Online Supplementary Material for Reliability Estimation of k-Out-of-n: G System with Model Uncertainty[‡]

Ziwen Gao* Dalei Yu† and Xinyu Zhang*

*Academy of Mathematics and Systems Science, Chinese Academy of Sciences, China 100190

ziwen@amss.ac.cn; xinyu@amss.ac.cn

†School of Mathematics and Statistics, Xi'an Jiaotong University China 710049 yudalei@126.com

This Supplementary Material includes the proofs for theoretical results and some numerical results. Section A1 provides the proof for Theorem 1, and Section A2 shows the proof for Theorem 2. Some numerical results are included in Section A3.

A1. Proof of Theorem 1

First, it is obvious that $\widehat{\mathbf{w}} = \operatorname{argmin}_{\mathbf{w}\in\mathcal{W}}\widetilde{\mathrm{CV}}(\mathbf{w})$. From Lemma 1 of Gao *et al.* [1], Theorem 1 is valid if the following holds:

$$\sup_{\mathbf{w}\in\mathcal{W}} \frac{|\mathrm{KL}(\mathbf{w}) - \mathrm{KL}^*(\mathbf{w})|}{\mathrm{KL}^*(\mathbf{w})} = o_p(1) \tag{A1}$$

and

$$\sup_{\mathbf{w}\in\mathcal{W}} \frac{|\widetilde{\mathrm{CV}}(\mathbf{w}) - \mathrm{KL}(\mathbf{w})|}{\mathrm{KL}^*(\mathbf{w})} = o_p(1). \tag{A2}$$

Under Condition 1, for each $\mathbf{w} \in \mathcal{W}$ and any fixed $\delta > 0$, one has

$$\Pr\left(\sum_{m=1}^{M} w_m \left\| T^{1/2}(\widehat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_m^*)\right\| > \delta\right) \le \frac{\sum_{m=1}^{M} w_m \mathbb{E}\left\| T^{1/2}(\widehat{\boldsymbol{\theta}}_m - \boldsymbol{\theta}_m^*)\right\|}{\delta} \le \frac{c_1}{\delta}.$$

Then, it is seen that under Conditions 1, 2 and 5, for each $\mathbf{w} \in \mathcal{W}$,

$$\begin{aligned} & \frac{|\mathrm{KL}(\mathbf{w}) - \mathrm{KL}^*(\mathbf{w})|}{\mathrm{KL}^*(\mathbf{w})} \\ \le\ & \xi_T^{-1} \sum_{t=1}^{T} \mathbb{E}_{\mathbf{X}^*} \left| \log\left[\sum_{m=1}^{M} w_m c_m\{\widehat{F}_1(X_{t,1}^*), \ldots, \widehat{F}_n(X_{t,n}^*); \widehat{\boldsymbol{\theta}}_m\}\right] - \log\left[\sum_{m=1}^{M} w_m c_m\{\widehat{F}_1(X_{t,1}^*), \ldots, \widehat{F}_n(X_{t,n}^*); \boldsymbol{\theta}_m^*\}\right]\right| \end{aligned}$$

[‡]The first two authors contributed equally to this work and should be considered co-first authors, and the corresponding author is Xinyu Zhang.

$$
\begin{aligned}
&= \xi_T^{-1}\sum_{t=1}^{T}\mathbb{E}_{\mathbf{X}^*}\left|\sum_{m=1}^{M} w_m \mathbf{v}_m^\top\left\{\widehat{F}_1(X_{t,1}^*),\ldots,\widehat{F}_n(X_{t,n}^*),\mathbf{w},\bar{\boldsymbol{\theta}}_m\right\}(\widehat{\boldsymbol{\theta}}_m-\boldsymbol{\theta}_m^*)\right| \\
&\le \xi_T^{-1}\sum_{t=1}^{T}\sum_{m=1}^{M} w_m\left\|\widehat{\boldsymbol{\theta}}_m-\boldsymbol{\theta}_m^*\right\|\mathbb{E}_{\mathbf{X}^*}\left\|\mathbf{v}_m\left\{\widehat{F}_1(X_{t,1}^*),\ldots,\widehat{F}_n(X_{t,n}^*),\mathbf{w},\bar{\boldsymbol{\theta}}_m\right\}\right\| \\
&\le c_1\xi_T^{-1}\sum_{t=1}^{T}\sum_{m=1}^{M} w_m\left\|\widehat{\boldsymbol{\theta}}_m-\boldsymbol{\theta}_m^*\right\| \\
&= O_p(T^{1/2}\xi_T^{-1}) \\
&= o_p(1), \qquad \text{(A3)}
\end{aligned}
$$

where $\bar{\boldsymbol{\theta}}_m$ lies between $\widehat{\boldsymbol{\theta}}_m$ and $\boldsymbol{\theta}_m^*$ and we have used the fact that $\left\|\widehat{\boldsymbol{\theta}}_m-\boldsymbol{\theta}_m^*\right\|$ is independent to $\mathbf{X}^*$.

In addition, for any $\mathbf{w}\in\mathcal{W}$,

$$
\begin{aligned}
&\frac{|\widetilde{\mathrm{CV}}(\mathbf{w})-\mathrm{KL}^*(\mathbf{w})|}{\mathrm{KL}^*(\mathbf{w})} \\
=\;& \{\mathrm{KL}^*(\mathbf{w})\}^{-1}\left|\sum_{l=1}^{J}\sum_{h=1}^{H}\log\left[\sum_{m=1}^{M} w_m c_m\left\{\widehat{F}_1(X_{(l-1)H+h,1}),\ldots,\widehat{F}_n(X_{(l-1)H+h,n});\widehat{\boldsymbol{\theta}}_{m,[-l]}\right\}\right]\right. \\
&\left.-\sum_{l=1}^{J}\sum_{h=1}^{H}\mathbb{E}_{\mathbf{X}^*}\left(\log\left[\sum_{m=1}^{M} w_m c_m\left\{\widehat{F}_1(X_{(l-1)H+h,1}^*),\ldots,\widehat{F}_n(X_{(l-1)H+h,n}^*);\boldsymbol{\theta}_m^*\right\}\right]\right)\right| \\
\le\;& \xi_T^{-1}\left|\sum_{l=1}^{J}\sum_{h=1}^{H}\log\left[\sum_{m=1}^{M} w_m c_m\left\{\widehat{F}_1(X_{(l-1)H+h,1}),\ldots,\widehat{F}_n(X_{(l-1)H+h,n});\widehat{\boldsymbol{\theta}}_{m,[-l]}\right\}\right]\right. \\
&\left.-\sum_{l=1}^{J}\sum_{h=1}^{H}\log\left[\sum_{m=1}^{M} w_m c_m\left\{\widehat{F}_1(X_{(l-1)H+h,1}),\ldots,\widehat{F}_n(X_{(l-1)H+h,n});\boldsymbol{\theta}_m^*\right\}\right]\right| \\
&+\xi_T^{-1}\left|\sum_{l=1}^{J}\sum_{h=1}^{H}\log\left\{\sum_{m=1}^{M} w_m c_m\left(\widehat{F}_1(X_{(l-1)H+h,1}),\ldots,\widehat{F}_n(X_{(l-1)H+h,n});\boldsymbol{\theta}_m^*\right)\right\}\right. \\
&\left.-\sum_{l=1}^{J}\sum_{h=1}^{H}\mathbb{E}_{\mathbf{X}^*}\left(\log\left[\sum_{m=1}^{M} w_m c_m\left\{\widehat{F}_1(X_{(l-1)H+h,1}^*),\ldots,\widehat{F}_n(X_{(l-1)H+h,n}^*);\boldsymbol{\theta}_m^*\right\}\right]\right)\right|. \qquad \text{(A4)}
\end{aligned}
$$

Similar to the proof of (A1), under Conditions 1, 2 and 5, for each $\mathbf{w}\in\mathcal{W}$, one has

$$
\begin{aligned}
&\xi_T^{-1}\left|\sum_{l=1}^{J}\sum_{h=1}^{H}\log\left[\sum_{m=1}^{M} w_m c_m\left\{\widehat{F}_1(X_{(l-1)H+h,1}),\ldots,\widehat{F}_n(X_{(l-1)H+h,n});\widehat{\boldsymbol{\theta}}_{m,[-l]}\right\}\right]\right. \\
&\left.-\sum_{l=1}^{J}\sum_{h=1}^{H}\log\left[\sum_{m=1}^{M} w_m c_m\left\{\widehat{F}_1(X_{(l-1)H+h,1}),\ldots,\widehat{F}_n(X_{(l-1)H+h,n});\boldsymbol{\theta}_m^*\right\}\right]\right| \\
&= o_p(1). \qquad \text{(A5)}
\end{aligned}
$$

In addition, note that

$$
\begin{aligned}
&\xi_T^{-1}\left|\sum_{l=1}^{J}\sum_{h=1}^{H}\log\left[\sum_{m=1}^{M} w_m c_m\left\{\widehat{F}_1(X_{(l-1)H+h,1}),\ldots,\widehat{F}_n(X_{(l-1)H+h,n});\boldsymbol{\theta}_m^*\right\}\right]\right. \\
&\left.-\sum_{l=1}^{J}\sum_{h=1}^{H}\mathbb{E}_{\mathbf{X}^*}\left(\log\left[\sum_{m=1}^{M} w_m c_m\left\{\widehat{F}_1(X_{(l-1)H+h,1}^*),\ldots,\widehat{F}_n(X_{(l-1)H+h,n}^*);\boldsymbol{\theta}_m^*\right\}\right]\right)\right|
\end{aligned}
$$

$$
\begin{aligned}
&\leq \xi_T^{-1}\left|\sum_{l=1}^{J}\sum_{h=1}^{H}\log\left[\sum_{m=1}^{M} w_m c_m\left\{\widehat{F}_1(X_{(l-1)H+h,1}),\ldots,\widehat{F}_n(X_{(l-1)H+h,n});\boldsymbol{\theta}_m^*\right\}\right]\right.\\
&\quad\left.-\sum_{l=1}^{J}\sum_{h=1}^{H}\log\left[\sum_{m=1}^{M} w_m c_m\left\{F_1(X_{(l-1)H+h,1}),\ldots,F_n(X_{(l-1)H+h,n});\boldsymbol{\theta}_m^*\right\}\right]\right|\\
&\quad+\xi_T^{-1}\left|\sum_{l=1}^{J}\sum_{h=1}^{H}\log\left[\sum_{m=1}^{M} w_m c_m\left\{F_1(X_{(l-1)H+h,1}),\ldots,F_n(X_{(l-1)H+h,n});\boldsymbol{\theta}_m^*\right\}\right]\right.\\
&\quad\left.-\sum_{l=1}^{J}\sum_{h=1}^{H}\mathbb{E}_{\mathbf{X}^*}\left(\log\left[\sum_{m=1}^{M} w_m c_m\left\{F_1(X^*_{(l-1)H+h,1}),\ldots,F_n(X^*_{(l-1)H+h,n});\boldsymbol{\theta}_m^*\right\}\right]\right)\right|\\
&\quad+\xi_T^{-1}\left|\sum_{l=1}^{J}\sum_{h=1}^{H}\mathbb{E}_{\mathbf{X}^*}\left(\log\left[\sum_{m=1}^{M} w_m c_m\left\{F_1(X^*_{(l-1)H+h,1}),\ldots,F_n(X^*_{(l-1)H+h,n});\boldsymbol{\theta}_m^*\right\}\right]\right)\right.\\
&\quad\left.-\sum_{l=1}^{J}\sum_{h=1}^{H}\mathbb{E}_{\mathbf{X}^*}\left(\log\left[\sum_{m=1}^{M} w_m c_m\left\{\widehat{F}_1(X^*_{(l-1)H+h,1}),\ldots,\widehat{F}_n(X^*_{(l-1)H+h,n});\boldsymbol{\theta}_m^*\right\}\right]\right)\right|. \qquad \text{(A6)}
\end{aligned}
$$

Under Conditions 2 and 5, by the DKW inequality after Dvoretzky, Kiefer and Wolfowitz [2, page 268], one has

$$
\begin{aligned}
&\xi_T^{-1}\left|\sum_{l=1}^{J}\sum_{h=1}^{H}\log\left[\sum_{m=1}^{M} w_m c_m\left\{\widehat{F}_1(X_{(l-1)H+h,1}),\ldots,\widehat{F}_n(X_{(l-1)H+h,n});\boldsymbol{\theta}_m^*\right\}\right]\right.\\
&\left.-\sum_{l=1}^{J}\sum_{h=1}^{H}\log\left[\sum_{m=1}^{M} w_m c_m\left\{F_1(X_{(l-1)H+h,1}),\ldots,F_n(X_{(l-1)H+h,n});\boldsymbol{\theta}_m^*\right\}\right]\right|\\
&\leq \xi_T^{-1}\sum_{l=1}^{J}\sum_{h=1}^{H}\sum_{m=1}^{M} w_m\sum_{j=1}^{n}\left|r_{m,j}(\bar{F}_1,\ldots,\bar{F}_n,\mathbf{w},\boldsymbol{\theta}_m^*)\{\widehat{F}_j(X_{(l-1)H+h,j})-F_j(X_{(l-1)H+h,j})\}\right|\\
&\leq \xi_T^{-1}\sum_{l=1}^{J}\sum_{h=1}^{H}c_1\sum_{j=1}^{n}\sup_{x\in\mathbb{R}^1}\left|\widehat{F}_j(x)-F_j(x)\right|\\
&= O_p(T^{1/2}\xi_T^{-1})\\
&= o_p(1), \qquad \text{(A7)}
\end{aligned}
$$

where $\bar{F}_j$ lies between $\widehat{F}_j(X_{(l-1)H+h,j})$ and $F_j(X_{(l-1)H+h,j})$. In addition, under Conditions 3 and 5, by the Central Limit Theorem, it is seen that

$$
\begin{aligned}
&\xi_T^{-1}\left|\sum_{l=1}^{J}\sum_{h=1}^{H}\log\left[\sum_{m=1}^{M} w_m c_m\left\{F_1(X_{(l-1)H+h,1}),\ldots,F_n(X_{(l-1)H+h,n});\boldsymbol{\theta}_m^*\right\}\right]\right.\\
&\left.-\sum_{l=1}^{J}\sum_{h=1}^{H}\mathbb{E}_{\mathbf{X}^*}\left(\log\left[\sum_{m=1}^{M} w_m c_m\left\{F_1(X^*_{(l-1)H+h,1}),\ldots,F_n(X^*_{(l-1)H+h,n});\boldsymbol{\theta}_m^*\right\}\right]\right)\right|\\
&= \xi_T^{-1}\left|\sum_{l=1}^{J}\sum_{h=1}^{H}\log\left[\sum_{m=1}^{M} w_m c_m\left\{F_1(X_{(l-1)H+h,1}),\ldots,F_n(X_{(l-1)H+h,n});\boldsymbol{\theta}_m^*\right\}\right]\right.\\
&\quad\left.-\sum_{l=1}^{J}\sum_{h=1}^{H}\mathbb{E}_{\mathbf{X}}\left(\log\left[\sum_{m=1}^{M} w_m c_m\left\{F_1(X_{(l-1)H+h,1}),\ldots,F_n(X_{(l-1)H+h,n});\boldsymbol{\theta}_m^*\right\}\right]\right)\right|\\
&= o_p(1). \qquad \text{(A8)}
\end{aligned}
$$

Similar to the derivation of (A7), one has

$$\begin{aligned}
&\xi_T^{-1}\left|\sum_{l=1}^{J}\sum_{h=1}^{H}\mathbb{E}_{\mathbf{X}^*}\left(\log\left[\sum_{m=1}^{M} w_m c_m\left\{F_1(X^*_{(l-1)H+h,1}),\ldots,F_n(X^*_{(l-1)H+h,n});\boldsymbol{\theta}^*_m\right\}\right]\right)\right.\\
&\left.-\sum_{l=1}^{J}\sum_{h=1}^{H}\mathbb{E}_{\mathbf{X}^*}\left(\log\left[\sum_{m=1}^{M} w_m c_m\left\{\widehat{F}_1(X^*_{(l-1)H+h,1}),\ldots,\widehat{F}_n(X^*_{(l-1)H+h,n});\boldsymbol{\theta}^*_m\right\}\right]\right)\right|\\
&\leq c_1 T\xi_T^{-1}\sum_{j=1}^{n}\sup_{x\in\mathbb{R}^1}\left|\widehat{F}_j(x)-F_j(x)\right|\\
&= o_p(1),
\end{aligned} \tag{A9}$$

where the first inequality is owning to the fact that the Kolmogorov-Smirnov statistic $\sup_{x\in\mathbb{R}^1}|\widehat{F}_j(x)-F_j(x)|$ is based on $\mathbf{X}$ and is irrelevant to $\mathbf{X}^*$. By combining (A7)–(A9) with (A6), we have

$$\begin{aligned}
&\xi_T^{-1}\left|\sum_{l=1}^{J}\sum_{h=1}^{H}\log\left[\sum_{m=1}^{M} w_m c_m\left\{\widehat{F}_1(X_{(l-1)H+h,1}),\ldots,\widehat{F}_n(X_{(l-1)H+h,n});\boldsymbol{\theta}^*_m\right\}\right]\right.\\
&\left.-\sum_{l=1}^{J}\sum_{h=1}^{H}\mathbb{E}_{\mathbf{X}^*}\left(\log\left[\sum_{m=1}^{M} w_m c_m\left\{\widehat{F}_1(X^*_{(l-1)H+h,1}),\ldots,\widehat{F}_n(X^*_{(l-1)H+h,n});\boldsymbol{\theta}^*_m\right\}\right]\right)\right|\\
&= o_p(1).
\end{aligned}$$

Then, by combining this with (A4) and (A5), it is readily seen that for any $\mathbf{w}\in\mathcal{W}$,

$$\frac{|\widetilde{\mathrm{CV}}(\mathbf{w})-\mathrm{KL}^*(\mathbf{w})|}{\mathrm{KL}^*(\mathbf{w})}=o_p(1). \tag{A10}$$

Now, by the pointwise convergence established in (A3) and (A10), Condition 4 and the proof of Theorem 2 of [3], we obtain (A1) and (A2). This completes the proof.

A2. Proof of Theorem 2

Let $\mathbf{w}_{m^*}$ be the vector whose m^*th component is one and the others are zeros, then, under Condition 2 and the third part of Condition 6, one has

$$\frac{1}{T}\mathrm{KL}^*(\mathbf{w}_{m^*})=o_p(1). \tag{A11}$$

In addition, under Conditions 1-3, by the proof of (A10), we know that for any $\mathbf{w}\in\mathcal{W}$,

$$\frac{|\widetilde{\mathrm{CV}}(\mathbf{w})-\mathrm{KL}^*(\mathbf{w})|}{T}=O_p(T^{-1/2}). \tag{A12}$$

Now, recall that for any $\mathbf{w},\mathbf{w}'\in\mathcal{W}$, one has, under Condition 6,

$$\begin{aligned}
&\frac{|\widetilde{\mathrm{CV}}(\mathbf{w})-\mathrm{KL}^*(\mathbf{w})-\widetilde{\mathrm{CV}}(\mathbf{w}')+\mathrm{KL}^*(\mathbf{w}')|}{T}\\
\leq\;&\frac{1}{T}\left|\sum_{l=1}^{J}\sum_{h=1}^{H}\log\left[\sum_{m=1}^{M} w_m c_m\left\{\widehat{F}_1(X_{(l-1)H+h,1}),\ldots,\widehat{F}_n(X_{(l-1)H+h,n});\widehat{\boldsymbol{\theta}}_{m,[-l]}\right\}\right]\right.\\
&\left.-\sum_{l=1}^{J}\sum_{h=1}^{H}\log\left[\sum_{m=1}^{M} w'_m c_m\left\{\widehat{F}_1(X_{(l-1)H+h,1}),\ldots,\widehat{F}_n(X_{(l-1)H+h,n});\widehat{\boldsymbol{\theta}}_{m,[-l]}\right\}\right]\right|
\end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{T}\left|\sum_{l=1}^{J}\sum_{h=1}^{H}\mathbb{E}_{\mathbf{X}^*}\left(\log\left[\sum_{m=1}^{M} w_m c_m\left\{\widehat{F}_1(X^*_{(l-1)H+h,1}),\ldots,\widehat{F}_n(X^*_{(l-1)H+h,n});\boldsymbol{\theta}^*_m\right\}\right]\right)\right.\\
&\left.-\sum_{l=1}^{J}\sum_{h=1}^{H}\mathbb{E}_{\mathbf{X}^*}\left(\log\left[\sum_{m=1}^{M} w'_m c_m\left\{\widehat{F}_1(X^*_{(l-1)H+h,1}),\ldots,\widehat{F}_n(X^*_{(l-1)H+h,n});\boldsymbol{\theta}^*_m\right\}\right]\right)\right|\\
\leq\ & 2c_1\|\mathbf{w}-\mathbf{w}'\|_1,
\end{aligned}
$$

where $\|\cdot\|_1$ is the L_1-norm. This indicates that $\{\widetilde{\mathrm{CV}}(\mathbf{w})-\mathrm{KL}^*(\mathbf{w})\}/T$ is equicontinuous. Then, similar to the proof of Theorem 1 in [3], by Corollary 2.2 of [4] and pointwise convergence established in (A12), one has

$$
\sup_{\mathbf{w}\in\mathcal{W}}\left|\frac{\widetilde{\mathrm{CV}}(\mathbf{w})-\mathrm{KL}^*(\mathbf{w})}{T}\right| = o_p(1). \tag{A13}
$$

In addition, since $\widehat{\mathbf{w}} = \mathrm{argmin}_{\mathbf{w}\in\mathcal{W}}\widetilde{\mathrm{CV}}(\mathbf{w})$, we have $\widetilde{\mathrm{CV}}(\widehat{\mathbf{w}}) \leq \widetilde{\mathrm{CV}}(\mathbf{w}_{m^*})$, which along with (A11) and (A13), implies

$$
\frac{\mathrm{KL}^*(\widehat{\mathbf{w}})}{T} \leq \frac{\mathrm{KL}^*(\mathbf{w}_{m^*})}{T} + o_p(1). \tag{A14}
$$

From (A11) and (A14), we obtain

$$
\frac{\mathrm{KL}^*(\widehat{\mathbf{w}})}{T} = o_p(1). \tag{A15}
$$

In addition, define $\mathbf{X}^*_t = (X^*_{t,1},\ldots,X^*_{t,n})$, i.e., $f(\mathbf{X}^*_t) = f(X^*_{t,1},\ldots,X^*_{t,n})$, and let

$$
\Xi = \frac{\prod_{t=1}^{T}\left\{\sum_{m=1}^{M}\widehat{w}_m f_m(\mathbf{X}^*_t;\boldsymbol{\theta}^*_m)\right\}}{\prod_{t=1}^{T}\left[\sum_{m=1}^{M}\widehat{w}_m c_m\{\widehat{F}_1(X^*_{t,1}),\ldots,\widehat{F}_n(X^*_{t,n});\boldsymbol{\theta}^*_m\}\prod_{j=1}^{n}\widehat{f}_j(X^*_{t,j})\right]}.
$$

Then, we have

$$
\begin{aligned}
\mathrm{KL}^*(\widehat{\mathbf{w}}) &= \mathbb{E}_{\mathbf{X}^*}\left[\log\left\{\prod_{t=1}^{T} f(\mathbf{X}^*_t)\right\} - \log\left\{\prod_{t=1}^{T}\widehat{f}_{\widehat{\mathbf{w}}}(\mathbf{X}^*_t;\boldsymbol{\theta}^*)\right\}\right]\\
&= \mathbb{E}_{\mathbf{X}^*}\left[\log\left\{\frac{\prod_{t=1}^{T} f(\mathbf{X}^*_t)}{\prod_{t=1}^{T}\left\{\sum_{m=1}^{M}\widehat{w}_m f_m(\mathbf{X}^*_t;\boldsymbol{\theta}^*_m)\right\}}\right\}\right] + \mathbb{E}_{\mathbf{X}^*}\{\log(\Xi)\}\\
&= \mathbb{E}_{\mathbf{X}^*}\left[\log\left\{\frac{\prod_{t=1}^{T} f(\mathbf{X}^*_t)}{\prod_{t=1}^{T}\left\{\sum_{m\in\mathcal{D}}\widehat{w}_m f_m(\mathbf{X}^*_t;\boldsymbol{\theta}^*_m) + \sum_{m\notin\mathcal{D}}\widehat{w}_m f_m(\mathbf{X}^*_t;\boldsymbol{\theta}^*_m)\right\}}\right\}\right] + \mathbb{E}_{\mathbf{X}^*}\{\log(\Xi)\}\\
&= \mathbb{E}_{\mathbf{X}^*}\left[\log\left\{\prod_{t=1}^{T}\frac{f(\mathbf{X}^*_t)}{\Gamma(\widehat{\mathbf{w}})f(\mathbf{X}^*_t) + \sum_{m\notin\mathcal{D}}\widehat{w}_m f_m(\mathbf{X}^*_t;\boldsymbol{\theta}^*_m)}\right\}\right] + \mathbb{E}_{\mathbf{X}^*}\{\log(\Xi)\},\\
&= \mathbb{E}_{\mathbf{X}^*}\left[\log\left\{\prod_{t=1}^{T}\frac{f(\mathbf{X}^*_t)}{f^\star(\widehat{\mathbf{w}}, X^*_{t,1},\ldots,X^*_{t,n};\boldsymbol{\theta}^*_0)}\right\}\right] + \mathbb{E}_{\mathbf{X}^*}\{\log(\Xi)\},
\end{aligned} \tag{A16}
$$

where the last second equality is from the definition of correctly specified model. For Ξ, as $\sup_{x\in\mathbb{R}^1}|\widehat{F}_j(x) - F_j(x)| = O_p(T^{-1/2})$ and $\sup_{x\in\mathbb{R}^1}|\widehat{f}_j(x) - f_j(x)| = o_p(1)$ for $j = 1,\ldots,n$ as $T\to\infty$, under Condition 6, we have

$$\begin{aligned} 0 &\leq \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{n}\mathbb{E}_{\mathbf{X}^*}\left[\log\left\{f_j(X_{t,j}^*)\right\}-\log\left\{\widehat{f}_j(X_{t,j}^*)\right\}\right] \\ &= \sum_{j=1}^{n}\mathbb{E}_{\mathbf{X}_1^*}\left[\log\left\{f_j(X_{1,j}^*)\right\}-\log\left\{\widehat{f}_j(X_{1,j}^*)\right\}\right] \\ &\leq \sum_{j=1}^{n}\mathbb{E}_{\mathbf{X}_1^*}\left\{\frac{f_j(X_{1,j}^*)}{\widehat{f}_j(X_{1,j}^*)}-1\right\} \\ &= \sum_{j=1}^{n}\int_{-\infty}^{\infty}\frac{f_j(x)-\widehat{f}_j(x)}{\widehat{f}_j(x)}f_j(x)dx \\ &\leq \max_{1\leq j\leq n}\sup_{x\in\mathbb{R}^1}\left|f_j(x)-\widehat{f}_j(x)\right|\sum_{j=1}^{n}\int_{-\infty}^{\infty}\frac{f_j(x)}{\widehat{f}_j(x)}dx \\ &= \max_{1\leq j\leq n}\sup_{x\in\mathbb{R}^1}\left|f_j(x)-\widehat{f}_j(x)\right|\sum_{j=1}^{n}\mathbb{E}_{\mathbf{X}_1^*}\left\{\widehat{f}_j^{-1}(X_{1,j}^*)\right\} \\ &= o_p(1), \end{aligned}$$

further,

$$\begin{aligned} \mathbb{E}_{\mathbf{X}^*}\left\{\frac{\log(\Xi)}{T}\right\} &= \frac{1}{T}\mathbb{E}_{\mathbf{X}^*}\left\{\log\left(\frac{\prod_{t=1}^{T}\left\{\sum_{m=1}^{M}\widehat{w}_m f_m(\mathbf{X}_t^*;\boldsymbol{\theta}_m^*)\right\}}{\prod_{t=1}^{T}\left[\sum_{m=1}^{M}\widehat{w}_m c_m\{\widehat{F}_1(X_{t,1}^*),\ldots,\widehat{F}_n(X_{t,n}^*);\boldsymbol{\theta}_m^*\}\prod_{j=1}^{n}\widehat{f}_j(X_{t,j}^*)\right]}\right)\right\} \\ &= \frac{1}{T}\mathbb{E}_{\mathbf{X}^*}\left\{\log\left(\frac{\prod_{t=1}^{T}\left[\sum_{m=1}^{M}\widehat{w}_m c_m\{F_1(X_{t,1}^*),\ldots,F_n(X_{t,n}^*);\boldsymbol{\theta}_m^*\}\prod_{j=1}^{n}f_j(X_{t,j}^*)\right]}{\prod_{t=1}^{T}\left[\sum_{m=1}^{M}\widehat{w}_m c_m\{\widehat{F}_1(X_{t,1}^*),\ldots,\widehat{F}_n(X_{t,n}^*);\boldsymbol{\theta}_m^*\}\prod_{j=1}^{n}\widehat{f}_j(X_{t,j}^*)\right]}\right)\right\} \\ &= o_p(1), \end{aligned}$$

which with (A15) indicates that

$$\frac{1}{T}\mathbb{E}_{\mathbf{X}^*}\left[\log\left\{\prod_{t=1}^{T}\frac{f(\mathbf{X}_t^*)}{f^{\star}(\widehat{\mathbf{w}},X_{t,1}^*,\ldots,X_{t,n}^*;\boldsymbol{\theta}_0^*)}\right\}\right]=o_p(1).$$

It is readily seen that for any $\epsilon>0$, under Condition 6, for any fixed $\epsilon>0$

$$\text{Pr}\left(1-\Gamma(\widehat{\mathbf{w}})>\epsilon\right) \leq \text{Pr}\left(\frac{1}{T}\mathbb{E}_{\mathbf{X}^*}\left[\log\left\{\prod_{t=1}^{T}\frac{f(\mathbf{X}_t^*)}{f^{\star}(\widehat{\mathbf{w}},X_{t,1}^*,\ldots,X_{t,n}^*;\boldsymbol{\theta}_0^*)}\right\}\right]>c_0\right)\to 0$$

and this concludes the proof.

A3. More Simulation Results

We supplement a simulation using 3-variate t-distribution as DGP to generate the observations. The degree of freedom and location parameter in the 3-variate t-distribution are set as 4 and $(1,2,3)^\top$, respectively. The positive definite scale matrix $\boldsymbol{\Sigma}=0.5(\mathbf{I}_3+\mathbf{1}_{3\times1}\mathbf{1}_{3\times1}^\top)$, where $\mathbf{I}_3$ is identity matrix and $\mathbf{1}_{3\times1}=(1,1,1)^\top$. The simulation results are shown in Table A1. The KL loss of 10-KLMA is significantly smaller than that of the other methods, while the L_1 loss and L_2 loss of all methods are relatively close.

TABLE A1
COMPARISON OF KL, L_1 AND L_2 LOSSES FOR DIFFERENT METHODS.

Loss	T		5-KLMA	10-KLMA	CVMA	CW	QMLE	BICMS	SBIC	EWMA
KL	200	Mean	13.60	**13.25**	15.86	16.91	13.86	19.94	18.79	17.90
		Median	13.76	**13.44**	16.49	16.09	14.03	19.67	18.55	17.41
	500	Mean	26.51	**26.00**	32.21	32.70	27.64	45.63	43.77	41.42
		Median	28.17	**27.79**	33.49	28.48	29.66	43.45	42.87	40.94
L_1	200	Mean	6.30	6.30	6.32	6.33	6.34	6.28	6.28	**6.26**
		Median	6.09	6.07	6.11	6.13	6.10	6.08	6.08	**6.04**
	500	Mean	5.32	5.32	5.34	5.34	5.34	5.31	5.31	**5.28**
		Median	5.35	5.34	5.37	5.36	5.35	5.36	5.36	**5.31**
L_2	200	Mean	**0.46**	**0.46**	**0.46**	0.47	**0.46**	**0.46**	**0.46**	0.47
		Median	**0.42**	**0.42**	**0.42**	0.43	**0.42**	**0.42**	**0.42**	0.43
	500	Mean	0.38	0.38	**0.37**	0.38	0.38	**0.37**	**0.37**	0.39
		Median	0.37	**0.36**	0.37	0.37	0.37	**0.36**	**0.36**	0.38

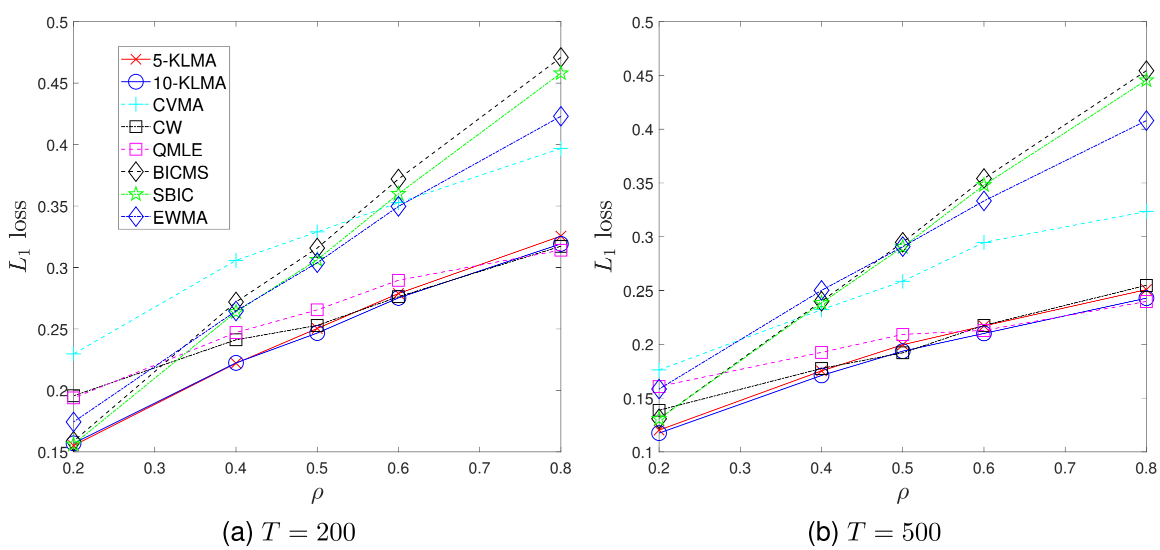


Fig. A1. L_1 loss comparison of different methods when $k = 1, n = 3$ and DGP is composed of Gumbel and Student-t copulas.

TABLE A2
COMPARISON OF KL LOSS FOR DIFFERENT METHODS BASED ON 500 REPETITIONS.

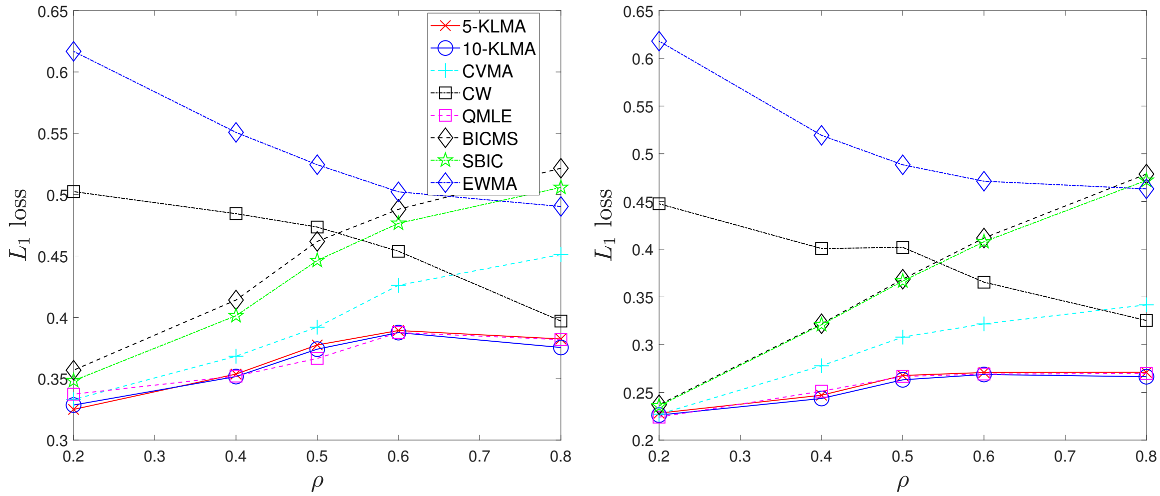
DGP	T	KL loss	5-KLMA	10-KLMA	CVMA	CW	QMLE	BICMS	SBIC	EWMA
Gumbel(2.5)+Normal(0.5)	200	Mean	4.72	**4.60**	9.50	9.01	5.27	11.30	9.83	5.90
		Median	4.56	**4.40**	11.16	8.29	5.03	10.73	9.28	5.40
		SD	2.15	2.16	4.30	8.23	2.93	2.91	2.91	1.51
	500	Mean	6.87	**6.32**	15.36	21.76	7.96	24.58	22.69	13.96
		Median	6.62	**6.10**	13.61	14.09	7.26	22.52	21.99	13.23
		SD	3.25	3.29	9.48	82.19	5.15	4.04	4.71	2.52
Gumbel(2.5)+Joe(3)	200	Mean	3.05	**3.02**	10.93	7.13	9.55	6.32	5.22	10.53
		Median	2.32	**2.29**	6.62	5.28	6.12	5.56	5.08	9.76
		SD	2.82	2.83	10.89	12.22	8.65	2.09	2.76	1.97
	500	Mean	3.02	**3.01**	10.86	12.63	18.99	13.71	11.83	25.25
		Median	**2.16**	2.27	7.64	10.89	8.97	12.95	12.54	24.45
		SD	2.64	2.50	11.24	23.71	21.12	2.01	3.99	2.83
Gumbel(2.5)+Frank(2)	200	Mean	5.57	**5.28**	12.50	8.96	6.15	13.29	12.04	9.07
		Median	4.96	**4.87**	10.05	7.54	5.19	11.26	10.80	8.48
		SD	3.00	2.80	8.97	13.65	6.23	4.17	3.22	1.58
	500	Mean	8.23	**7.62**	19.23	14.32	10.40	28.11	27.18	21.76
		Median	7.73	**7.19**	14.67	12.67	8.03	25.94	25.85	21.02
		SD	3.82	3.78	14.65	36.50	11.56	5.81	4.62	2.35
Normal(0.5)+Joe(3)	200	Mean	4.83	**4.62**	9.93	8.19	6.23	9.85	8.99	5.71
		Median	4.58	**4.44**	8.73	7.45	4.93	7.73	7.58	5.12
		SD	2.47	2.47	6.38	8.15	5.66	4.48	3.35	1.77
	500	Mean	7.50	**6.36**	12.87	13.05	10.22	19.20	18.94	13.49
		Median	6.76	**5.83**	10.34	12.17	7.12	18.22	18.20	12.67
		SD	4.08	3.77	10.06	18.94	9.54	4.00	2.69	2.62
Normal(0.5)+Frank(2)	200	Mean	1.76	**1.75**	2.49	5.23	2.56	3.15	2.20	3.16
		Median	1.51	**1.50**	2.00	2.60	2.05	2.50	1.81	2.85
		SD	1.35	1.33	2.31	23.77	2.89	1.85	1.72	0.86
	500	Mean	**2.05**	2.06	3.79	8.20	4.60	5.79	4.53	7.30
		Median	**1.60**	1.61	3.96	5.69	3.18	5.57	4.30	6.86
		SD	1.52	1.53	2.20	28.48	15.57	2.40	2.47	1.36
Joe(3)+Frank(2)	200	Mean	5.77	**5.34**	11.03	7.16	7.07	12.00	11.05	9.37
		Median	5.48	**5.10**	9.55	7.20	6.00	10.30	10.03	8.83
		SD	2.67	2.63	7.51	4.36	4.55	3.44	2.56	1.60
	500	Mean	9.50	**8.23**	16.26	15.81	13.03	25.66	25.15	22.49
		Median	9.13	**8.05**	15.43	8.45	10.01	24.34	24.28	21.93
		SD	4.41	4.32	9.94	59.28	10.32	3.90	2.89	2.27

Note: The smallest loss in each Mean row and Median row is in boldface type.
Abbreviations: Gumbel(2.5)+Normal(0.5): 0.5Gumbel(2.5)+0.5Normal(0.5), others are similar. SD: standard deviation.

TABLE A3
COMPARISON OF L_2 LOSS FOR DIFFERENT METHODS BASED ON 500 REPETITIONS ($\times 10^{-3}$).

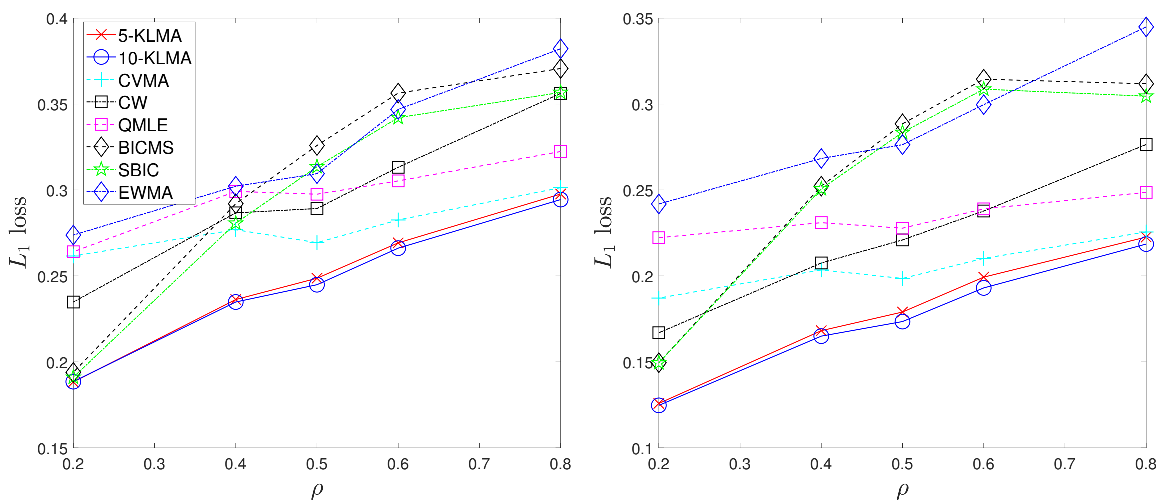
DGP	T	L_2 loss	5-KLMA	10-KLMA	CVMA	CW	QMLE	BICMS	SBIC	EWMA
Gumbel(2.5)+Normal(0.5)	200	Mean	0.97	**0.95**	1.31	2.19	1.09	2.25	1.89	1.31
		Median	0.73	**0.72**	1.16	1.83	0.87	2.13	2.05	1.11
		SD	0.83	0.78	0.86	2.26	0.89	0.94	0.94	0.53
	500	Mean	0.53	**0.48**	0.85	1.73	0.64	2.01	1.83	1.08
		Median	0.40	**0.36**	0.74	1.41	0.50	2.08	2.06	1.00
		SD	0.45	0.41	0.55	2.58	0.51	0.49	0.64	0.21
Gumbel(2.5)+Joe(3)	200	Mean	0.56	**0.55**	2.27	1.29	1.85	1.37	1.13	2.18
		Median	**0.39**	0.41	1.45	1.19	1.11	1.29	1.21	2.10
		SD	0.53	0.53	2.20	0.93	2.05	0.34	0.50	0.33
	500	Mean	0.21	**0.21**	0.96	1.08	1.56	1.28	1.11	2.11
		Median	0.15	**0.15**	0.65	1.03	0.56	1.28	1.23	2.07
		SD	0.20	0.19	1.04	0.55	2.01	0.19	0.37	0.16
Gumbel(2.5)+Frank(2)	200	Mean	1.54	**1.44**	2.93	2.67	1.62	3.05	2.64	1.54
		Median	1.18	**1.09**	2.40	2.00	1.14	2.15	2.00	1.28
		SD	1.39	1.23	2.36	2.79	1.96	2.01	1.73	0.69
	500	Mean	0.83	**0.74**	1.80	1.65	0.98	2.28	2.13	1.21
		Median	0.63	**0.59**	1.27	1.23	0.70	1.81	1.79	1.10
		SD	0.69	0.60	1.56	2.61	1.44	1.28	1.10	0.28
Normal(0.5)+Joe(3)	200	Mean	1.40	1.33	3.03	3.12	2.02	2.21	1.93	**0.58**
		Median	1.04	0.94	2.87	2.49	1.49	1.13	1.17	**0.26**
		SD	1.28	1.24	2.16	2.69	1.82	2.40	1.87	0.90
	500	Mean	0.83	0.69	1.56	1.76	1.41	1.26	1.22	**0.34**
		Median	0.63	0.48	1.04	0.97	0.87	0.99	1.00	**0.18**
		SD	0.69	0.65	1.39	2.14	1.39	0.81	0.70	0.42
Normal(0.5)+Frank(2)	200	Mean	1.36	1.34	**1.33**	2.67	1.45	1.92	1.48	4.20
		Median	0.96	0.92	**0.81**	1.01	0.81	1.00	0.85	3.56
		SD	1.44	1.39	1.98	6.61	2.20	2.22	1.66	1.75
	500	Mean	**0.57**	0.58	0.72	1.76	0.84	1.15	0.94	3.98
		Median	0.36	**0.35**	0.55	0.74	0.42	0.75	0.67	3.69
		SD	0.67	0.67	0.72	4.20	2.70	1.36	1.09	1.16
Joe(3)+Frank(2)	200	Mean	2.05	**1.92**	3.92	3.48	3.04	3.36	2.95	1.96
		Median	1.58	**1.46**	2.94	2.94	2.54	1.93	2.08	1.51
		SD	1.66	1.56	3.15	2.71	2.27	2.70	2.10	1.25
	500	Mean	1.27	**1.09**	2.03	2.22	2.20	2.13	2.02	1.60
		Median	1.03	**0.86**	1.50	1.44	1.54	1.53	1.56	1.40
		SD	0.92	0.86	1.64	3.76	1.90	1.66	1.25	0.59

Note: The smallest loss in each Mean row and Median row is in boldface type.
Abbreviations: Gumbel(2.5)+Normal(0.5): 0.5Gumbel(2.5)+0.5Normal(0.5), others are similar. SD: standard deviation.



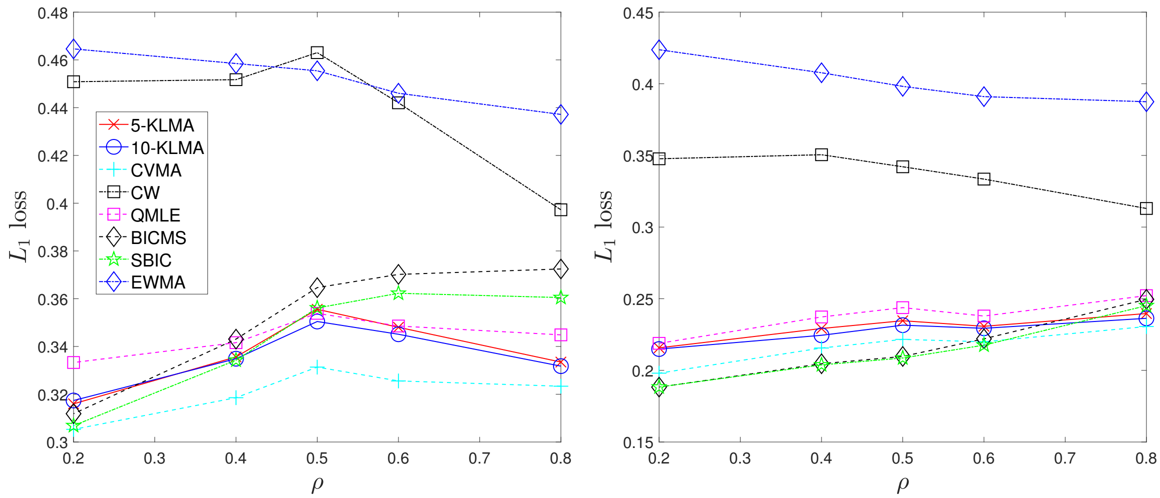
(a) $T = 200$ (b) $T = 500$

Fig. A2. L_1 loss comparison of different methods when $k = 1, n = 3$ and DGP is composed of Normal and Student-t copulas.



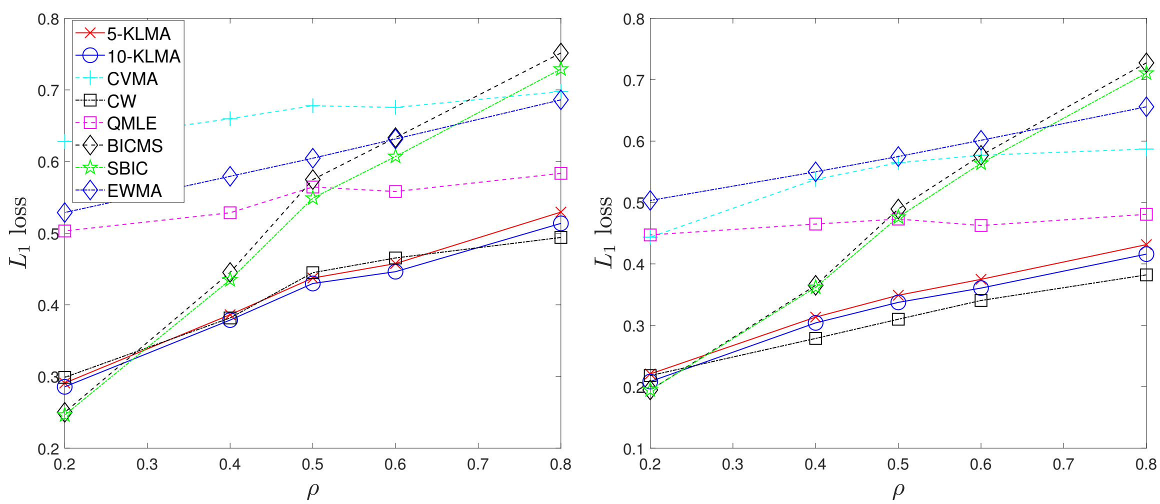
(a) $T = 200$ (b) $T = 500$

Fig. A3. L_1 loss comparison of different methods when $k = 3, n = 3$ and DGP is composed of Gumbel and Student-t copulas.



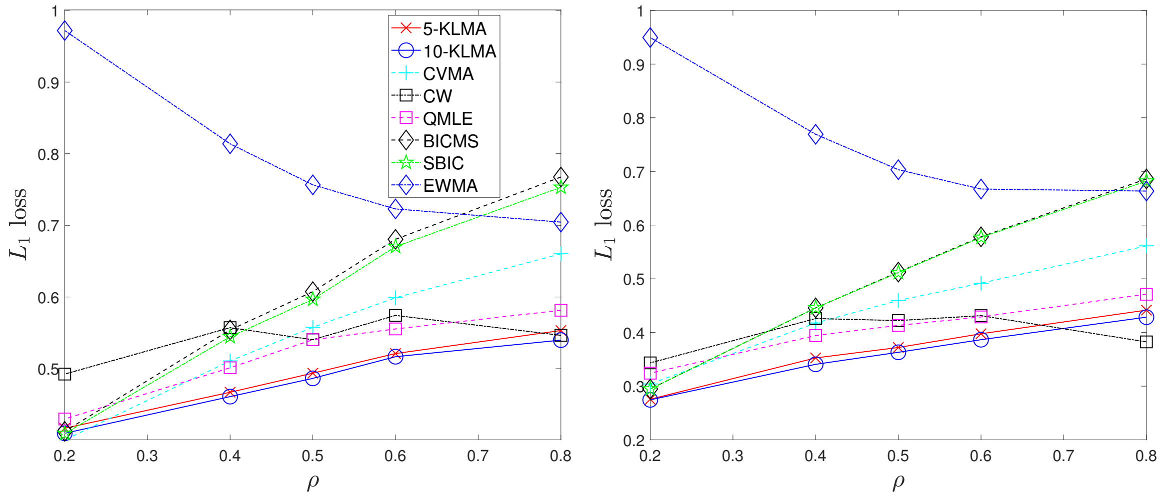
(a) $T = 200$ (b) $T = 500$

Fig. A4. L_1 loss comparison of different methods when $k = 3, n = 3$ and DGP is composed of Normal and Student-t copulas



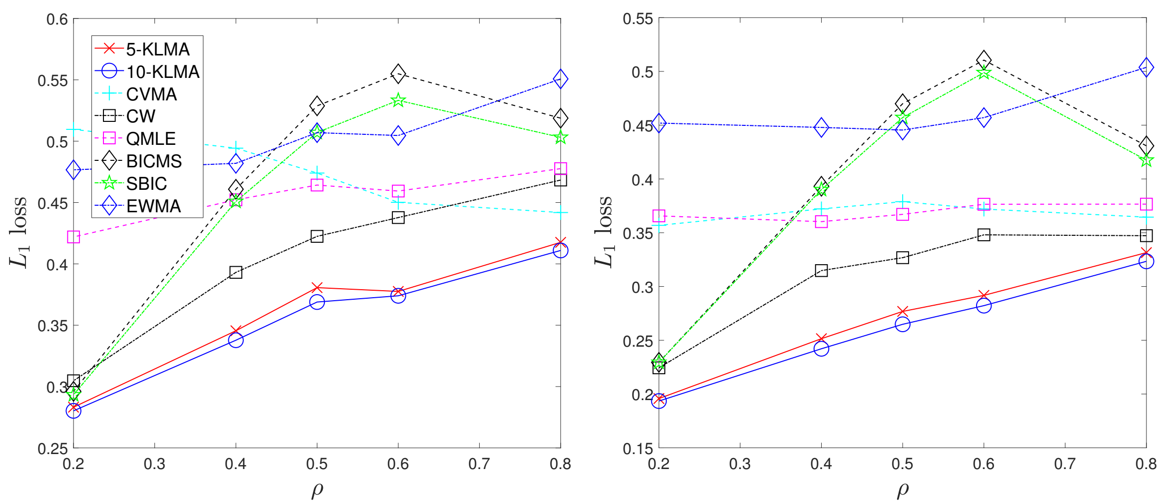
(a) $T = 200$ (b) $T = 500$

Fig. A5. L_1 loss comparison of different methods when $k = 2, n = 4$ and DGP is composed of Gumbel and Student-t copulas.



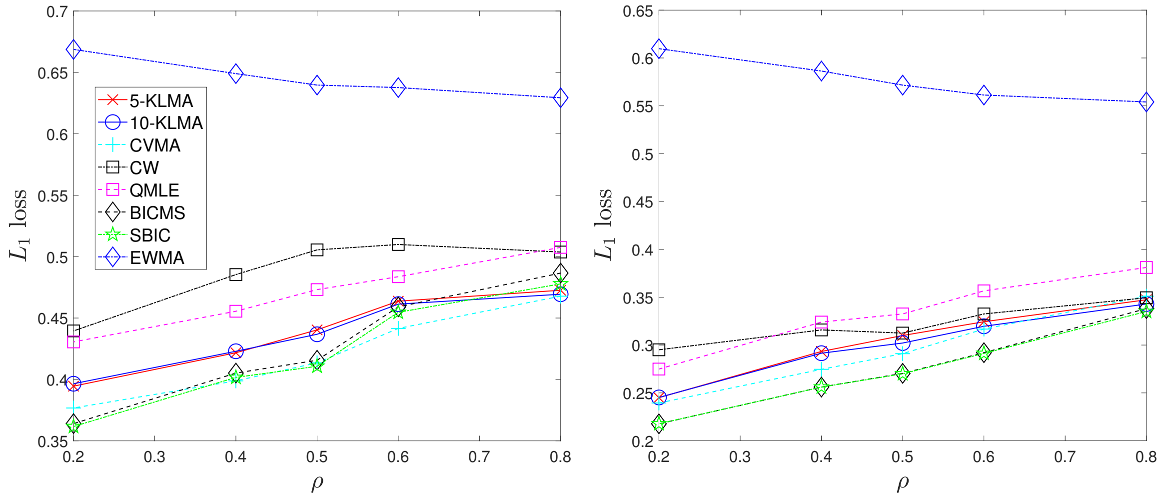
(a) $T = 200$ (b) $T = 500$

Fig. A6. L_1 loss comparison of different methods when $k = 2, n = 4$ and DGP is composed of Normal and Student-t copulas.



(a) $T = 200$ (b) $T = 500$

Fig. A7. L_1 loss comparison of different methods when $k = 4, n = 4$ and DGP is composed of Gumbel and Student-t copulas.



(a) $T = 200$ (b) $T = 500$

Fig. A8. L_1 loss comparison of different methods when $k = 4, n = 4$ and DGP is composed of Normal and Student-t copulas.

TABLE A4
COMPARISON OF KL LOSS FOR DIFFERENT METHODS BASED ON 500 REPETITIONS. THE BEST, SECOND BEST, AND THIRD BEST METHODS IN EACH CASE ARE FLAGGED BY ①, ② AND ③ RESPECTIVELY.

DGP	KL loss	10-KLMA	CVMA	CW	QMLE	BICMS	SBIC	EWMA
10-KLMA	Mean	15.86①	64.63	22.50③	19.92②	90.42	87.86	69.56
	Median	15.56①	75.27	19.47③	19.15②	86.92	86.47	68.93
	SD	5.51	38.40	22.78	6.30	7.61	7.30	3.31
CVMA	Mean	7.53①	55.96	9.93③	8.33②	169.21	168.68	165.66
	Median	2.89②	14.70	2.95③	2.49①	167.91	167.87	165.39
	SD	8.67	71.30	18.45	11.33	6.67	4.72	4.35
CW	Mean	8.65①	50.47	10.13③	9.73②	187.01	185.94	177.78
	Median	4.23②	8.63	4.38③	3.91①	185.32	185.16	177.31
	SD	9.59	72.17	20.06	12.88	7.04	5.75	4.45
QMLE	Mean	7.34①	46.88	9.08③	8.35②	169.13	168.44	164.96
	Median	2.99③	7.30	2.37①	2.61②	167.92	167.88	164.62
	SD	9.76	67.18	14.73	13.16	6.52	4.39	3.94
BICMS	Mean	0.93③	2.32	6.67	2.44	0.51①	0.51①	32.68
	Median	0.62	0.65	0.50③	1.22	0.26①	0.26①	32.34
	SD	1.13	5.32	16.92	3.40	0.70	0.70	1.78
SBIC	Mean	0.96③	2.06	6.19	2.33	0.60①	0.60①	31.47
	Median	0.59③	0.69	0.55	1.24	0.32①	0.32①	31.05
	SD	1.08	3.89	18.21	3.24	0.76	0.76	1.64
EWMA	Mean	3.63②	23.82	12.20	10.68③	12.57	11.61	2.96①
	Median	3.28②	23.79	7.25③	8.01	10.81	10.75	2.55①
	SD	1.97	12.28	56.39	8.07	5.78	3.95	1.86

SD: standard deviation.

TABLE A5
COMPARISON OF L_2 LOSS FOR DIFFERENT METHODS BASED ON 500 REPETITIONS. THE BEST, SECOND BEST, AND THIRD BEST METHODS IN EACH CASE ARE FLAGGED BY ①, ② AND ③ RESPECTIVELY.

DGP	L_2 loss ($\times 10^{-3}$)	10-KLMA	CVMA	CW	QMLE	BICMS	SBIC	EWMA
10-KLMA	Mean	6.87①	29.87	14.20③	7.20②	17.45	16.56	17.25
	Median	4.38①	24.32	7.43③	5.14②	12.06	11.62	14.07
	SD	7.53	25.31	52.79	6.78	14.78	14.20	13.15
CVMA	Mean	4.30①	9.35	8.68③	5.08②	20.34	20.28	31.11
	Median	2.46①	5.85	3.01③	2.49②	18.57	18.36	27.17
	SD	5.14	9.84	30.08	7.32	11.71	11.74	15.28
CW	Mean	4.04①	8.51③	8.62	5.20②	21.13	20.89	24.66
	Median	2.14①	4.84	2.88③	2.40②	18.44	18.16	21.03
	SD	4.95	10.29	41.97	7.00	12.49	12.59	14.61
QMLE	Mean	3.90①	7.89	6.37③	4.98②	18.84	18.65	28.28
	Median	2.12①	3.85	2.50③	2.17②	16.83	16.27	25.36
	SD	5.36	10.18	25.82	9.36	11.57	11.57	14.47
BICMS	Mean	1.48③	2.68	4.03	3.37	1.25①	1.25①	26.44
	Median	0.83③	1.10	0.84	1.25	0.65①	0.65①	24.50
	SD	2.46	4.87	8.97	5.78	1.72	1.72	7.06
SBIC	Mean	1.49③	2.57	3.95	3.65	1.22①	1.22①	25.82
	Median	0.73	0.94	0.70③	1.34	0.55①	0.56②	24.21
	SD	2.41	4.44	12.38	6.54	1.80	1.80	6.42
EWMA	Mean	7.20①	37.32	14.73	11.04	10.05③	9.39②	10.71
	Median	4.74①	35.89	5.44②	9.41	6.98	6.48③	8.89
	SD	7.44	22.96	88.63	9.95	8.80	8.12	8.60

SD: standard deviation.

REFERENCES

[1] Y. Gao, X. Zhang, S. Wang, T. Chong, and G. Zou, "Frequentist model averaging for threshold models," *Annals of the Institute of Statistical Mathematics*, vol. 71, no. 2, pp. 275–306, 2019.

[2] A. W. van der Vaart, *Asymptotic Statistics*, ser. Cambridge Series in Statistical and Probabilistic Mathematics. New York, New York: Cambridge Univeristy Press, 2000.

[3] D. Yu, X. Zhang, and H. Liang, "Unified optimal model averaging with a general loss function based on cross-validation," *Available at SSRN*, 2022.

[4] W. K. Newey, "Uniform convergence in probability and stochastic equicontinuity," *Econometrica*, vol. 59, no. 4, pp. 1161–1167, 1991.