(k_1),F_2(k_2)}$ carries over the differentiability as a function of them. Denote $\alpha_{21}(s,t) = R_2(K)\sigma^2(s,t)f_{H(s,t),n}(s,t)$. The Taylor series expansion of $\sigma^2 f_{F_1(k_1),F_2(k_2)}$ yields

$$\mathrm{Var}\left[\hat{m}_2(s,t)\right] = \frac{1}{n\epsilon_n^2}\alpha_{21}(s,t) + o\left(\frac{1}{n\epsilon_n^2}\right).$$

Note that the term $\hat{m}_2(s,t)$ is the summation of a triangular array. To establish the asymptotic distribution of $\hat{m}_2(s,t)$, we now verify that Lyapunov's central limit theorem holds. Denote

$$w_{ni}(s,t) = \frac{1}{n\epsilon_n^2}K\left[\epsilon_n^{-1}\left(H_{i1}(s)-s\right), {\epsilon_n}^{-1}(H_{i2}(t)-t)\right]\{Y_i(\beta) - C\left[H_{i1}(s), H_{i2}(t)\right]\},$$

then $\hat{m}_2(s,t) = \sum_{i=1}^n w_{ni}$. It is sufficient to show

$$\frac{\sum_{i=1}^n \mathrm{E}\left|w_{ni}(s,t)\right|^3}{\left(\sum_{i=1}^n \mathrm{Var}\left[w_{ni}(s,t)\right]\right)^{3/2}} \to 0. \tag{A.12}$$

Note that

$$\mathrm{Var}\left[w_{ni}(s,t)\right] = \frac{1}{n^2\epsilon_n^2}\alpha_{21}(s,t) + o\left(\frac{1}{n^2\epsilon_n^2}\right),$$
$$\mathrm{E}\left|w_{ni}(s,t)\right|^3 \leq \frac{f_{H(s,t),n}(s,t)}{n^3\epsilon_n^4}\int K(u,v)^3 \mathrm{d}u\mathrm{d}v + o\left(\frac{1}{n^3\epsilon_n^4}\right) = O\left(\frac{1}{n^3\epsilon_n^4}\right).$$

Hence, (A.12) follows immediately, and the Lyapunov condition is satisfied. Therefore,

$$\sqrt{n\epsilon_n^2}\frac{\hat{m}_2(s,t)}{\sqrt{\alpha_{21}(s,t)}} \to_d N(0,1).$$

Since $f_{H(s,t),n}(s,t) \to f_{H(s,t)}(s,t)$,

$$\sqrt{n\epsilon_n^2}\hat{m}_2(s,t) \to_d N\left(0, R_2(K)C(s,t)\left[1-C(s,t)\right]f_{H(s,t)}(s,t)\right).$$

By Slutsky's theorem, substitution of $\hat{f}_{H(s,t)}(s,t)$ by $f_{H(s,t)}(s,t)$ yields

$$\sqrt{n\epsilon_n^2}\frac{\hat{m}_2(s,t)}{\hat{f}_{H(s,t)}(s,t)} \to_d N\left(0, \frac{C(s,t)\left[1-C(s,t)\right]R_2(K)}{f_{H(s,t)}(s,t)}\right). \tag{A.13}$$

Bias. The mean of $\hat{m}_1(s,t)$ is

$$\mathrm{E}\left[\hat{m}_1(s,t)\right] = \frac{1}{\epsilon_n^2}\mathrm{E}\left\{K\left[\epsilon_n^{-1}\left(H_{i1}(s)-s\right), {\epsilon_n}^{-1}(H_{i2}(t)-t)\right]\left[C\left[H_{i1}(s), H_{i2}(t)\right] - C(s,t)\right]\right\}.$$

Since C satisfies Lipschitz condition as in Assumption 2.4, given $|(H_{i1}(s)-s, H_{i2}(t)-t)| < \epsilon_n$, under which $K\left[\epsilon_n^{-1}\left(H_{i1}(s)-s\right), {\epsilon_n}^{-1}(H_{i2}(t)-t)\right]$ is nonzero,

$$\left|C\left[H_{i1}(s), H_{i2}(t)\right] - C(s,t)\right| \leq \alpha_1\epsilon_n^2.$$

Similar arguments used to show (A.2) lead to the approximation

$$
\begin{aligned}
\mathrm{E}\left[\hat{m}_1(s,t)\right] \approx & \frac{1}{\epsilon_n^2} \sum_{k_1=0}^{a_n(s)} \sum_{k_2=0}^{b_n(t)} \int K\left(\frac{a-s}{\epsilon_n}, \frac{b-t}{\epsilon_n}\right) \left[C(a,b) - C(s,t)\right] f_{F_1(k_1),F_2(k_2)}(a,b)\mathrm{d}a\mathrm{d}b + o(\epsilon_n^2) \\
= & \sum_{k_1=0}^{a_n(s)} \sum_{k_2=0}^{b_n(t)} \int K(u,v) \left[C(s+u\epsilon_n, t+v\epsilon_n) - C(s,t)\right] \cdot \\
& f_{F_1(k_1),F_2(k_2)}(s,t)(s+u\epsilon_n, t+v\epsilon_n)\mathrm{d}u\mathrm{d}v + o(\epsilon_n^2).
\end{aligned}
$$

Taking the product of the Taylor series expansions for C up to second order and first order for $f_{F_1(k_1),F_2(k_2)}$ at (s,t), one gets

$$
\begin{aligned}
\mathrm{E}\left[\hat{m}_1(s,t)\right] = \sum_{k_1=0}^{a_n(s)} \sum_{k_2=0}^{b_n(t)} \int & \left\{ K(u,v)\epsilon_n^2 \left[\frac{1}{2} C_{11}(s,t) f_{F_1(k_1),F_2(k_2)}(s,t)u^2 + \right.\right. \\
& \frac{1}{2} C_{22}(s,t) f_{F_1(k_1),F_2(k_2)}(s,t)v^2 + C_1(s,t) f_{F_1(k_1),F_2(k_2),1}(s,t)u^2 + \\
& \left.\left. C_2(s,t) f_{F_1(k_1),F_2(k_2),2}(s,t)v^2 \right] \right\} \mathrm{d}u\mathrm{d}v + o(\epsilon_n^2).
\end{aligned}
$$

Let $f_{H(s,t),n,j}(s,t), j = 1,2$ denote the partial derivatives of $f_{H(s,t),n}(s,t)$ and define

$$
\zeta_n(s,t) = \frac{1}{2} C_{11}(s,t) + \frac{1}{2} C_{22}(s,t) + \frac{C_1(s,t) f_{H(s,t),n,1}(s,t)}{f_{H(s,t),n}(s,t)} + \frac{C_2(s,t) f_{H(s,t),n,2}(s,t)}{f_{H(s,t),n}(s,t)}.
$$

Recall the notation $\kappa_2 = \int u^2 K(u,v)\mathrm{d}u$, hence we obtain

$$
\mathrm{E}\left[\hat{m}_1(s,t)\right] = \kappa_2 \zeta_n(s,t) f_{H(s,t),n}(s,t)\epsilon_n^2 + o(\epsilon_n^2). \tag{A.14}
$$

We now compute the variance of $\hat{m}_1(s,t)$

$$
\begin{aligned}
&\mathrm{Var}\left[\hat{m}_1(s,t)\right]\\
=&\frac{1}{n\epsilon_n^4}\mathrm{Var}\left\{K\left[\epsilon_n^{-1}\left(H_{i1}(s)-s\right),{\epsilon_n}^{-1}(H_{i2}(t)-t)\right]\left[C\left[H_{i1}(s),H_{i2}(t)\right]-C(s,t)\right]\right\}\\
=&\frac{1}{n\epsilon_n^4}\Big(\mathrm{E}\left\{K\left[\epsilon_n^{-1}\left(H_{i1}(s)-s\right),{\epsilon_n}^{-1}(H_{i2}(t)-t)\right]\left[C\left[H_{i1}(s),H_{i2}(t)\right]-C(s,t)\right]\right\}^2\\
-&\left\{\mathrm{E}\left(K\left[\epsilon_n^{-1}\left(H_{i1}(s)-s\right),{\epsilon_n}^{-1}(H_{i2}(t)-t)\right]\left[C\left[H_{i1}(s),H_{i2}(t)\right]-C(s,t)\right]\right)\right\}^2\Big).
\end{aligned}
\tag{A.15}
$$

The first term of (A.15) is

$$
\begin{aligned}
&\frac{1}{n\epsilon_n^4}\mathrm{E}\left\{K\left[\epsilon_n^{-1}\left(H_{i1}(s)-s\right),{\epsilon_n}^{-1}(H_{i2}(t)-t)\right]\left[C\left[H_{i1}(s),H_{i2}(t)\right]-C(s,t)\right]\right\}^2\\
=&\frac{1}{n\epsilon_n^4}\mathrm{E}\left(\sum_{k_1=0}^{a_n(s)}\sum_{k_2=0}^{b_n(t)}K\left(\frac{F_{i1}(k_1)-s}{\epsilon_n},\frac{F_{i2}(k_2)-t}{\epsilon_n}\right)\left[C\left[F_{i1}(k_1),F_{i2}(k_2)\right]-C(s,t)\right]\right)^2+o\left(\frac{1}{n}\right)\\
=&\frac{1}{n\epsilon_n^2}\sum_{k_1=0}^{a_n(s)}\sum_{k_2=0}^{b_n(t)}\int K(u,v)^2\left[C(s+u\epsilon_n,t+v\epsilon_n)-C(s,t)\right]^2 f_{F_1(k_1),F_2(k_2)}(s+u\epsilon_n,t+v\epsilon_n)\mathrm{d}u\mathrm{d}v\\
&+o\left(\frac{1}{n}\right),
\end{aligned}
$$

where the residuals are obtained by similar techniques used to compute $\mathrm{E}\left[\hat{m}_1(s,t)\right]$. By Taylor expansions, we have $\left[C(s+u\epsilon_n,t+v\epsilon_n)-C(s,t)\right]^2 f_{F_1(k_1),F_2(k_2)}(s+u\epsilon_n,t+v\epsilon_n)=O(\epsilon_n^2)$. Thus, the first term of (A.15) is of order $O(1/n)$.

Then, we check the second term of (A.15) . From Equation (A.14),

$$
\frac{1}{n\epsilon_n^4}\left\{\mathrm{E}\left(K\left[\epsilon_n^{-1}\left(H_{i1}(s)-s\right),{\epsilon_n}^{-1}(H_{i2}(t)-t)\right]\left[C\left[H_{i1}(s),H_{i2}(t)\right]-C(s,t)\right]\right)\right\}^2=O\left(\frac{1}{n}\right).
\tag{A.16}
$$

Therefore,

$$
\mathrm{Var}\left(\sqrt{n\epsilon_n^2}\hat{m}_1(s,t)\right)=O(\epsilon_n^2)\to 0.
\tag{A.17}
$$

Equations (A.14) and (A.17) entail that

$$\sqrt{n\epsilon_n^2}\left(\hat{m}_1(s,t)-\kappa_2\zeta_n(s,t)f_{H(s,t),n}(s,t)\epsilon_n^2\right)\to_p 0.$$

Recall $\zeta(s,t)$ defined in (10), it can be easily seen that $\zeta_n(s,t)\to\zeta(s,t)$. As $a_n(s)$ and $b_n(t)$ go to infinity, we have

$$\sqrt{n\epsilon_n^2}\left(\hat{m}_1(s,t)-\kappa_2\zeta(s,t)f_{H(s,t)}(s,t)\epsilon_n^2\right)\to_p 0.$$

Together with (A.1), it follows

$$\sqrt{n\epsilon_n^2}\left(\frac{\hat{m}_1(s,t)}{\hat{f}_{H(s,t)}(s,t)}-\kappa_2\frac{f_{H(s,t)}(s,t)}{\hat{f}_{H(s,t)}(s,t)}\zeta(s,t)\epsilon_n^2\right)\to_p 0.$$

Recall $n\epsilon_n^6=O(1)$, hence we also have

$$\sqrt{n\epsilon_n^2}\left(\kappa_2\frac{f_{H(s,t)}(s,t)}{\hat{f}_{H(s,t)}(s,t)}\zeta(s,t)\epsilon_n^2-\kappa_2\zeta(s,t)\epsilon_n^2\right)\to_p 0. \tag{A.18}$$

Therefore,

$$\sqrt{n\epsilon_n^2}\left(\frac{\hat{m}_1(s,t)}{\hat{f}_{H(s,t)}(s,t)}-\kappa_2\zeta(s,t)\epsilon_n^2\right)\to_p 0. \tag{A.19}$$

Summing up (A.13) and (A.19) finishes the proof.

□

Hence the AMSE of $\hat{C}(s,t)$ is $(C(s,t)\left[1-C(s,t)\right]R_2(K))/(n\epsilon_n^2 f_{H(s,t)}(s,t))+\kappa_2^2\zeta(s,t)^2\epsilon_n^4$.

The above theorems guarantee the identifiability of underlying copula. With Assumptions 2.2, 2.1, 2.3, and 2.4, if there exists another copula $\tilde{C}$ compatible with data, the pointwise difference between C and $\tilde{C}$ at $(s,t) \in V$ is

$$\left|C(s,t) - \tilde{C}(s,t)\right|^2 \leq \mathrm{E}\left|\hat{C}(s,t) - C(s,t)\right|^2 + \mathrm{E}\left|\hat{C}(s,t) - \tilde{C}(s,t)\right|^2 \to 0.$$

Since $\tilde{C}(s,t) - C(s,t)$ does not change with n, it has to be that $\tilde{C}(s,t) - C(s,t) = 0$ for any $(s,t) \in V$. That is, the copula is identifiable at V.

A.3 Proof of Theorem 2.3

We now demonstrate the asymptotic properties of the copula estimator defined in (6) when the marginal parameters are unknown and the estimates are plugged in. We first analyze the numerator and denominator of the estimator separately in Lemmas A.2 and A.4. Finally, the distribution of the copula estimator follows as stated in Theorem 2.3.

For the following results Lemmas A.2, A.3, and A.4, we prove them with uniform kernel. The proof for other compacted supported kernels is a trivial extension from the uniform kernel with Lipschitz conditions of the kernel function. Let $A_n(s,t)$ denote the neighborhood of (s,t) with radius ϵ_n.

Assumption 2.6 indicates that for any $\xi > 0$ there exists $\gamma_\xi > 0$ such that for n big enough,

$$P(\hat{\beta} \notin B(\beta, n^{-1/2}\gamma_\xi)) < \xi$$

where $B(\beta, d)$ is a neighborhood of β with radius d.

Lemma A.2 (Consistency of Denominator)**.** *Under Assumptions 2.1, 2.2, 2.5, and 2.6,*

$$\frac{1}{n\epsilon_n^2} \sum_{i=1}^{n} \left\{ 1\left(H_i(s,t;\hat{\beta}) \in A_n(s,t)\right) - 1\left[H_i(s,t) \in A_n(s,t)\right] \right\} \to_p 0. \tag{A.20}$$

Therefore,

$$\frac{1}{n\epsilon_n^2} \sum_{i=1}^{n} K\left[{\epsilon_n}^{-1}(H_{i1}(s;\hat{\beta}) - s), {\epsilon_n}^{-1}(H_{i2}(s;\hat{\beta}) - t)\right] \to_p f_{H(s,t)}(s,t). \tag{A.21}$$

Proof. Different aspects of the method of proof can be found in Sukhatme (1958), Randles (1984), and Frees (1995b). Denote

$$l_{ni}(s,t;\theta) = 1\left(H_i(s,t;\theta) \in A_n(s,t)\right)$$

and

$$S_n(\theta;s,t) = \frac{1}{n\epsilon_n^2}\sum_{i=1}^{n}\left[l_{ni}(s,t;\beta) - l_{ni}(s,t;\theta)\right].$$

For arbitrary $\epsilon > 0$, for any $\xi > 0$, let γ_ξ be the constant in Assumption 2.6. To show (A.20), we calculate the probability

$$\begin{aligned}
P\left(\left|S_n(\hat{\beta};s,t)\right| > \epsilon\right) &= P\left(\left|S_n(\hat{\beta};s,t)\right| > \epsilon, \hat{\beta} \in B(\beta, n^{-1/2}\gamma_\xi)\right) \\
&\quad + P\left(\left|S_n(\hat{\beta};s,t)\right| > \epsilon, \hat{\beta} \notin B(\beta, n^{-1/2}\gamma_\xi)\right) \\
&\le P\left(\sup_{\theta \in B(\beta, n^{-1/2}\gamma_\xi)} |S_n(\theta;s,t)| > \epsilon\right) + P\left(\hat{\beta} \notin B(\beta, n^{-1/2}\gamma_\xi)\right) \\
&:= M_1 + M_2.
\end{aligned}$$

By Assumption 2.6,

$$M_2 < \xi/2.$$

We now check M_1. Since $|l_{ni}(s,t;\beta) - l_{ni}(s,t;\theta)| \le 1$, using a similar method as in Frees (1995a), we can see there are two cases $|l_{ni}(s,t;\beta) - l_{ni}(s,t;\theta)|$ can be 1, i.e.,

$$|l_{ni}(s,t;\beta) - l_{ni}(s,t;\theta)| \le \begin{cases} 1(|H_i(s,t;\theta) - (s,t)| > \epsilon_n, |H_i(s,t) - (s,t)| \le \epsilon_n) := J_1, \\ 1(|H_i(s,t;\theta) - (s,t)| \le \epsilon_n, |H_i(s,t) - (s,t)| > \epsilon_n) := J_2. \end{cases}$$

For the first case J_1,

$$J_1 \le 1(\epsilon_n < |H_i(s,t;\theta) - (s,t)|).$$

Further, subtracting $|H_i(s,t)-(s,t)|$ from both sides yields

$$J_1 \leq 1\left(\epsilon_n - |H_i(s,t)-(s,t)| < |H_i(s,t;\theta)-(s,t)| - |H_i(s,t)-(s,t)|\right).$$

Similarly, for the second case J_2

$$J_2 \leq 1\left(\epsilon_n \geq |H_i(s,t;\theta)-(s,t)|\right).$$

So we have

$$J_2 \leq 1\left(|H_i(s,t)-(s,t)| - \epsilon_n \leq |H_i(s,t)-(s,t)| - |H_i(s,t;\theta)-(s,t)|\right).$$

Since if $J_1 = 1$, $|H_i(s,t)-(s,t)| \leq \epsilon_n$, and when $J_2 = 1$, $|H_i(s,t)-(s,t)| > \epsilon_n$, we can summarize these two cases obtaining

$$\begin{aligned} |l_{ni}(s,t;\beta) - l_{ni}(s,t;\theta)| &\leq 1\left(||H_i(s,t)-(s,t)| - \epsilon_n| \leq ||H_i(s,t)-(s,t)| - |H_i(s,t;\theta)-(s,t)||\right) \\ &\leq 1\left(||H_i(s,t)-(s,t)| - \epsilon_n| \leq \alpha_2 |\beta-\theta|\right), \end{aligned} \tag{A.22}$$

where the second inequality is due to Assumption 2.5.

Now we take supremum with respect to θ over $B(\beta, n^{-1/2}\gamma_\xi)$, defining

$$\eta_n(s,t;X_i) = \sup_{\theta \in B(\beta, n^{-1/2}\gamma_\xi)} |l_{ni}(s,t;\theta) - l_{ni}(s,t;\beta)|.$$

From (A.22), there exists a constant α_2 such that

$$\eta_n(s,t;X_i) \leq 1\left(||H_i(s,t)-(s,t)| - \epsilon_n| \leq \alpha_2 n^{-1/2}\gamma_\xi\right).$$

By Assumption 2.2, the density of $H_i(s,t)$ is bounded. Thus $|H_i(s,t)-(s,t)| - \epsilon_n$ which is the linear transformation result from $H_i(s,t)$ also has a bounded density. For all n, i, and (s,t), there exists a constant α_3 such that

$$\mathrm{E}\eta_n(s,t;X_i) \leq \alpha_3 \gamma_\xi^2 n^{-1}. \tag{A.23}$$

Note that $\sup_{\theta \in B(\beta, n^{-1/2}\gamma_\xi)} |S_n(\theta;s,t)| \leq 1/(n\epsilon_n^2) \sum_{i=1}^n \eta_n(s,t;X_i)$. Therefore,

$$\begin{aligned}
M_1 &\leq P\left(\frac{1}{n\epsilon_n^2}\sum_{i=1}^n \eta_n(s,t;X_i) > \epsilon\right) \\
&\leq P\left(\left(\frac{1}{n\epsilon_n^2}\sum_{i=1}^n [\eta_n(s,t;X_i) - \mathrm{E}\eta_n(s,t;X_i)] + \frac{1}{n{\epsilon_n}^2}\sum_{i=1}^n \mathrm{E}\eta_n(s,t;X_i)\right) > \epsilon\right)
\end{aligned}$$

From (A.23),

$$\frac{1}{n\epsilon_n^2}\sum_{i=1}^n \mathrm{E}\eta_n(s,t;X_i) \leq \frac{\alpha_3\gamma_\xi^2}{n\epsilon_n^2} \to 0.$$

Hence, when n gets large,

$$\frac{1}{n\epsilon_n^2}\sum_{i=1}^n \mathrm{E}\eta_n(s,t;X_i) < \epsilon/2.$$

By applying the Chebyshev's inequality,

$$\begin{aligned}
M_1 &\leq P\left(\frac{1}{n\epsilon_n^2}\sum_{i=1}^n [\eta_n(s,t;X_i) - \mathrm{E}\eta_n(s,t;X_i)] > \epsilon/2\right) \\
&\leq \frac{1}{(\epsilon/2)^2}\mathrm{Var}\left(\frac{1}{n\epsilon_n^2}\sum_{i=1}^n [\eta_n(s,t;X_i) - \mathrm{E}\eta_n(s,t;X_i)]\right).
\end{aligned}$$

Note that $\eta_n(s,t;X_i)^2 = \eta_n(s,t;X_i)$. By applying (A.23) we have

$$M_1 \leq \frac{1}{(\epsilon/2)^2}\frac{1}{(n\epsilon_n^2)^2}\sum_{i=1}^n \left[\mathrm{E}\eta_n(s,t;X_i)^2 - [\mathrm{E}\eta_n(s,t;X_i)]^2\right] \leq \frac{1}{(\epsilon/2)^2(n\epsilon_n^2)^2}\alpha_3\gamma_\xi^2.$$

Therefore, when n is large enough $M_1 < \xi/2$. Now (A.20) follows from the fact that for arbitrary ϵ and $\xi > 0$,

$$P\left(\left|S_n(\hat{\beta};s,t)\right| > \epsilon\right) < \xi.$$

Finally, note that (A.21) follows from (A.1) and (A.20), and the proof is finished. □

Define

$$h_{ni}(s,t;\theta) = 1\left(H_i(s,t;\theta) \in A_n(s,t)\right) Y_i(\theta),$$

where $Y_i(\theta)$ is defined in (5).

Lemma A.3. *Under Assumption 2.5, there exists a constant α_4 such that for all n and i,*

$$\mathrm{E} \sup_{\theta \in B(\beta,d)} |h_{ni}(s,t;\theta) - h_{ni}(s,t;\beta)| \leq \alpha_4 d^2.$$

Proof. We first note that

$$|h_{ni}(s,t;\theta) - h_{ni}(s,t;\beta)| \leq |l_{ni}(s,t;\beta) - l_{ni}(s,t;\theta)| + |Y_i(\theta) - Y_i(\beta)|.$$

From (5),

$$Y_i(\theta) = 1\left(Y_{i1} \leq F_1^{(-1)}\left(H_1(s;\theta_1);\theta_1\right), Y_{i2} \leq F_2^{(-1)}\left(H_2(t;\theta_2);\theta_2\right)\right).$$

As in Figure 1, when θ approaches β with distance d small enough, there exists integer k such that

$$M_1^k \leq X_1'\beta_1 < M_1^{k+1}$$
$$M_1^k \leq X_1'\theta_1 < M_1^k,$$

with probability 1. Hence, almost surely,

$$F_1^{(-1)}\left(H_1(s;\beta_1);\beta_1\right) = F_1^{(-1)}\left(H_1(s;\theta_1);\theta_1\right) = k.$$

By a similar argument, with small d, $F_2^{(-1)}\left(H_2(t;\beta_2);\beta_2\right) = F_2^{(-1)}\left(H_2(t;\theta_2);\theta_2\right)$. Hence, $Y_i(\theta) = Y_i(\beta)$ almost surely when d is small enough. Thus $|h_{ni}(s,t;\theta) - h_{ni}(s,t;\beta)| \leq$

$|l_{ni}(s,t;\beta) - l_{ni}(s,t;\theta)|$. Recall (A.23) in the proof of Lemma A.2, one obtains when d is small enough, there exists constant α_4 such that for all n, i and (s,t),

$$\mathrm{E} \sup_{\theta \in B(\beta,d)} |h_{ni}(s,t;\theta) - h_{ni}(s,t;\beta)| \leq \alpha_4 d^2,$$

as required. □

Lemma A.4. *With Assumptions 2.5 and 2.6,*

$$\frac{1}{\sqrt{n\epsilon_n^2}} \sum_{i=1}^{n} \left[h_{ni}(s,t;\hat{\beta}) - h_{ni}(s,t;\beta) \right] \to_p 0 \tag{A.24}$$

Proof. Denote

$$\begin{aligned} R_{ni}(\theta;s,t) &= h_{ni}(s,t;\theta) - h_{ni}(s,t;\beta) \\ &= 1\left(H_i(s,t;\theta) \in A_n(s,t)\right) Y_i(\theta) - 1\left(H_i(s,t) \in A_n(s,t)\right) Y_i(\beta), \end{aligned}$$

and

$$Q_n(\theta;s,t) = \frac{1}{\sqrt{n\epsilon_n^2}} \sum_{i=1}^{n} R_{ni}(\theta;s,t).$$

For any $\epsilon > 0$, for any $\xi > 0$, let γ_ξ be the constant in Assumption 2.6. To show (A.24), we check the probability

$$\begin{aligned} P\left(\left| Q_n(\hat{\beta};s,t) \right| > \epsilon \right) &= P\left(\left| Q_n(\hat{\beta};s,t) \right| > \epsilon, \hat{\beta} \in B(\beta, n^{-1/2}\gamma_\xi) \right) \\ &\quad + P\left(\left| Q_n(\hat{\beta};s,t) \right| > \epsilon, \hat{\beta} \notin B(\beta, n^{-1/2}\gamma_\xi) \right) \\ &\leq P\left(\sup_{\theta \in B(\beta, n^{-1/2}\gamma_\xi)} |Q_n(\theta;s,t)| > \epsilon \right) + P\left(\hat{\beta} \notin B(\beta, n^{-1/2}\gamma_\xi) \right) \\ &:= I_1 + I_2 \end{aligned}$$

By Assumption 2.6, $I_2 < \xi/2$.

Now define $L_{ni}(s,t) = \sup_{\theta \in B(\beta, n^{-1/2}\gamma_\xi)} |R_{ni}(\theta; s,t)|$. By Lemma A.3,

$$\begin{aligned} \mathrm{E}L_{ni}(s,t) &\le \mathrm{E}\left(\sup_{\theta \in B(\beta, n^{-1/2}\gamma_\xi)} |h_{ni}(s,t;\theta) - h_{ni}(s,t;\theta)| \right) \\ &\le \alpha_4 (\gamma_\xi/\sqrt{n})^2. \end{aligned} \tag{A.25}$$

At the same time,

$$\sup_{\theta \in B(\beta, n^{-1/2}\gamma_\xi)} |Q_n(\theta; s,t)| \le \frac{1}{\sqrt{n\epsilon_n^2}} \sum_{i=1}^n L_{ni}(s,t),$$

hence we have

$$\begin{aligned} & P\left(\sup_{\theta \in B(\beta, n^{-1/2}\gamma_\xi)} |Q_n(\theta; s,t)| > \epsilon \right) \\ \le & P\left(\frac{1}{\sqrt{n\epsilon_n^2}} \sum_{i=1}^n L_{ni}(s,t) > \epsilon \right) \\ \le & P\left(\frac{1}{\sqrt{n\epsilon_n^2}} \sum_{i=1}^n [L_{ni}(s,t) - \mathrm{E}L_{ni}(s,t)] + \frac{1}{\sqrt{n\epsilon_n^2}} \sum_{i=1}^n \mathrm{E}L_{ni}(s,t) > \epsilon \right). \end{aligned}$$

It follows (A.25) that

$$\frac{1}{\sqrt{n\epsilon_n^2}} \sum_{i=1}^n \mathrm{E}L_{ni}(s,t) \le \alpha_4 \gamma_\xi^2 \frac{1}{\sqrt{n\epsilon_n^2}}.$$

Therefore, when n is large enough,

$$\frac{1}{\sqrt{n\epsilon_n^2}} \sum_{i=1}^n \mathrm{E}L_{ni}(s,t) < \epsilon/2.$$

Noting that

$$\mathrm{Var}\left(\frac{1}{\sqrt{n\epsilon_n^2}} \sum_{i=1}^n [L_{ni}(s,t) - \mathrm{E}L_{ni}(s,t)] \right) = \frac{1}{n\epsilon_n^2} \sum_{i=1}^n \left(\mathrm{E}L_{ni}(s,t)^2 - [\mathrm{E}L_{ni}(s,t)]^2 \right).$$

Since $|R_{ni}(\theta; s, t)| \leq 1$, one obtains $L_{ni}(s,t)^2 \leq L_{ni}(s,t)$. From (A.25),

$$\mathrm{Var}\left(\frac{1}{\sqrt{n\epsilon_n^2}}\sum_{i=1}^{n}\left[L_{ni}(s,t) - \mathrm{E}L_{ni}(s,t)\right]\right) \leq \frac{1}{n\epsilon_n^2}\alpha_4\gamma_\xi^2.$$

By Chebyshev's inequality

$$\begin{aligned} & P\left(\frac{1}{\sqrt{n\epsilon_n^2}}\sum_{i=1}^{n}\left[L_{ni}(s,t) - \mathrm{E}L_{ni}(s,t)\right] > \epsilon/2\right) \\ \leq & \frac{1}{(\epsilon/2)^2}\mathrm{Var}\left(\frac{1}{\sqrt{n\epsilon_n^2}}\sum_{i=1}^{n}\left[L_{ni}(s,t) - \mathrm{E}L_{ni}(s,t)\right]\right) \to 0. \end{aligned}$$

The theorem now follows $Q_n(\hat{\beta}; s, t) \to_p 0$. □

Proof of Theorem 2.3

Proof. Recall from Lemma A.2, we have $(n\pi\epsilon_n^2)^{-1}\sum_{i=1}^n l_{ni}(s,t;\hat{\beta})$ is also a consistent estimator of $f_{H(s,t)}(s,t)$. Following the proof of Theorem 2.2, replacing $\hat{f}_{H(s,t)}$ by its approximation $(n\pi\epsilon_n^2)^{-1}\sum_{i=1}^n l_{ni}(s,t;\hat{\beta})$ in (A.13) and (A.19), from Slutsky's theorem, we have

$$\sqrt{n\epsilon_n^2}\frac{\hat{m}_2(s,t)}{(n\pi\epsilon_n^2)^{-1}\sum_{i=1}^n l_{ni}(s,t;\hat{\beta})} \to_d N\left(0, \frac{R_2(K)}{f_{H(s,t)}(s,t)}C(s,t)\left[1 - C(s,t)\right]\right), \tag{A.26}$$

and

$$\sqrt{n\epsilon_n^2}\left(\frac{\hat{m}_1(s,t)}{(n\pi\epsilon_n^2)^{-1}\sum_{i=1}^n l_{ni}(s,t;\hat{\beta})} - \kappa_2\zeta(s,t)\epsilon_n^2\frac{f_{H(s,t)}(s,t)}{(n\pi\epsilon_n^2)^{-1}\sum_{i=1}^n l_{ni}(s,t;\hat{\beta})}\right) \to_p 0.$$

Using Lemma A.2 together with the fact that $n\epsilon_n^6 = O_p(1)$, we have

$$\sqrt{n\epsilon_n^2}\left(\kappa_2\zeta(s,t)\epsilon_n^2\frac{f_{H(s,t)}(s,t)}{(n\pi\epsilon_n^2)^{-1}\sum_{i=1}^n l_{ni}(s,t;\hat{\beta})} - \kappa_2\zeta(s,t)\epsilon_n^2\right) \to_p 0.$$

This leads to

$$\sqrt{n\epsilon_n^2}\left(\frac{\hat{m}_1(s,t)}{(n\pi\epsilon_n^2)^{-1}\sum_{i=1}^n l_{ni}(s,t;\hat{\beta})} - \kappa_2\zeta(s,t)\epsilon_n^2\right) \to_p 0. \tag{A.27}$$

Summing up (A.26) and (A.27), we have

$$\sqrt{n\epsilon_n^2}\left(\frac{\sum_{i=1}^n h_{ni}(s,t;\beta)}{\sum_{i=1}^n l_{ni}(s,t;\hat{\beta})} - C(s,t) - \kappa_2\zeta(s,t)\epsilon_n^2\right) \to_d N\left(0, \frac{C(s,t)\left[1-C(s,t)\right]R_2(K)}{f_{H(s,t)}(s,t)}\right). \tag{A.28}$$

It follows Lemmas A.2 and A.4 that

$$\sqrt{n\epsilon_n^2}\left(\frac{\sum_{i=1}^n h_{ni}(s,t;\hat{\beta})}{\sum_{i=1}^n l_{ni}(s,t;\hat{\beta})} - \frac{\sum_{i=1}^n h_{ni}(s,t;\beta)}{\sum_{i=1}^n l_{ni}(s,t;\hat{\beta})}\right) \to_p 0. \tag{A.29}$$

Summing up (A.28) and (A.29) yields

$$\sqrt{n\epsilon_n^2}\left(\hat{C}(s,t;\hat{\beta}) - C(s,t) - \kappa_2\zeta(s,t)\epsilon_n^2\right) \to_d N\left(0, \frac{C(s,t)\left[1-C(s,t)\right]R_2(K)}{f_{H(s,t)}(s,t)}\right),$$

as required. □

B. Additional Simulations

B.1 Finite Sample Performance under Other Marginal Settings

Here we include additional simulations to evaluate the finite sample performance of the proposed estimator under other marginal settings. As mentioned in Section 3.2, the estimator performs comparably across different levels of dependence. Hence here we only employ the Gaussian copula with high dependence as the underlying model.

Negative binomial. We use the same mean structure as the Poisson outcomes in Section 3.1, and dispersion parameters as 0.7 and 1.3 respectively for $j = 1, 2$. Figure B.1 includes the contour plots of the estimator, and it follows the same pattern that the estimator is more accurate when the data are less discrete. Though by comparing each plot in Figure B.1 with the corresponding plot of Poisson variables with same marginal mean level in Figure 5, we see that it is more difficult to identify copulas for negative binomial variables when overdispersion is pronounced.

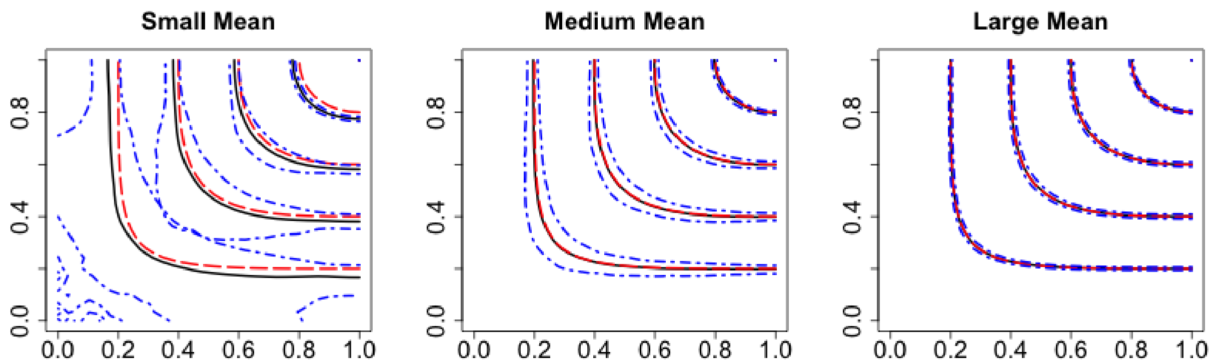

Figure B.1: Contour plots of the nonparametric estimator for negative binomial outcomes with sample size 5000. The average of the estimator over 500 replications is given by the solid lines, while the dash-dot symbols give the corresponding 95% confidence interval for every other copula value, and the dashed lines give the underlying copulas.

Mixed outcomes. So far, the simulations we have done are for the cases in which both margins have same mean level. It is of interest to see the combination of variables with different levels of discreteness. Figure B.2 includes two examples. The left panel displays the case when Y_1 is a binary and Y_2 is a Poisson variable with medium marginal mean. Compared with Figures 5 and 6, we can see the variance and bias are smaller than these of two binary variables but bigger than two Poisson variables with medium means. This phenomenon is also reflected in the ISE values summarized in Table B.1 that the ISE value of the mixed case 0.481×10^{-3} is smaller than the ISE of two binary outcomes 2.098×10^{-3} and bigger than the value of two Poisson variables at medium mean level 0.103×10^{-3} in Table 1. The difference between the two margins becomes significant at boundaries. The right panel includes the results of two Poisson variables at different marginal mean levels, and we can draw the same conclusion.

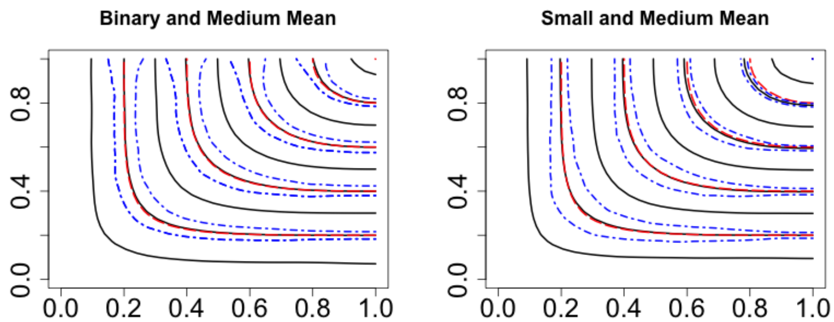

Figure B.2: Contour plots of the nonparametric estimator for mixed outcomes with sample size 5000.

Table B.1: ISE of additional simulations (multiplied by 1000).

Margins	Binary		Binary and Medium Poisson		Small and Medium Poisson	
	Average	sd	Average	sd	Average	sd
ISE	2.098	0.524	0.481	0.117	0.271	0.098

B.2 Copula Identification Using Probability of Zeros

In principle, the copula can be identified using only the probability of $(0,0)$ due to the fact $F_{\mathbf{Y}|\mathbf{X}}(0,0|X_1,X_2) = C(F_1(0|X_1), F_2(0|X_2))$. Assuming that the explanatory variables have sufficient range so that the probabilities of zeros span the interval $(0,1)$, the corresponding estimator is of the form

$$\hat{C}_0(s,t;\beta) = \frac{\sum_{i=1}^{n} K\left[\left(F_1(0|X_{i1}) - s\right)/\epsilon_n, \left(F_2(0|X_{i2}) - t\right)/\epsilon_n\right] 1(Y_{i1}=0, Y_{i2}=0)}{\sum_{i=1}^{n} K\left[\left(F_1(0|X_{i1}) - s\right)/\epsilon_n, \left(F_2(0|X_{i2}) - t\right)/\epsilon_n\right]}, \qquad \text{(B.1)}$$

which is an application of the Nadaraya-Watson estimator. From its established asymptotic results (Hansen 2009), the variance of $\hat{C}_0(s,t;\beta)$ is of the form

$$\frac{C(s,t)\left[1 - C(s,t)\right] R_2(K)}{n\epsilon_n^2 f_{F_1(0|X_1),F_2(0|X_2)}(s,t)}.$$

Theorem 2.2 underscores the benefit of employing our nonstandard estimator. From the form of $f_{H(s,t;\mathbf{X})}(s,t)$ in (9), we can see $f_{H(s,t;\mathbf{X})}(s,t) > f_{F_1(0|X_1),F_2(0|X_2)}(s,t)$. Thus, we have smaller variance by applying the proposed estimator. Intuitively, instead of applying a variable $(F_1(0|X_1), F_2(0|X_2))$, we use many variables $(F_1(k_1|X_1), F_2(k_2|X_2))$, $k_1 = 0, \ldots, k_2 = 0, \ldots$ for copula estimation, which increases the efficiency.

Now we illustrate this point through a simulation study. Figure B.3 displays the results for copula identification using only zeros under the medium marginal mean level with high dependence. Compared with the middle columns in Figures 3 and 5, it is clear that the proposed nonstandard nonparametric estimator has smaller bias and variance.

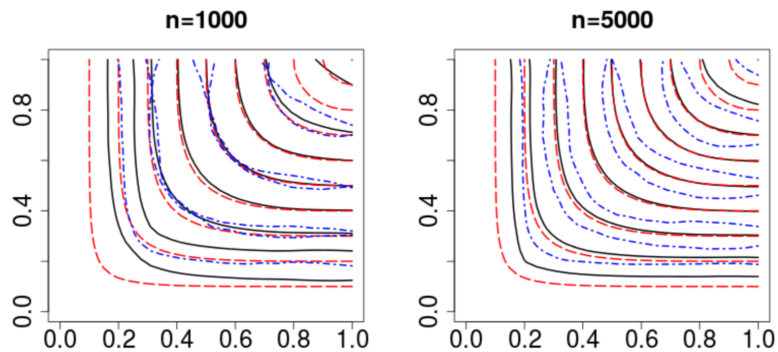

Figure B.3: Contour plots of the estimator (B.1) using the probability of $(0, 0)$. The average of the estimator over 500 replications is given by the solid lines, while the dash-dot symbols give the corresponding 95% confidence interval, and the dashed lines give the underlying copulas.

B.3 Copula Specification and Diagnosis

We now explore the usage of the nonparametric estimator as a diagnostic tool under different scenarios. For each of the simulations, given the generated data, we first fit the marginal models. Then, we plug the marginal estimates in (6) to obtain our nonparametric estimator. Meanwhile, different parametric copulas are fit through MLE. Finally, we compare the parametric copulas with our nonparametric estimator. We generate the data using Gaussian (no tail dependencies), Clayton (lower tail dependence), and Joe (upper tail dependence) copulas to explore the impact of tail dependence, and here we only present results for Poisson variables.

Case 1. Gaussian copulas. We first analyze the generated data from Gaussian copulas, the most commonly used copulas without tail dependence. We include a representative graphical summary of the results under medium means in Figures B.4 and B.5. Due to space limitations, the results of other scenarios are summarized numerically in Table B.2. Under low dependence, the dashed lines (corresponding to the fitted parametric copulas) across plots in Figure B.4 are hardly distinguishable due to the fact that they are all similar to the independence copula. As a result, their distances with the nonparametric estimator are comparable. Therefore, the choice of parametric copulas is not essential when the dependence is very weak.

In contrast, under high dependence as in Figure B.5, we can exclude the Gumbel and Joe copulas due to the large discrepancy with the nonparametric estimator in the center of the graphs, and the Clayton copula is wide apart towards right upper corner. Recall that Gaussian and Frank copulas do not have tail dependence. Gumbel and Joe copulas have upper tail dependence, while a Clayton copula has lower tail dependence. Hence, when the dependence is strong, we can rule out copulas with wrong types of tail dependencies, and the graphical comparison with our nonparametric estimator suggests improvement. Due to the similarity in the Gaussian and Frank copulas, the choice between these two copulas is difficult and probably not that important. It is also noticeable that among copulas with upper tail dependence, the Joe copula has more significant distance than the Gumbel copula with the nonparametric estimator, which can be explained by the stronger tail dependence of the Joe copula.

Figure B.6 displays the graphical results under small marginal means and high dependence. Due to large bias and variance in the nonparametric estimator, as demonstrated in Section 3.2, all the copulas are inside the confidence intervals. Hence, the wrong models cannot be rejected statistically, and it is hard to make conclusions about copula specification.

Table B.2 summarizes the results numerically. As an example, when the sample size is

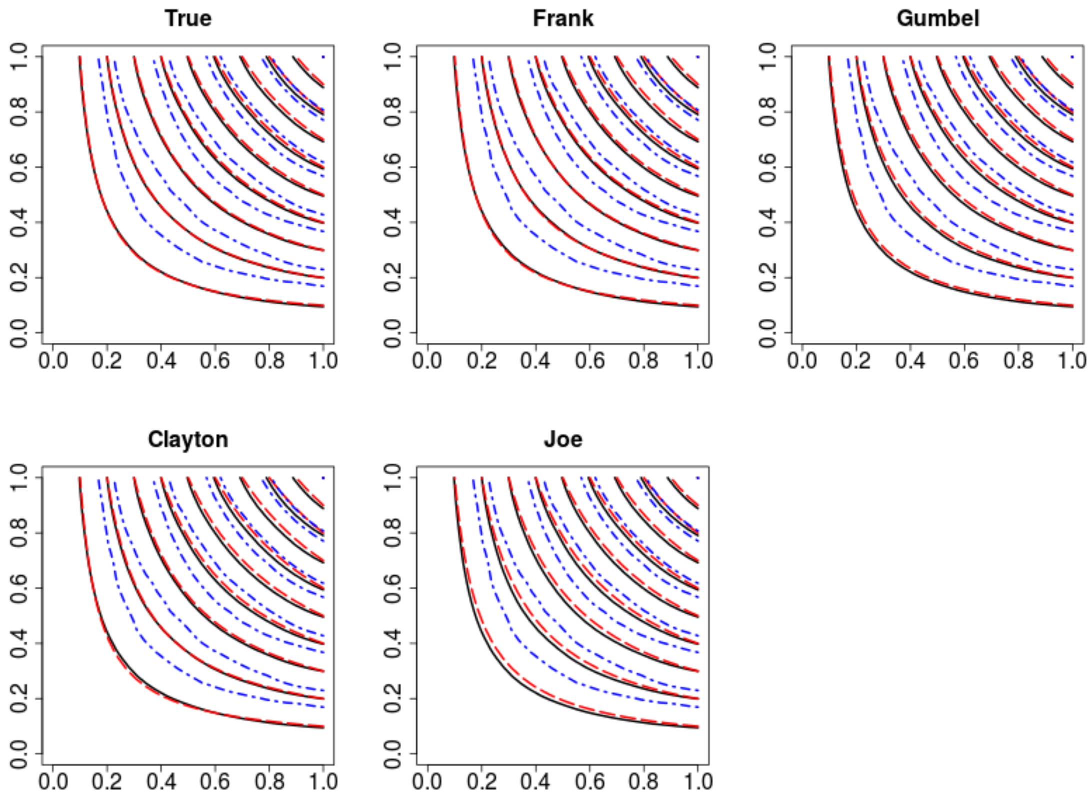

Figure B.4: Contour plots of the nonparametric estimator compared with several parametric copulas under medium means and low dependence. The estimator is given by the solid lines, and the dash-dot symbols give the corresponding confidence intervals. The fitted parametric copulas are given by a dashed line. These plots are based on a sample size of 1000.

$n = 1000$, under small marginal means and low dependence, the average distance of the fitted Gaussian copula with our nonparametric estimator is 2.865×10^{-3} with standard deviation 2.081×10^{-3} over the 500 replications, while the fitted Frank copula has an average distance 2.834×10^{-3} with standard deviance 2.095×10^{-3}. Consistent with Figure B.6, with high level of discreteness, the distances between different parametric copulas with the nonparametric estimator are high and comparable (first three rows of each block). Thus, it is difficult to pick up a copula. When the marginal means are at medium and large levels, the strength of dependence plays an important role in the model specification. Under low dependence, we are unable to distinguish most of the copulas in terms of the distance, except the Joe copula shows worse fitting at large mean level. With stronger dependence, we can rule out

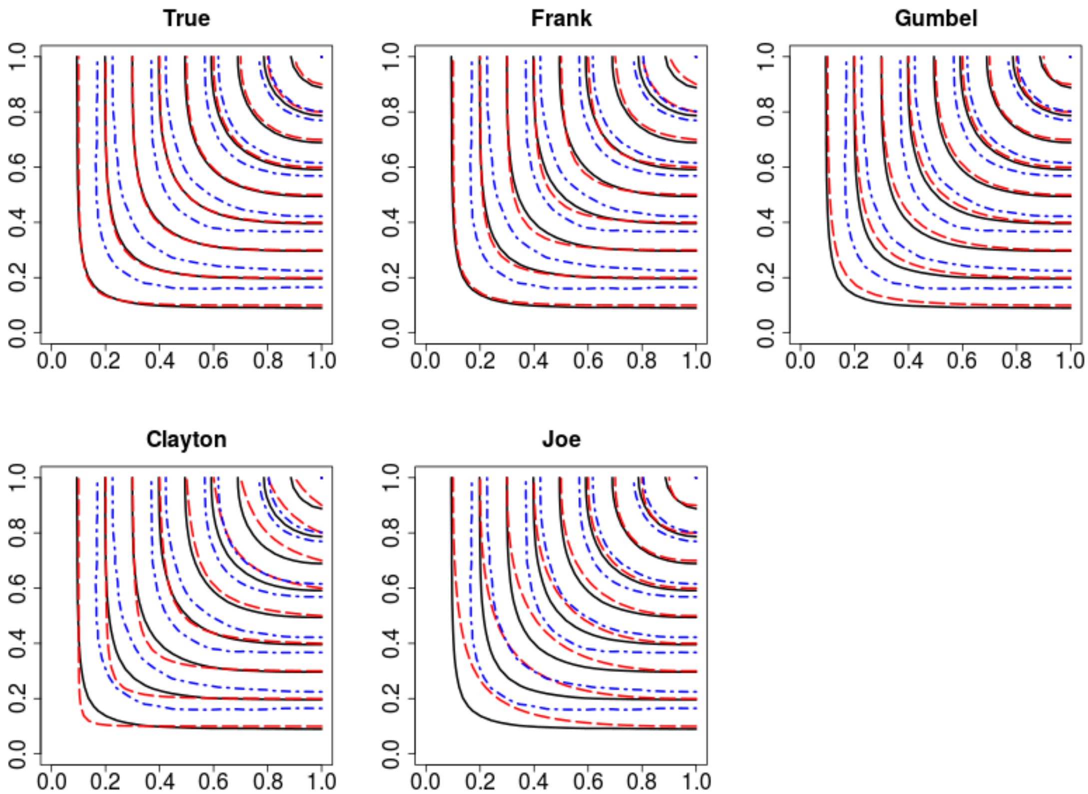

Figure B.5: Contour plots of the nonparametric estimator compared with several parametric copulas under medium means and high dependence. These plots are based on a sample size of 1000.

the Gumbel, Joe, and Clayton copulas, especially under high dependence, where the true model outperforms alternative models clearly. Again, the Gaussian and Frank copulas are generally indistinguishable, except that the difference is more noticeable with large marginal means and high dependence.

Case 2. Clayton copulas. To further explore the impact of tail dependence, we next consider copulas with lower tail dependence. Table B.3 portrays the results of the Clayton copulas. As we concluded the choice of copulas is not essential under low dependence or small marginal means, we omit the corresponding results here. In all the scenarios, the true model has smallest distance with the nonparametric estimator. Meanwhile, we can rule out the Gumbel and Joe copulas easily as they have opposite tail dependence structures of the

Table B.2: Distances of different parametric copulas with the nonparametric estimator when Gaussian copulas are the underlying (multiplied by 1000).

			True		Frank		Gumbel		Clayton		Joe	
n	Marginal Mean	Dependence	Average	sd	Average	sd	Average	sd	Average	sd	Average	sd
1000	Small	Low	2.865	2.081	2.834	2.095	3.026	2.266	2.809	1.928	3.092	2.309
		Moderate	3.061	2.233	2.985	2.151	3.523	2.635	3.045	1.853	3.855	2.853
		High	3.547	2.652	3.495	2.519	4.049	3.205	3.633	2.288	4.994	3.893
	Medium	Low	0.331	0.118	0.323	0.102	0.356	0.139	0.351	0.114	0.394	0.169
		Moderate	0.330	0.125	0.330	0.110	0.409	0.187	0.433	0.146	0.587	0.282
		High	0.352	0.150	0.392	0.158	0.432	0.221	0.552	0.190	0.846	0.408
	Large	Low	0.088	0.040	0.081	0.033	0.090	0.039	0.096	0.049	0.115	0.055
		Moderate	0.088	0.040	0.087	0.035	0.112	0.051	0.174	0.085	0.236	0.098
		High	0.091	0.047	0.114	0.052	0.128	0.064	0.425	0.144	0.459	0.149
5000	Small	Low	0.857	0.311	0.843	0.291	0.953	0.391	0.863	0.261	1.000	0.423
		Moderate	0.878	0.334	0.849	0.285	1.171	0.537	0.986	0.281	1.425	0.661
		High	0.974	0.429	1.014	0.402	1.286	0.682	1.204	0.445	2.039	1.054
	Medium	Low	0.107	0.030	0.104	0.025	0.126	0.046	0.124	0.031	0.157	0.064
		Moderate	0.101	0.030	0.107	0.025	0.165	0.069	0.189	0.043	0.325	0.118
		High	0.103	0.031	0.142	0.040	0.171	0.078	0.281	0.060	0.565	0.177
	Large	Low	0.018	0.009	0.017	0.007	0.026	0.012	0.032	0.016	0.050	0.023
		Moderate	0.018	0.009	0.023	0.009	0.048	0.020	0.110	0.035	0.173	0.046
		High	0.019	0.010	0.043	0.016	0.059	0.023	0.358	0.061	0.393	0.069

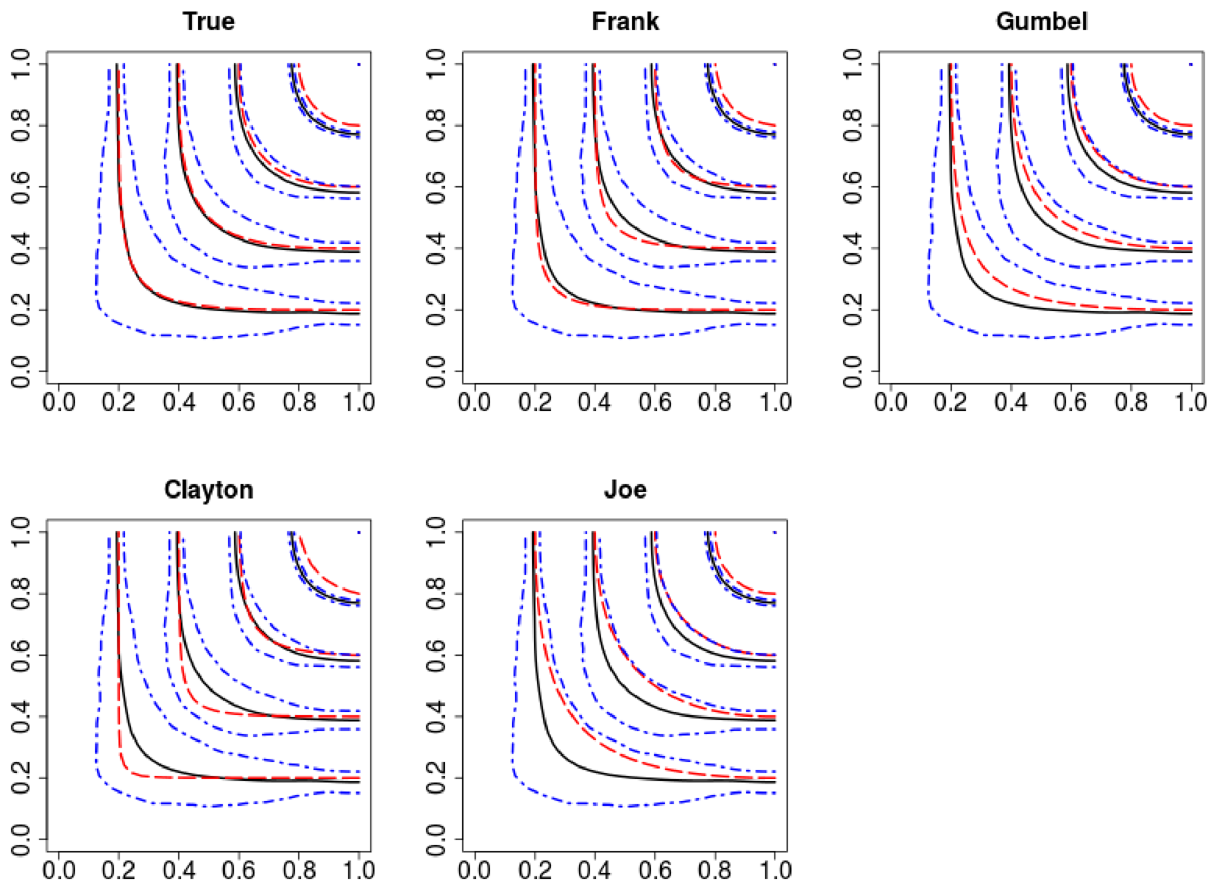

Figure B.6: Contour plots of the nonparametric estimator compared with several parametric copulas under small means and high dependence. These plots are based on a sample size of 5000.

Clayton copulas. The Frank copulas are far apart when the dependence is high.

Case 3. Joe copulas. Now we use copulas with upper tail dependence as the underlying models. Table B.4 shows the results when the Joe copula is the data generating mechanism. The true model has smallest distance. Meanwhile, the Gumbel copulas are better than the Frank and Clayton copulas due to the fact that both Gumbel and Joe copulas have upper tail dependencies.

To summarize, first, the selection of copula is more important with large marginal means and high dependence. Second, overall, our nonparametric estimator is likely to exclude

Table B.3: Distances of different parametric copulas with the nonparametric estimator when Clayton copulas are the underlying (multiplied by 1000).

			True		Frank		Gumbel		Joe	
n	Marginal Mean	Dependence	Average	sd	Average	sd	Average	sd	Average	sd
1000	Medium	Moderate	0.332	0.121	0.364	0.159	0.629	0.292	0.929	0.392
		High	0.349	0.146	0.430	0.236	1.131	0.519	2.287	0.792
	Large	Moderate	0.089	0.040	0.134	0.055	0.224	0.089	0.487	0.161
		High	0.091	0.048	0.202	0.081	0.483	0.165	1.511	0.340
5000	Medium	Moderate	0.106	0.030	0.143	0.049	0.353	0.120	0.613	0.173
		High	0.113	0.035	0.204	0.073	0.859	0.224	1.974	0.361
	Large	Moderate	0.018	0.008	0.075	0.024	0.164	0.037	0.423	0.071
		High	0.019	0.010	0.139	0.034	0.420	0.073	1.448	0.156

Table B.4: Distances of different parametric copulas with the nonparametric estimator when Joe copulas are the underlying (multiplied by 1000).

			True		Frank		Gumbel		Clayton	
n	Marginal Mean	Dependence	Average	sd	Average	sd	Average	sd	Average	sd
1000	Medium	Moderate	0.311	0.117	0.454	0.145	0.328	0.107	0.646	0.160
		High	0.320	0.129	0.540	0.210	0.391	0.156	0.956	0.247
	Large	Moderate	0.093	0.044	0.176	0.066	0.112	0.050	0.420	0.143
		High	0.085	0.041	0.233	0.087	0.142	0.060	1.473	0.246
5000	Medium	Moderate	0.091	0.026	0.251	0.061	0.126	0.036	0.398	0.066
		High	0.092	0.027	0.318	0.087	0.174	0.054	0.694	0.098
	Large	Moderate	0.018	0.009	0.109	0.026	0.044	0.016	0.353	0.061
		High	0.019	0.009	0.164	0.037	0.075	0.022	1.414	0.125

copulas with wrong tail behaviors, especially those with opposite tail dependence structures of the underlying model. In the situations where it seems ambiguous among copulas, we suggest expanding the candidate pool.

B.4 Selection of Bandwidth

First, to demonstrate sensitivity of the proposed estimator to different bandwidths under different scenarios, Figure B.7 portrays the contour plots of the nonparametric estimator with different bandwidths under different marginal mean levels. It appears that the bandwidth plays a more important role in the small and medium marginal mean settings than in the large mean cases where the estimator is not as sensitive to the selection of bandwidth. Therefore, we do not emphasize bandwidth selection for large mean cases in this section.

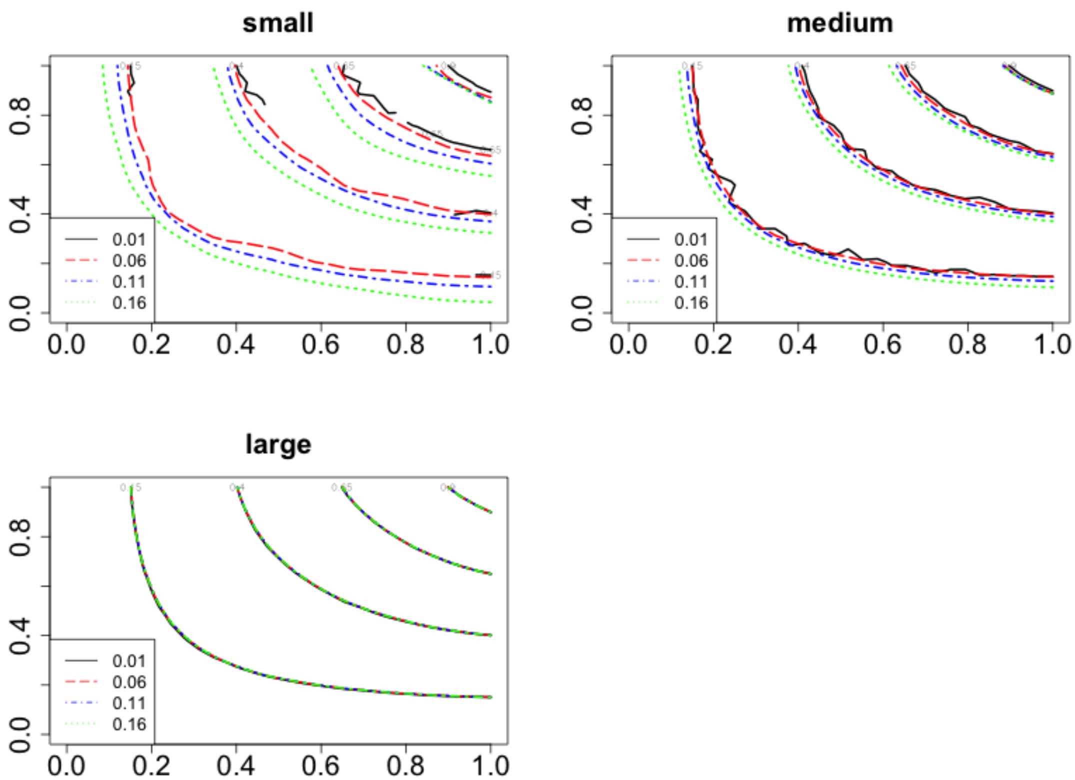

Figure B.7: Contour plots of the nonparametric estimator with different bandwidths at moderate dependence. Sample size: 1000.

We conduct a simulation study to assess the proposed bandwidth selector (Section 2.4) by comparing it with the benchmark selector minimizing the ISE values. In addition, we include the results using the independence copula as the working copula in our procedure. Tables B.5, B.6, and B.7 report the numerical results. We compare the selected bandwidths from different selectors and the resulted ISE values. We do not concern the low dependence scenarios here, since the independence copula is close to the truth in these cases and can be used as the working copula without doubt.

Table B.5 shows the results with Gaussian copulas as the underlying dependence structures. For example, when the data are generated with small marginal means and a Gaussian copula at moderate dependence level, the minimizer of the ISE values gives a bandwidth 90.653×10^{-3} on average with standard deviation 12.440×10^{-3}. With the selected bandwidths, the ISE value of our nonparametric estimator are 2.703×10^{-3} on average with standard deviation 1.202×10^{-3}. We see that the proposed procedure returns bandwidths close to the results of the benchmark selector across different marginal means, dependence levels, and sample sizes, while using the independence copula tends to undersmooth significantly, especially when there is high dependence. Intuitively, the discrepancy between the underlying copula and independence is large under high dependence. Therefore, we suggest not using the independence copula alternative for bandwidth selection when significant dependence is detected.

While Gaussian and Frank copulas have same tail dependence properties, to evaluate the proposed bandwidth selector when the underlying copula has different tail dependencies, we conduct the simulation using Gumbel and Clayton copulas to generate the data. As portrayed in Tables B.6 and B.7, it is not surprising that the proposed selector performs not as good as when Gaussian copulas are the underlying models (Table B.5). For Gumbel copulas, the selector using a Frank copula gives larger bandwidths while for Clayton copulas the bandwidths are smaller than the benchmark values. However, we think the selector performs satisfactorily even with misspecification of tail dependence, which is reflected in the ISE values of the resulted copulas.

References

Frees, E. W. (1995a) "Assessing cross-sectional correlation in panel data," *Journal of Econometrics*, Vol. 69, pp. 393–414.

——— (1995b) "Semiparametric estimation of warranty costs," *Journal of Nonparametric Statistics*, Vol. 5, pp. 103–122.

Hansen, B. E. (2009) "Lecture notes on nonparametrics."

Randles, R. H. (1984) "On tests applied to residuals," *Journal of the American Statistical Association*, Vol. 79, pp. 349–354.

Sukhatme, B. V. (1958) "Testing the hypothesis that two populations differ only in location," *The Annals of Mathematical Statistics*, pp. 60–78.

Table B.5: Bandwidths and resulting ISE values of different selectors for data generated by Gaussian copulas (multiplied by 1000).

			Small Mean						Medium Mean					
			Benchmark		Frank		Independent		Benchmark		Frank		Independent	
Dependence	n		Average	sd	Average	sd	Average	sd	Average	sd	Average	sd	Average	sd
Moderate	1000	ϵ_n	90.653	12.440	90.989	12.871	77.347	9.392	74.132	11.369	74.500	11.429	55.968	7.728
		ISE	2.703	1.202	2.718	1.201	2.855	1.170	0.330	0.119	0.333	0.119	0.390	0.148
	5000	ϵ_n	67.095	5.636	68.653	5.783	55.768	4.723	53.732	4.986	53.142	5.424	38.626	7.887
		ISE	0.852	0.335	0.856	0.335	0.937	0.331	0.096	0.026	0.096	0.026	0.127	0.044
High	1000	ϵ_n	98.779	15.995	102.021	16.831	73.516	10.405	77.189	13.036	77.226	13.972	56.447	9.446
		ISE	2.904	1.392	2.915	1.391	3.531	1.463	0.329	0.123	0.330	0.123	0.414	0.168
	5000	ϵ_n	73.495	7.404	76.653	8.910	52.695	7.960	56.532	5.963	55.058	6.723	38.111	14.678
		ISE	0.929	0.432	0.935	0.431	1.234	0.432	0.097	0.028	0.097	0.028	0.197	0.167

Table B.6: Bandwidths and resulting ISE values of different selectors for data generated by Gumbel copulas (multiplied by 1000).

			Small Mean						Medium Mean					
			Benchmark		Frank		Independent		Benchmark		Frank		Independent	
Dependence	n		Average	sd	Average	sd	Average	sd	Average	sd	Average	sd	Average	sd
Moderate	1000	ϵ_n	91.453	10.994	98.232	11.429	78.947	10.112	74.979	10.713	79.032	10.988	57.332	8.801
		ISE	2.801	1.588	2.861	1.580	2.945	1.565	0.318	0.127	0.326	0.128	0.371	0.130
	5000	ϵ_n	68.189	6.480	75.895	7.928	55.347	5.285	55.021	5.507	56.200	8.418	37.042	8.014
		ISE	0.812	0.392	0.847	0.388	0.915	0.385	0.092	0.025	0.094	0.025	0.135	0.052
High	1000	ϵ_n	99.032	14.815	105.011	16.155	75.242	13.271	79.768	12.244	82.753	15.498	58.989	11.332
		ISE	3.313	2.172	3.349	2.162	3.888	2.116	0.334	0.122	0.338	0.122	0.414	0.149
	5000	ϵ_n	73.958	8.385	80.021	10.893	52.189	8.202	58.521	6.583	58.521	10.046	37.742	14.240
		ISE	0.849	0.317	0.874	0.314	1.172	0.403	0.096	0.031	0.098	0.032	0.197	0.132

Table B.7: Bandwidths and resulting ISE values of different selectors for data generated by Clayton copulas (multiplied by 1000).

			Small Mean						Medium Mean					
			True		Frank		Independent		True		Frank		Independent	
Dependence	n		Average	sd	Average	sd	Average	sd	Average	sd	Average	sd	Average	sd
Moderate	1000	ϵ_n	93.305	11.133	89.726	10.153	81.600	9.662	74.795	11.602	72.511	9.739	58.032	8.908
		ISE	2.973	1.683	3.003	1.676	3.110	1.652	0.326	0.125	0.331	0.125	0.376	0.135
	5000	ϵ_n	68.274	6.106	65.579	5.845	57.705	6.108	54.395	6.795	53.695	5.902	39.732	8.234
		ISE	0.895	0.309	0.902	0.307	0.968	0.308	0.102	0.027	0.103	0.027	0.133	0.046
High	1000	ϵ_n	100.253	16.051	94.989	15.740	76.926	15.293	78.037	12.718	75.753	11.164	57.884	10.681
		ISE	3.570	3.838	3.606	3.831	4.170	3.750	0.334	0.154	0.337	0.154	0.413	0.171
	5000	ϵ_n	73.326	9.430	69.411	8.260	54.042	8.277	56.826	7.825	57.637	7.050	39.768	14.752
		ISE	1.006	0.388	1.021	0.386	1.287	0.434	0.105	0.030	0.107	0.030	0.196	0.135