Supplementary Material: Global Depths for Irregular Observed Multivariate Functional Data

Zhuo Qu[1], Wenlin Dai[2] and Marc G. Genton[1]

September 6, 2022

This supplementary material provides the results of Models 1-4 in the Simulation Studies and the proofs of theorems. In Section 1, we first present one simulation of all the above models with contamination but no sparseness. Next, we demonstrate the robustness of depths from extracting the median and central region, keeping the rank association for nonoutliers, and simple outlier detection for the above models. In Section 2, we first recall the definitions of global multivariate functional integrated depth and global multivariate functional extremal depth, respectively. Then we show the proofs of Theorems 1-3.

---

[1]Statistics Program, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia. E-mail: zhuo.qu@kaust.edu.sa, marc.genton@kaust.edu.sa

[2]Institute of Statistics and Big Data, Renmin University of China, Beijing 100872, China. E-mail: wenlin.dai@ruc.edu.cn
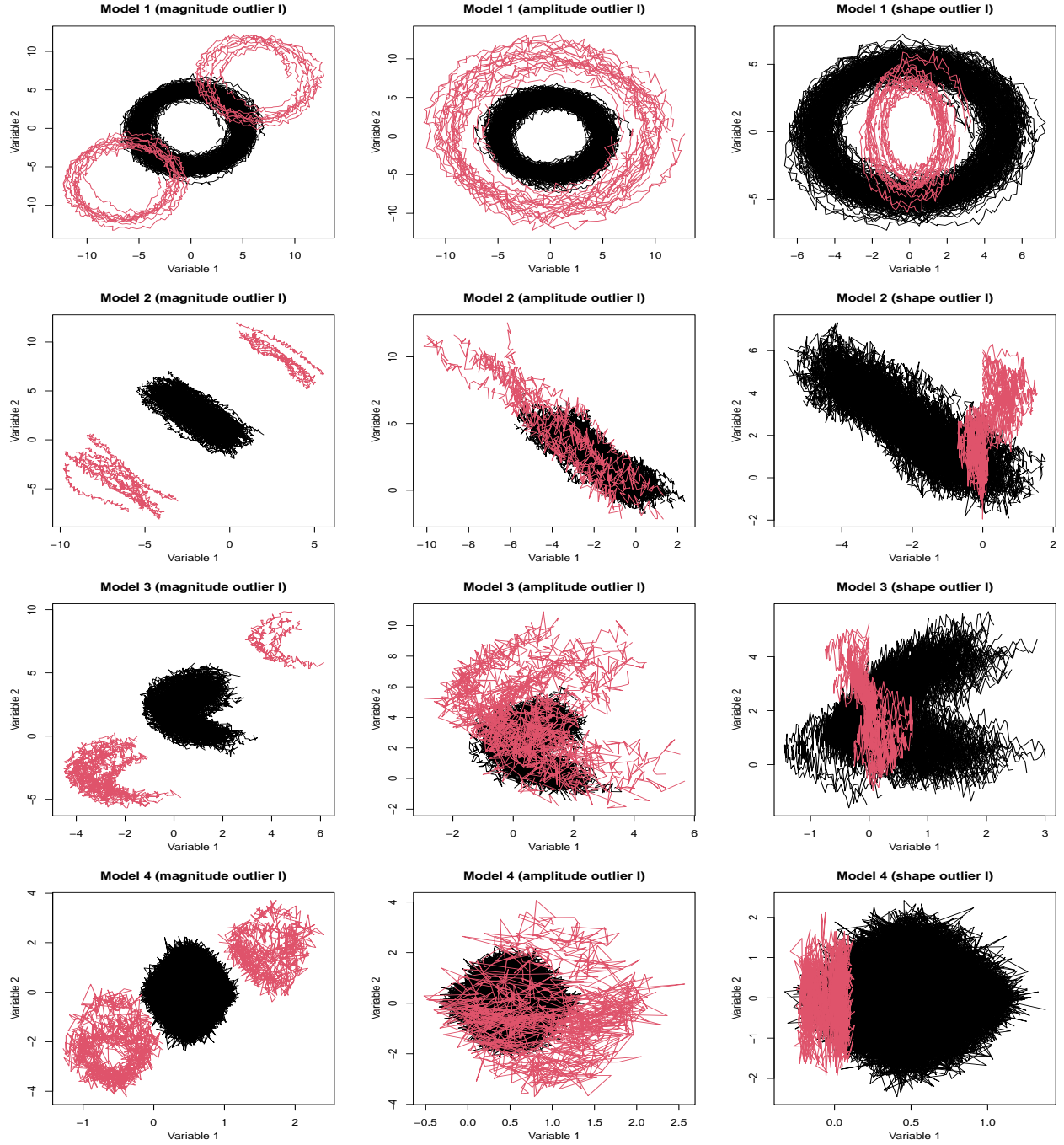
1 Simulation Study

Figure S1: Panels from top to bottom show Models 1-4 contaminated by different types of outliers. Each panel displays the model contaminated by a 10% proportion of magnitude outlier I, amplitude outlier I, and shape outlier I, respectively, from left to right. The non-outlying and outlying curves are colored in black and red, respectively.

Figure S2: The left to tight shows Model 1 with magnitude outlier I, amplitude outlier I, and shape outlier I, respectively. ASE of the estimated central curve (the first row), of the 50%-dispersion curve (the second row), and the outlier proportion in the lowest 10% depth region (the third row), the Spearman correlation (the fourth row), and the running time (the fifth row). The sparseness type is point sparseness with a dense case (0%) and a high sparseness case (40%-60%).
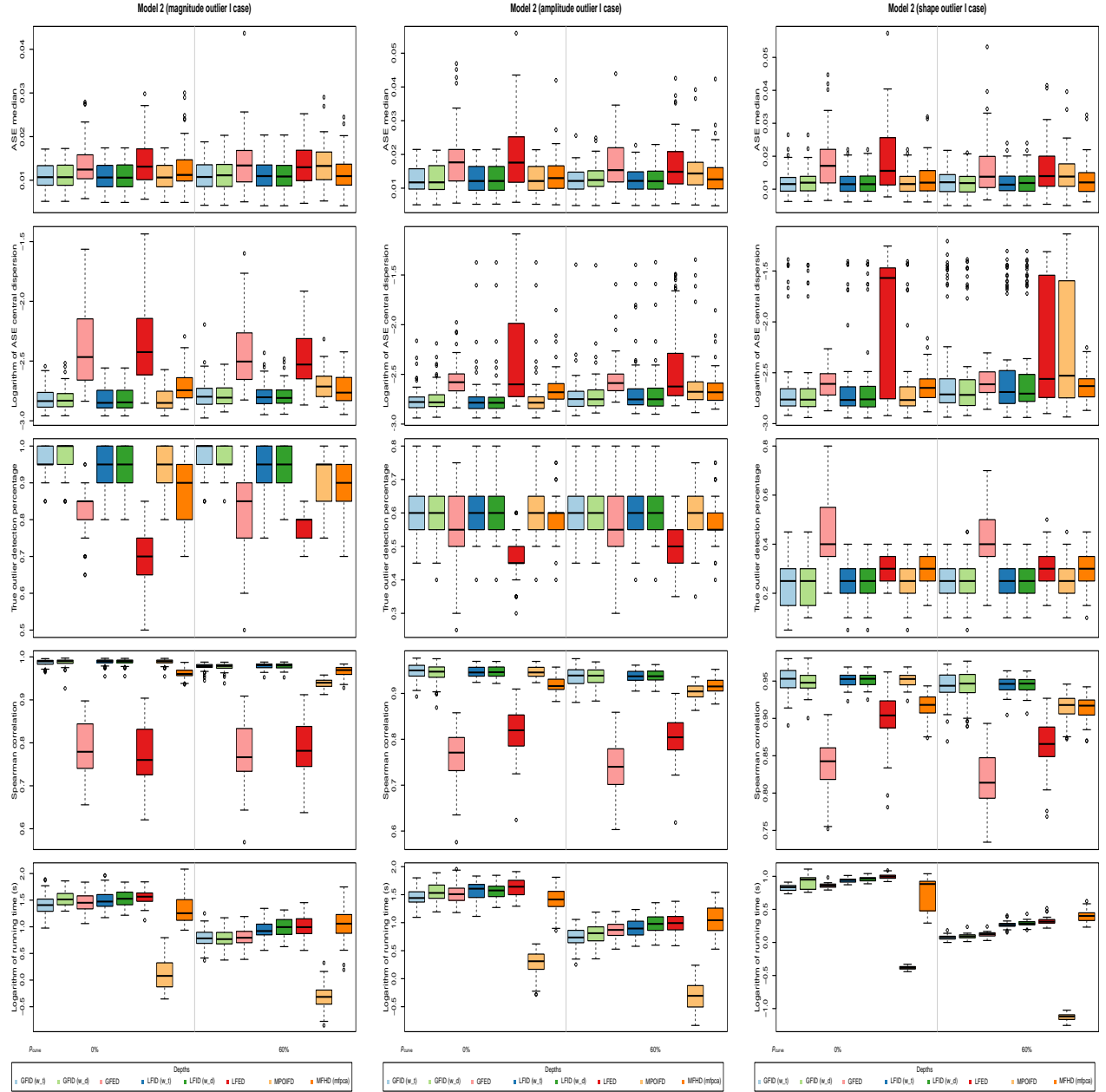
Figure S3: The left to tight shows Model 2 with magnitude outlier I, amplitude outlier I, and shape outlier I, respectively. ASE of the estimated central curve (the first row), of the 50%-dispersion curve (the second row), and the outlier proportion in the lowest 10% depth region (the third row), the Spearman correlation (the fourth row), and the running time (the fifth row). The sparseness type is point sparseness with a dense case (0%) and a high sparseness case (40%-60%).

Figure S4: The left to tight shows Model 3 with magnitude outlier I, amplitude outlier I, and shape outlier I, respectively. ASE of the estimated central curve (the first row), of the 50%-dispersion curve (the second row), and the outlier proportion in the lowest 10% depth region (the third row), the Spearman correlation (the fourth row), and the running time (the fifth row). The sparseness type is point sparseness with a dense case (0%) and a high sparseness case (40%-60%).
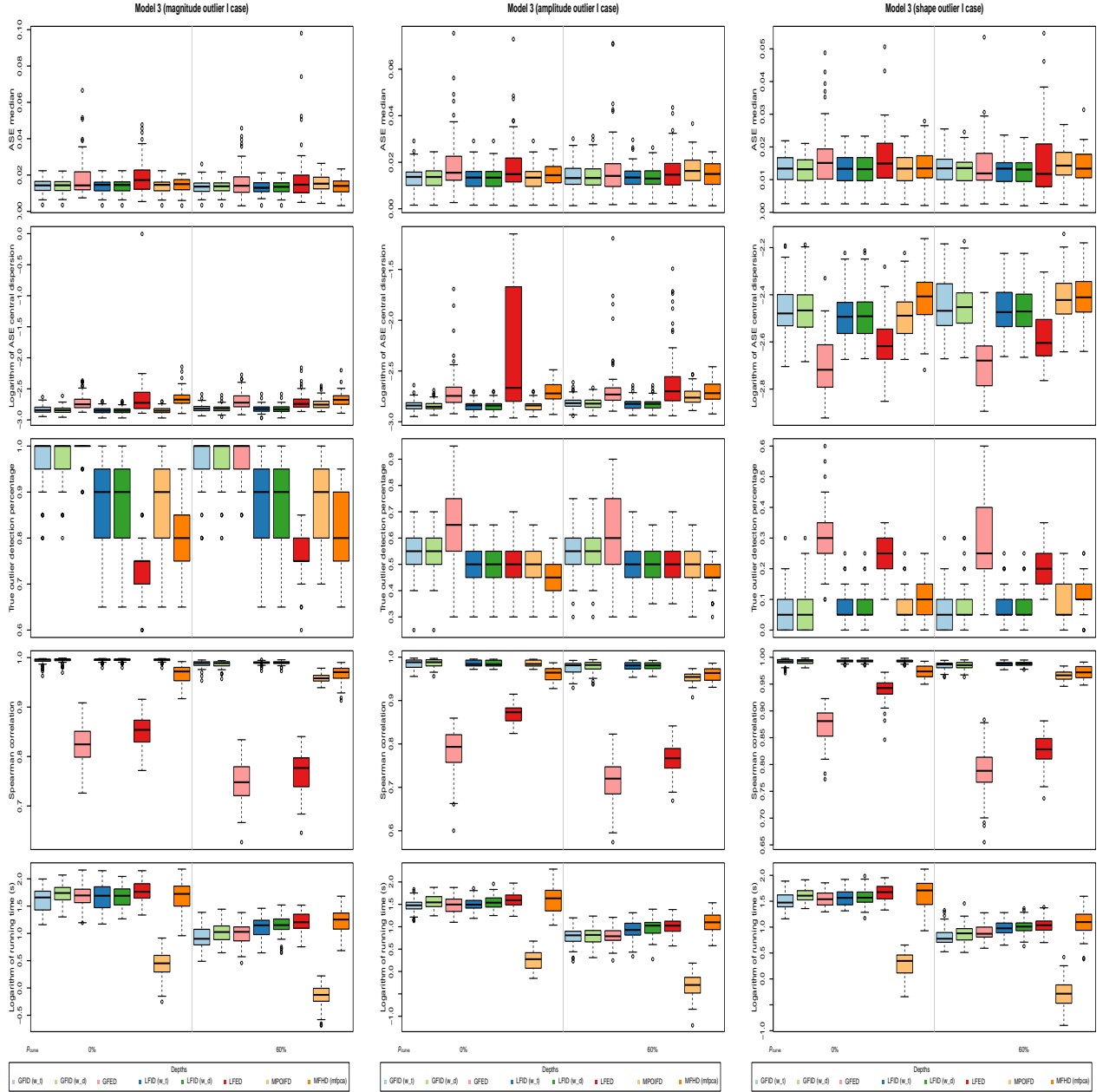
Figure S5: The left to tight shows Model 4 with magnitude outlier I, amplitude outlier I, and shape outlier I, respectively. ASE of the estimated central curve (the first row), of the 50%-dispersion curve (the second row), and the outlier proportion in the lowest 10% depth region (the third row), the Spearman correlation (the fourth row), and the running time (the fifth row). The sparseness type is point sparseness with a dense case (0%) and a high sparseness case (40%-60%).
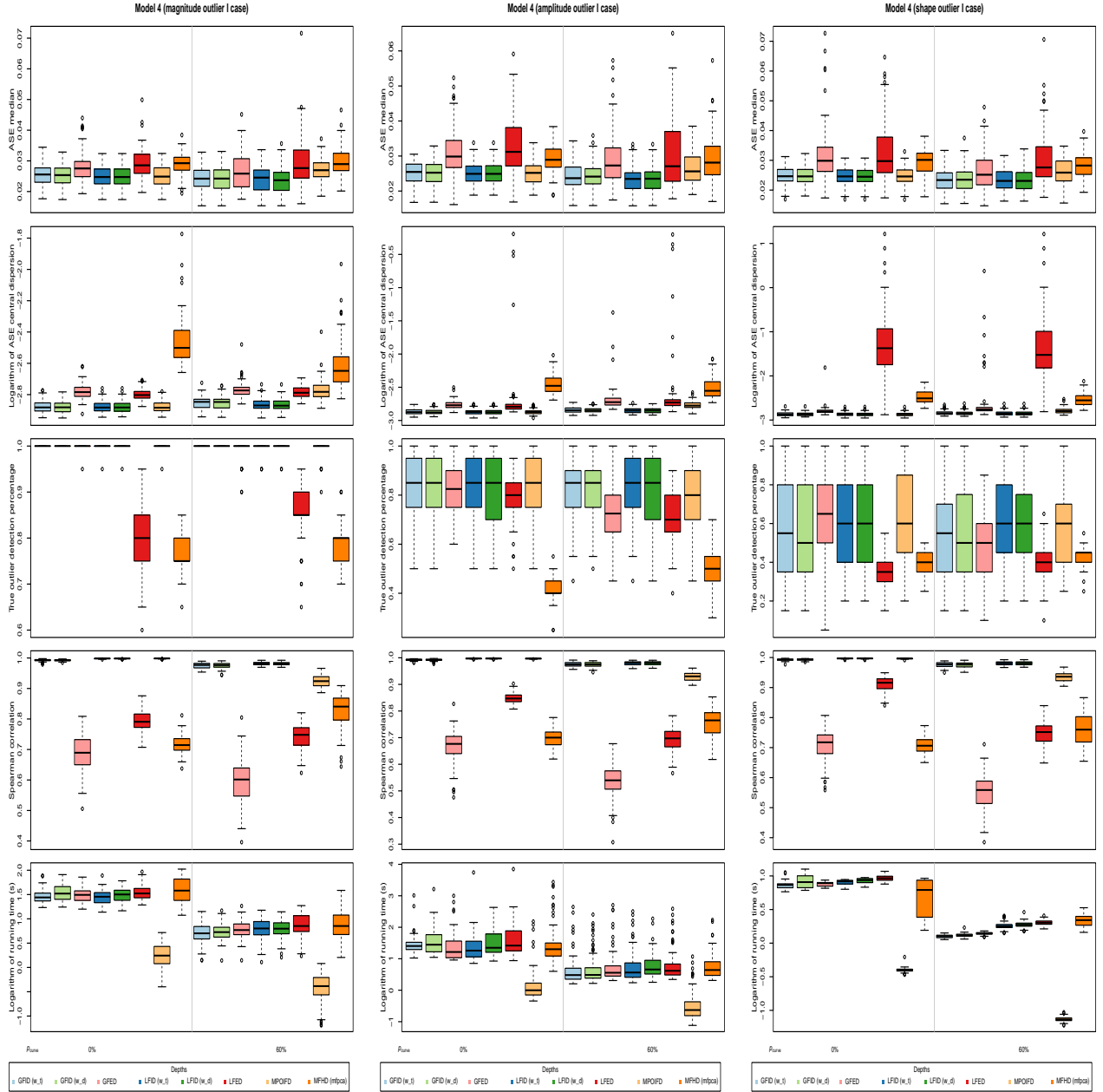
2 Proofs

Before the proof, recall the definition of GMFD.

Definition 1 *Global Multivariate Functional Depth* (GMFD, Population Version).

Assume Conditions (A), (B.1) and (C), and the stochastic process $\boldsymbol{\mathcal{Y}}^*(t)$ across time t is with cdf $F_{\boldsymbol{S}}$. From an arbitrary $\boldsymbol{X} \in \mathcal{C}^p(\mathcal{T})$, we obtain $\boldsymbol{X}^*(t) \in \mathbb{R}^p$ ($t \in \mathcal{T}$).

*1. The **global multivariate functional integrated depth** (GMFID) is*

$$GMFID(\boldsymbol{X}; F_{\boldsymbol{\mathcal{Y}}}, \beta) = \int_{\mathcal{T}} D(\boldsymbol{X}^*(t); F_{\boldsymbol{S}}) \cdot w_\beta(t; F_{\boldsymbol{S}}) \mathrm{d}t, \tag{1}$$

where $w_\beta(t; F_{\boldsymbol{S}})$ is the definition of $w_\beta(t; F_{\boldsymbol{\mathcal{Y}}(t)})$ with the cdf $F_{\boldsymbol{S}}$ replacing $F_{\boldsymbol{\mathcal{Y}}^*(t)}$.

*2. The **global multivariate functional extremal depth** (GMFED) of $\boldsymbol{X}$ is*

$$MFED(\boldsymbol{X}; F_{\boldsymbol{\mathcal{Y}}}) = 1 - P(\boldsymbol{X} \prec \boldsymbol{Y}) = P(\boldsymbol{X} \succeq \boldsymbol{Y}), \text{ where } \boldsymbol{Y} \sim F_{\boldsymbol{\mathcal{Y}}}, \tag{2}$$

where $\Psi_{\boldsymbol{X}^*}(r) = \int_{\mathcal{T}} \mathbf{1}\{D(\boldsymbol{X}^*(t); F_{\boldsymbol{S}}) \leq r\} \mathrm{d}t$.

Proof of Theorem 1

1) We have $(\boldsymbol{AX} + \widetilde{\boldsymbol{X}})^*(t) = (\boldsymbol{A}\Sigma_t\boldsymbol{A}^\top)^{-1/2}\{\boldsymbol{AX}(t) + \widetilde{\boldsymbol{X}}(t) - \boldsymbol{A}\boldsymbol{\mu_X}(t) - \widetilde{\boldsymbol{X}}(t)\} = (\boldsymbol{A}\Sigma_t\boldsymbol{A}^\top)^{-1/2}\boldsymbol{A}\{\boldsymbol{X}(t) - \boldsymbol{\mu_X}(t)\} = (\boldsymbol{A}\Sigma_t\boldsymbol{A}^\top)^{-1/2}\boldsymbol{A}\Sigma_t^{1/2}\Sigma_t^{-1/2}\{\boldsymbol{X}(t) - \boldsymbol{\mu_X}(t)\} = \boldsymbol{B}_t\boldsymbol{X}^*(t)$, where $\boldsymbol{B}_t = (\boldsymbol{A}\Sigma_t\boldsymbol{A}^\top)^{-1/2}\boldsymbol{A}\Sigma_t^{1/2}$.
Besides, $\boldsymbol{X}^*(t) \in \mathbb{R}^p$ for $t \in \mathcal{T}$ follow the same distribution of $\boldsymbol{S}$ satisfying $\mathrm{E}(\boldsymbol{S}) = \mathbf{0}$ and $\mathrm{var}(\boldsymbol{S}) = \mathbf{I}_p$ with cdf $F_{\boldsymbol{S}}$.

We can obtain $\mathrm{E}\{(\boldsymbol{AX} + \widetilde{\boldsymbol{X}})^*(t)\} = \mathbf{0}$, $\mathrm{var}\{(\boldsymbol{AX} + \widetilde{\boldsymbol{X}})^*(t)\} = \boldsymbol{B}_t\mathrm{var}\{\boldsymbol{X}^*(t)\}\boldsymbol{B}_t^\top = \boldsymbol{B}_t\boldsymbol{B}_t^\top = (\boldsymbol{A}\Sigma_t\boldsymbol{A}^\top)^{-1/2}\boldsymbol{A}\Sigma_t^{1/2}\Sigma_t^{1/2}\boldsymbol{A}^\top(\boldsymbol{A}\Sigma_t\boldsymbol{A}^\top)^{-1/2} = (\boldsymbol{A}\Sigma_t\boldsymbol{A}^\top)^{-1/2}\boldsymbol{A}\Sigma_t\boldsymbol{A}^\top(\boldsymbol{A}\Sigma_t\boldsymbol{A}^\top)^{-1/2} = \mathbf{I}_p$.
Hence, $(\boldsymbol{AX} + \widetilde{\boldsymbol{X}})^*(t)$ also follow the distribution of $\boldsymbol{S}$. From the affine invariance of D, we have $D((\boldsymbol{AX} + \widetilde{\boldsymbol{X}})^*(t); F_{(\boldsymbol{A\mathcal{Y}} + \widetilde{\boldsymbol{\mathcal{Y}}})^*(t)}) = D(\boldsymbol{X}^*(t); F_{\boldsymbol{\mathcal{Y}}^*(t)})$. Then $w_\beta(t; F_{(\boldsymbol{A\mathcal{Y}} + \widetilde{\boldsymbol{\mathcal{Y}}})^*}) = w(t; F_{\boldsymbol{\mathcal{Y}}^*})$, $\Psi_{(\boldsymbol{AX} + \widetilde{\boldsymbol{X}})^*}(d) = \Psi_{\boldsymbol{X}^*}(d)$ for $0 \leq d \leq 1$.

Therefore, $GMFD(\boldsymbol{X}; F_{\boldsymbol{\mathcal{Y}}}, \beta) = GMFD(\boldsymbol{AX} + \widetilde{\boldsymbol{X}}; F_{\boldsymbol{A\mathcal{Y}} + \widetilde{\boldsymbol{\mathcal{y}}}}, \beta)$.

2) For a center of symmetry $\boldsymbol{\Theta} \in \mathcal{C}^p(\mathcal{T})$ such that $P_{\boldsymbol{Y}-\boldsymbol{\Theta}} = P_{\boldsymbol{\Theta}-\boldsymbol{Y}}$ for a stochastic process $\boldsymbol{Y}$, it holds that for each $t \in \mathcal{T}$, $\boldsymbol{\Theta}(t)$ is a center of symmetry for $P_{\boldsymbol{Y}(t)}$. From definition 2.1 (ii) in Zuo & Serfling (2000a), $D(\boldsymbol{\Theta}(t); F_{\boldsymbol{\mathcal{Y}}(t)}) = \sup\limits_{\boldsymbol{X}(t) \in \mathbb{R}^p} D(\boldsymbol{X}(t); F_{\boldsymbol{\mathcal{Y}}(t)})$ for $t \in \mathcal{T}$. From the affine invariance of D, we have $D(\boldsymbol{\Theta}^*(t); F_{\boldsymbol{\mathcal{y}}^*(t)}) = \sup\limits_{\boldsymbol{X}^*(t) \in \mathbb{R}^p} D(\boldsymbol{X}^*(t); F_{\boldsymbol{\mathcal{y}}^*(t)})$ with $\boldsymbol{X}^*(t) = \boldsymbol{\Sigma}_t^{-1/2}\{\boldsymbol{X}(t) - \boldsymbol{\mu}(t)\}$, and $\boldsymbol{\Theta}^*(t) = \boldsymbol{\Sigma}_t^{-1/2}\{\boldsymbol{\Theta}(t) - \boldsymbol{\mu}(t)\}$. Hence, $\boldsymbol{\Theta}^*(t)$ is a center of symmetry for $P_{\boldsymbol{Y}^*(t)}$.

According to the definition of GMFID in Equation (1), since D has maximality at the center $\boldsymbol{\Theta}(t)$, GMFID has maximality at $\boldsymbol{\Theta}$.

According to the definition of GMFED in Equation (1), for any $0 < r < 1$, there exists $\mathcal{U}_r \subset \mathcal{T}$, $\forall t \in \mathcal{U}_r$, $D(\boldsymbol{\Theta}^*(t), F_{\boldsymbol{\mathcal{y}}^*(t)}) \leq r$; and $\forall t \in \mathcal{T} \setminus \mathcal{U}_r$, $D(\boldsymbol{\Theta}^*(t), F_{\boldsymbol{\mathcal{y}}^*(t)}) > r$. That is, $\Psi_{\boldsymbol{\Theta}^*}(r) = \frac{|\mathcal{U}_r|}{|\mathcal{T}|}$. If there exists $\boldsymbol{X}^*$ such that $\boldsymbol{X}^* \succ \boldsymbol{\Theta}^*$, then we have an $0 < r_0 < 1$ such that: when $0 < r < r_0$, we have $\Psi_{\boldsymbol{X}^*}(r) = \Psi_{\boldsymbol{\Theta}^*}(r)$, and $\Psi_{\boldsymbol{X}^*}(r_0) < \Psi_{\boldsymbol{\Theta}^*}(r_0)$. There exists $\mathcal{U}_{r_0} \subset \mathcal{T}$, for $t \in \mathcal{U}_{r_0}$, such that $D(\boldsymbol{\Theta}^*(t), F_{\boldsymbol{\mathcal{y}}^*(t)}) \leq r_0$, and when $t \in \mathcal{T} \setminus \mathcal{U}_{r_0}$, we have $D(\boldsymbol{\Theta}^*(t), F_{\boldsymbol{\mathcal{y}}^*(t)}) > r_0$. Hence, $\Psi_{\boldsymbol{\Theta}^*}(r_0) = \frac{|\mathcal{U}_{r_0}|}{|\mathcal{T}|}$.

When $t \in \mathcal{U}_{r_0}$, we have $D(\boldsymbol{X}^*(t); F_{\boldsymbol{\mathcal{y}}^*(t)}) \leq D(\boldsymbol{\Theta}^*(t), F_{\boldsymbol{\mathcal{y}}^*(t)}) \leq r_0$. Then, $\Psi_{\boldsymbol{X}^*}(r_0) \geq \Psi_{\boldsymbol{\Theta}^*}(r_0)$ conflicts with $\Psi_{\boldsymbol{X}^*}(r_0) < \Psi_{\boldsymbol{\Theta}^*}(r_0)$ when $\boldsymbol{X}^* \succ \boldsymbol{\Theta}^*$. According to the ranking criteria, $\forall \boldsymbol{X}^* \in \mathbb{R}^p$, $\boldsymbol{X}^* \preceq \boldsymbol{\Theta}^*$, and $\boldsymbol{\Theta}^*$ is the one with highest depth 1 in GMFED.

3) From the proof of (2) in Theorem **??** that $\boldsymbol{\Theta}^*(t)$ is the center of $F_{\boldsymbol{\mathcal{y}}^*(t)}$ and the monotonicity relative to the deepest point of D, we have $D(\boldsymbol{X}^*(t); F_{\boldsymbol{\mathcal{y}}^*(t)}) \leq D(\boldsymbol{\Theta}^*(t) + a\{\boldsymbol{X}^*(t) - \boldsymbol{\Theta}^*(t)\}; F_{\boldsymbol{\mathcal{y}}^*(t)})$ for $a \in [0, 1]$.

The weight $w_\beta(t; F_{\boldsymbol{\mathcal{y}}^*(t)})$ does not change, from the definition of Equation (1), hence $GMFID(\boldsymbol{X}; F_{\boldsymbol{\mathcal{y}}}, \beta) \leq GMFID(\boldsymbol{\Theta} + a(\boldsymbol{X} - \boldsymbol{\Theta}); F_{\boldsymbol{\mathcal{y}}}, \beta)$.

For any $0 < r < 1$, there exists a $\mathcal{U}_r \subset \mathcal{T}$, such that $\Psi_{\boldsymbol{\Theta}^*+a(\boldsymbol{X}^*-\boldsymbol{\Theta}^*)}(r) = \frac{|\mathcal{U}_r|}{|\mathcal{T}|}$, that is, $\forall t \in \mathcal{U}_r$, $D(\boldsymbol{\Theta}^*(t) + a\{\boldsymbol{X}^*(t) - \boldsymbol{\Theta}^*(t)\}; F_{\boldsymbol{\mathcal{y}}^*(t)}) \leq r$; $\forall t \in \mathcal{T} \setminus \mathcal{U}_r$, $D(\boldsymbol{\Theta}^*(t) + a\{\boldsymbol{X}^*(t) - \boldsymbol{\Theta}^*(t)\}; F_{\boldsymbol{\mathcal{y}}^*(t)}) > r$. If there exists $\boldsymbol{X}$ such that $GMFED(\boldsymbol{X}; F_{\boldsymbol{\mathcal{y}}}) > GMFED(\boldsymbol{\Theta} + a(\boldsymbol{X} - \boldsymbol{\Theta}); F_{\boldsymbol{\mathcal{y}}})$, there exists $\boldsymbol{X}^*$ such that $\boldsymbol{X}^* \succ \boldsymbol{\Theta}^* + a(\boldsymbol{X}^* - \boldsymbol{\Theta}^*)$, then we have an $0 < r_0 < 1$ such that: when $0 < r < r_0$, we have $\Psi_{\boldsymbol{X}^*}(r) = \Psi_{\boldsymbol{\Theta}^*}(r)$, and $\Psi_{\boldsymbol{X}^*}(r_0) < \Psi_{\boldsymbol{\Theta}^*+a(\boldsymbol{X}^*-\boldsymbol{\Theta}^*)}(r_0)$. There exists $\mathcal{U}_{r_0} \subset \mathcal{T}$, for

2

$t \in \mathcal{U}_{r_0}$, such that $D(\boldsymbol{\Theta}^*(t) + a(\boldsymbol{X}^*(t) - \boldsymbol{\Theta}^*(t)), F_{\boldsymbol{\mathcal{Y}}^*(t)}) \leq r_0$, and when $t \in \mathcal{T} \setminus \mathcal{U}_{r_0}$, we have

$D(\boldsymbol{\Theta}^*(t) + a(\boldsymbol{X}^* - \boldsymbol{\Theta}^*), F_{\boldsymbol{\mathcal{Y}}^*(t)}) > r_0$. Hence, $\Psi_{\boldsymbol{\Theta}^*}(r_0) = \frac{|\mathcal{U}_{r_0}|}{|\mathcal{T}|}$.

When $t \in \mathcal{U}_{r_0}$, we have $D(\boldsymbol{X}^*(t); F_{\boldsymbol{\mathcal{Y}}^*(t)}) \leq D(\boldsymbol{\Theta}^*(t) + a\{\boldsymbol{X}^*(t) - \boldsymbol{\Theta}^*(t)\}; F_{\boldsymbol{\mathcal{Y}}^*(t)}) \leq r_0$.

Thus $\Psi_{\boldsymbol{X}^*}(r_0) \geq \Psi_{\boldsymbol{\Theta}^*+a(\boldsymbol{X}^*-\boldsymbol{\Theta}^*)}(r_0)$. Hence, $\boldsymbol{X}^* \preceq \boldsymbol{\Theta}^* + a(\boldsymbol{X}^* - \boldsymbol{\Theta}^*)$ and conflicts with

$\boldsymbol{X}^* \succ \boldsymbol{\Theta}^* + a(\boldsymbol{X}^* - \boldsymbol{\Theta}^*)$. Hence, $GMFED(\boldsymbol{X}; F_{\boldsymbol{\mathcal{Y}}}) \leq GMFED(\boldsymbol{\Theta} + a(\boldsymbol{X} - \boldsymbol{\Theta}); F_{\boldsymbol{\mathcal{Y}}})$.

4) From the vanishing at infinity of D, for $t \in \mathcal{T}$ and $\boldsymbol{X}_n(t) \in \mathbb{R}^p$, if $\|\boldsymbol{X}_n(t)\| \to \infty$,

then $\lim_{n\to\infty} D(\boldsymbol{X}_n(t); F_{\boldsymbol{\mathcal{Y}}(t)}) = 0$. From the affine invariance of D, $\lim_{n\to\infty} D(\boldsymbol{X}_n(t); F_{\boldsymbol{\mathcal{Y}}(t)}) = $

$\lim_{n\to\infty} D(\boldsymbol{X}_n^*(t); F_{\boldsymbol{\mathcal{Y}}^*(t)}) = 0$.

Since we have $\lim_{n\to\infty} D(\boldsymbol{X}_n^*(t); F_{\boldsymbol{\mathcal{Y}}^*(t)}) = 0$ for almost all time points $t \in \mathcal{T}$, according to

Equation (1), $\lim_{n\to\infty} GMFID(\boldsymbol{X}_n; F_{\boldsymbol{\mathcal{Y}}}, \beta) = 0$.

Nex, $\forall C > 0$, $\exists n' > 0$, when $n > n'$ such that $D(\boldsymbol{X}_n^*(t), F_{\boldsymbol{\mathcal{Y}}^*(t)}) \leq C$ for almost all points

$t \in \mathcal{T}$. Let $C = \frac{1}{n}$. From the right continuity of Ψ, $\lim_{\frac{1}{n}\to 0^+} \Psi_{\boldsymbol{X}_n^*}(1/n) = \Psi_{\boldsymbol{X}^*}(0) = 1$, hence $\boldsymbol{X}_n^*$ is

the most extremal curve when $n \to \infty$. According to Equation (1), $\lim_{n\to\infty} GMFED(\boldsymbol{X}_n; F_{\boldsymbol{\mathcal{Y}}_n}) =$

$\lim_{n\to\infty} 1/n = 0$.

Proof of Theorem 2

1) We use Berge's maximum theorem (e.g. Abalo & Kostreva 2005, Theorem 1). Consider the

function $D : \mathcal{T} \times \mathbb{R}^p \to \mathbb{R} : (t, \boldsymbol{x}) \mapsto D(\boldsymbol{x}, F_{\boldsymbol{\mathcal{Y}}(t)})$. Due to the continuity of $\boldsymbol{x}(t)$ and D, the

continuity of $\boldsymbol{\Theta}$ follows.

2) For any $t \in \mathcal{T}$, there exists a point $\boldsymbol{\Theta}(t) \in \mathbb{R}^p$ with maximum depth for $F_{\boldsymbol{\mathcal{Y}}(t)}$ such that

$D(\boldsymbol{\Theta}(t), F_{\boldsymbol{\mathcal{Y}}(t)}) = \max_{\boldsymbol{X}(t)\in\mathbb{R}^p} D(\boldsymbol{X}(t), F_{\boldsymbol{\mathcal{Y}}(t)})$. From the affine invariance of D, the deepest point $\boldsymbol{\Theta}$

in $\boldsymbol{\mathcal{Y}}$ ensures the deepest point $\boldsymbol{\Theta}^*$ in $\boldsymbol{\mathcal{Y}}^*$. From Definition 1, we have $GMFID(\boldsymbol{X}; F_{\boldsymbol{\mathcal{Y}}}, \beta) \leq$

$GMFID(\boldsymbol{\Theta}; F_{\boldsymbol{\mathcal{Y}}}, \beta)$ for $\boldsymbol{X} \in \mathcal{C}^p(\mathcal{T})$.

Similar to the proof in 2) in Theorem **??**, if there exists $\boldsymbol{\Theta}$ with maximum depth at every time

point, then $GMFED(\boldsymbol{\Theta}; F_{\boldsymbol{\mathcal{Y}}}) \leq GMFED(\boldsymbol{X}; F_{\boldsymbol{\mathcal{Y}}})$ for $\boldsymbol{X} \in \mathcal{C}^p(\mathcal{T})$.

3) We prove by contradiction that a deepest curve implies deepest points at every $t \in \mathcal{T}$.

We denote by $\widetilde{\boldsymbol{\theta}}_t = \widetilde{\boldsymbol{\Theta}}(t)$ the deepest point of $P_{\boldsymbol{\mathcal{Y}}}$. Suppose that there exists a $t_1 \in \mathcal{T}$ with

3

$D(\boldsymbol{\Theta}(t_1); F_{\boldsymbol{\mathcal{Y}}(t_1)}) < D(\widetilde{\boldsymbol{\theta}}_{t_1}; F_{\boldsymbol{\mathcal{Y}}(t_1)})$. According to Lemma 2 in [Claeskens et al. (2014)](): $\exists \delta_0 > 0$ with for $|t t_1| < \delta : D(\boldsymbol{\Theta}(t); F_{\boldsymbol{\mathcal{Y}}(t)}) < D(\widetilde{\boldsymbol{\theta}}_{t_1}; F_{\boldsymbol{\mathcal{Y}}(t)}) < D(\widetilde{\boldsymbol{\theta}}_t; F_{\boldsymbol{\mathcal{Y}}(t)}) \leq \widetilde{r}$.

For GMFID, we have $GMFID(\boldsymbol{\Theta}; F_{\boldsymbol{\mathcal{Y}}}, \beta) = \int_{\mathcal{T}} D(\boldsymbol{\Theta}^*(t); F_{\boldsymbol{\mathcal{S}}}) w_\beta(t; F_{\boldsymbol{\mathcal{S}}}) \mathrm{d}t$

$< \int_{|t-t_1|>\delta_0} D(\boldsymbol{\Theta}^*(t); F_{\boldsymbol{\mathcal{S}}}) w_\beta(t; F_{\boldsymbol{\mathcal{S}}}) \mathrm{d}t + \int_{t_1-\delta_0}^{t_1+\delta_0} D(\widetilde{\boldsymbol{\theta}}(t); F_{\boldsymbol{\mathcal{S}}}) w_\beta(t; F_{\boldsymbol{\mathcal{S}}}) \mathrm{d}t$

$\leq GMFID(\widetilde{\boldsymbol{\Theta}}; F_{\boldsymbol{\mathcal{Y}}}, \beta)$, which is in contradiction with the fact that $\boldsymbol{\Theta}$ is a deepest curve.

For $GMFED$, we can easily get $\boldsymbol{\Psi}_{\widetilde{\boldsymbol{\Theta}}}(r) = \boldsymbol{\Psi}_{\boldsymbol{\Theta}}(r)$ for $0 \leq r \leq \inf_{|t-t_1|<\delta_0} D(\boldsymbol{\Theta}(t); F_{\boldsymbol{\mathcal{Y}}(t)})$, and $\boldsymbol{\Psi}_{\boldsymbol{\Theta}}(r) > \boldsymbol{\Psi}_{\widetilde{\boldsymbol{\Theta}}}(r)$ for $r > \inf_{|t-t_1|<\delta_0} D(\boldsymbol{\Theta}(t); F_{\boldsymbol{\mathcal{Y}}(t)})$. Hence, $\boldsymbol{\Theta} \prec \widetilde{\boldsymbol{\Theta}}$, which also conflicts with the fact that $\boldsymbol{\Theta}$ is the deepest curve.

Finite sample GMFID Calculation of Definition 3

In order to apply Definition 1 to the curve observations at a grid of time points, we use the linear interpolation at the unobserved time points and the middle time points between $[t_j^N, t_{j+1}^N]$ $(j = 1, \ldots, l_N)$ with the average of the function values. This yields a sample of continuous p-variate stochastic processes $\widetilde{\boldsymbol{\mathcal{Y}}}_i$ and $\widetilde{\boldsymbol{\mathcal{Y}}}_i^*$, $(i = 1, \ldots, N)$ on the interval $\widetilde{T}_N = [t_1^N, t_{l_N}^N]$ of which the lth $(l = 1, \ldots, p)$ component is defined by

$$
\widetilde{\mathcal{Y}}_i^{(l)}(t) = \begin{cases} Y_i^{(l)}(t_j^N) \frac{t_j^N + t_{j+1}^N - 2t}{t_{j+1}^N - t_j^N} \mathbf{1}_{T_i}(t_j^N) + \overline{Y^{(l)}}(t_j^N)\{1 - \frac{t_j^N + t_{j+1}^N - 2t}{t_{j+1}^N - t_j^N} \mathbf{1}_{T_i}(t_j^N)\} & t \in [t_j^N, (t_j^N + t_{j+1}^N)/2], \\ Y_i^{(l)}(t_{j+1}^N) \frac{2t - t_j^N - t_{j+1}^N}{t_{j+1}^N - t_j^N} \mathbf{1}_{T_i}(t_{j+1}^N) + \overline{Y^{(l)}}(t_j^N)\{1 - \frac{2t - t_j^N - t_{j+1}^N}{t_{j+1}^N - t_j^N} \mathbf{1}_{T_i}(t_{j+1}^N)\} & t \in [(t_j^N + t_{j+1}^N)/2, t_{j+1}^N], \end{cases}
$$

where $\mathbf{1}_A(x) := \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A, \end{cases}$ and $\overline{Y^{(l)}}(t_j^N) = \frac{\sum_{i=1}^N Y_i^{(l)}(t_j^N) \mathbf{1}_{T_i}(t_j^N)}{\sum_{i=1}^N \mathbf{1}_{T_i}(t_j^N)}$ for $j = 1, \ldots, l_N$. The empirical cumulative distribution function of this sample is denoted by $F_{\widetilde{\boldsymbol{\mathcal{Y}}}, N}$. Then

$$
\widetilde{\mathcal{Y}}_i^{(l)*}(t) = \begin{cases} \widehat{Y}_i^{(l)*}(t_j^N) \frac{t_j^N + t_{j+1}^N - 2t}{t_{j+1}^N - t_j^N} \mathbf{1}_{T_i}(t_j^N) + \overline{\widehat{Y}^{(l)*}}(t_j^N)\{1 - \frac{t_j^N + t_{j+1}^N - 2t}{t_{j+1}^N - t_j^N} \mathbf{1}_{T_i}(t_j^N)\} & t \in [t_j^N, (t_j^N + t_{j+1}^N)/2], \\ \widehat{Y}_i^{(l)*}(t_{j+1}^N) \frac{2t - t_j^N - t_{j+1}^N}{t_{j+1}^N - t_j^N} \mathbf{1}_{T_i}(t_{j+1}^N) + \overline{\widehat{Y}^{(l)*}}(t_j^N)\{1 - \frac{2t - t_j^N - t_{j+1}^N}{t_{j+1}^N - t_j^N} \mathbf{1}_{T_i}(t_{j+1}^N)\} & t \in [(t_j^N + t_{j+1}^N)/2, t_{j+1}^N], \end{cases}
$$

where $\widehat{\boldsymbol{Y}}^*(t_j^N) = (\widehat{Y}^{(1)*}(t_j^N), \ldots, \widehat{Y}^{(p)*}(t_j^N))^\top$, $\overline{\widehat{Y}^{(l)*}}(t_j^N) = \frac{\sum_{i=1}^N \widehat{Y}_i^{(l)*}(t_j^N) \mathbf{1}_{T_i}(t_j^N)}{\sum_{i=1}^N \mathbf{1}_{T_i}(t_j^N)}$ for $j = 1, \ldots, l_N$. The empirical cumulative distribution function of this sample is denoted by $F_{\widetilde{\boldsymbol{\mathcal{Y}}}^*, N}$.

Note the definition of the p-variate processes $\widetilde{\boldsymbol{\mathcal{Y}}}_i$ and the affine invariance of the depth function $D(\widetilde{\boldsymbol{X}}^*(t_j^N); F_{\widetilde{\boldsymbol{\mathcal{Y}}}^*(t_j^N), N}) = D(\widehat{\boldsymbol{X}}^*(t_j^N); F_{\widehat{\boldsymbol{\mathcal{Y}}}^*(t_j^N), N})$. In addition, for $\boldsymbol{X} \in \mathcal{C}^p(\mathcal{T})$ from the stochastic

4

process $\boldsymbol{\mathcal{Y}}$ with cdf $F_{\boldsymbol{\mathcal{Y}}}$, and $\boldsymbol{A} \in \mathbb{R}^{p\times p}$ with $\det(\boldsymbol{A}) \neq 0$ and $\widetilde{\boldsymbol{X}} \in \mathcal{C}^p(\mathcal{T})$, $\boldsymbol{AX} + \widetilde{\boldsymbol{X}}$ on $\mathcal{C}^p(\mathcal{T})$ is with cdf $F_{\boldsymbol{A\mathcal{Y}}+\widetilde{\boldsymbol{\mathcal{y}}}}$. For $t \in \mathcal{T}$, $\mathrm{vol}\{D_\beta(t, \boldsymbol{\mathcal{Y}}(t))\} = |\det(\boldsymbol{A})| \cdot \mathrm{vol}\{D_\beta(t, \boldsymbol{A\mathcal{Y}}(t) + \widetilde{\boldsymbol{\mathcal{Y}}}(t))\}$.

For $t \in [t_j^N, (t_j^N + t_{j+1}^N)/2]$, $\boldsymbol{A}_{N_j}(t) = \frac{t_j^N + t_{j+1}^N - 2t}{t_{j+1}^N - t_j^N}\boldsymbol{I}_p$; for $t \in [(t_j^N + t_{j+1}^N)/2, t_{j+1}^N]$, $\det\{\boldsymbol{A}_{N_j}(t)\} = -\frac{t_j^N + t_{j+1}^N - 2t}{t_{j+1}^N - t_j^N}\boldsymbol{I}_p$. Hence, for $t \in [t_j^N, t_{j+1}^N]$, $|\det\{\boldsymbol{A}_{N_j}(t)\}| = |\frac{t_j^N + t_{j+1}^N - 2t}{t_{j+1}^N - t_j^N}|^p$.

Indeed, when $w \propto$ the volume of the depth range,

$GMFID(\boldsymbol{X}; F_{\widetilde{\boldsymbol{\mathcal{y}}}_N}, \beta) = \sum_{j=1}^{l_N - 1} \int_{t_j^N}^{t_{j+1}^N} D(\widehat{\boldsymbol{X}}^*(t_j^N); F_{\widehat{\boldsymbol{\mathcal{y}}}^*(t_j^N),N}) w_\beta(t; F_{\widetilde{\boldsymbol{\mathcal{y}}}^*(t),N}) dt$

$= \sum_{j=1}^{l_N - 1} \{ D(\widehat{\boldsymbol{X}}^*(t_j^N); F_{\widehat{\boldsymbol{\mathcal{y}}}^*(t_j^N),N}) \frac{\mathrm{vol}\{D_\beta(t_j^N, F_{\widehat{\boldsymbol{\mathcal{y}}}^*(t_j^N),N})\}}{\int_{\widetilde{T}_N} \mathrm{vol}\{D_\beta(u, F_{\widetilde{\boldsymbol{\mathcal{y}}}^*(u),N})\} du} \int_{t_j^N}^{(t_j^N + t_{j+1}^N)/2} |\det\{\boldsymbol{A}_{N_j}(t)\}| dt +$

$D(\widehat{\boldsymbol{X}}^*(t_{j+1}^N); F_{\widehat{\boldsymbol{\mathcal{y}}}^*(t_{j+1}^N),N}) \frac{\mathrm{vol}\{D_\beta(t_{j+1}^N, F_{\widehat{\boldsymbol{\mathcal{y}}}^*(t_{j+1}^N),N})\}}{\int_{\widetilde{T}_N} \mathrm{vol}\{D_\beta(u, F_{\widetilde{\boldsymbol{\mathcal{y}}}^*(u),N})\} du} \int_{(t_j^N + t_{j+1}^N)/2}^{t_{j+1}^N} |\det\{\boldsymbol{A}_{N_j}(t)\}| dt \}$.

Knowing $\int_{t_j^N}^{(t_j^N + t_{j+1}^N)/2} |\det\{\boldsymbol{A}_{N_j}(t)\}| dt = \int_{(t_j^N + t_{j+1}^N)/2}^{t_{j+1}^N} |\det\{\boldsymbol{A}_{N_j}(t)\}| dt = \frac{t_{j+1}^N - t_j^N}{2(p+1)}$, and

$\int_{\widetilde{T}_N} \mathrm{vol}\{D_\beta(u, F_{\widetilde{\boldsymbol{\mathcal{y}}}^*(u),N})\} du = \sum_{j=1}^{l_N - 1} \{ \mathrm{vol}\{D_\beta(t_j^N, F_{\widehat{\boldsymbol{\mathcal{y}}}^*(t_j^N),N})\} \frac{t_{j+1}^N - t_j^N}{2(p+1)} +$

$\mathrm{vol}\{D_\beta(t_{j+1}^N, F_{\widehat{\boldsymbol{\mathcal{y}}}^*(t_{j+1}^N),N})\} \frac{t_{j+1}^N - t_j^N}{2(p+1)} \} = \sum_{j=1}^{l_N} \mathrm{vol}\{D_\beta(t_j^N, F_{\widehat{\boldsymbol{\mathcal{y}}}^*(t_j^N),N})\} \frac{t_{j+1}^N - t_{j-1}^N}{2(p+1)}$.

Hence, GMFID is transformed to

$GMFID(\boldsymbol{X}; F_{\widetilde{\boldsymbol{\mathcal{y}}}_N}, \beta) = \sum_{j=1}^{l_N} D(\widehat{\boldsymbol{X}}^*(t_j^N); F_{\widehat{\boldsymbol{\mathcal{y}}}^*(t_j^N),N}) \frac{\mathrm{vol}\{D_\beta(t_j^N, F_{\widehat{\boldsymbol{\mathcal{y}}}^*(t_j^N),N})\}(t_{j+1}^N t_{j-1}^N)}{\sum_{j=1}^{l_N} \mathrm{vol}\{D_\beta(t_j^N, F_{\widehat{\boldsymbol{\mathcal{y}}}^*(t_j^N),N})\}(t_{j+1}^N t_{j1}^N)}$.

Proof of Lemma 1

1) From the strong law of large numbers, the pointwise sample mean function $\overline{\boldsymbol{Y}}(t_{i,k})$ converges almost surely towards the pointwise mean $\mathrm{E}\{\boldsymbol{\mathcal{Y}}(t_{i,k})\}$, and the pointwise sample covariance $\boldsymbol{Q}_{t_{i,k}}$ converges in probability towards the variance $\boldsymbol{\Sigma}_{t_{i,k}}$, i.e., $\overline{\boldsymbol{Y}}(t_{i,k}) \xrightarrow{a.s.} \mathrm{E}\{\boldsymbol{\mathcal{Y}}(t_{i,k})\}$, and $\boldsymbol{Q}_{t_{i,k}} \xrightarrow{a.s.}$ $\boldsymbol{\Sigma}_{t_{i,k}}$.

When $\delta \to 0$, $\lim_{\delta \to 0} \overline{Y^{(l)}}(t_{i,k}) = \frac{\lim_{\delta \to 0} \sum_{g=1}^{N} \sum_{h=1}^{n_g} Y_{g,h}^{(l)} \mathbf{1}(t_{g,h} \in r_\delta(t_{i,k}))}{\lim_{\delta \to 0} \sum_{g=1}^{N} \sum_{h=1}^{n_g} \mathbf{1}(t_{g,h} \in r_\delta(t_{i,k}))} = \frac{\sum_{g=1}^{N} Y_g^{(l)}(t_{i,k}) m_g}{\sum_{g=1}^{N} m_g} = \frac{1}{M_N} \sum_{g=1}^{N} m_g Y_g^{(l)}(t_{i,k})$ with $m_g = \sum_{h=1}^{n_g} \mathbf{1}(t_{g,h} \in r_\delta(t_{i,k}))$, $M_N = \sum_{g=1}^{N} m_g$, and $0 \leq m_g \leq 1$.

Let $M_N \to \infty$ when $N \to \infty$. If $\sup_{N \geq 1} \frac{N m_N}{M_N} < \infty$, and $\sup_{N \geq 2} \sum_{i=1}^{N-1} \frac{i|m_{i+1} - m_i|}{M_N} < \infty$, and $\frac{1}{N} \sum_{g=1}^{N} Y_g^{(l)}(t_{i,k}) \xrightarrow{a.s.} \mathrm{E}\{\boldsymbol{\mathcal{Y}}^{(l)}(t_{i,k})\}$, according to Theorem 1 (Etemadi 2006), we have $\overline{Y^{(l)}}(t_{i,k}) \xrightarrow{a.s.} \mathrm{E}\{\boldsymbol{\mathcal{Y}}^{(l)}(t_{i,k})\}$ as $\delta \to 0$ and $N \to \infty$.

2) Similarly, we can prove $\overline{Y^{(l)} Y^{(m)}}(t_{i,k}) \xrightarrow{a.s.} \mathrm{E}\{\boldsymbol{\mathcal{Y}}^{(l)}(t_{i,k}) \boldsymbol{\mathcal{Y}}^{(m)}(t_{i,k})\}$. Hence, we have for any $1 \leq l, m \leq p$, $\overline{Y^{(l)} Y^{(m)}}(t_{i,k}) - \overline{Y^{(l)}}(t_{i,k}) \overline{Y^{(m)}}(t_{i,k}) \xrightarrow{a.s.} \mathrm{cov}\{Y^{(l)}(t_{i,k}), Y^{(m)}(t_{i,k})\}$ as $\delta \to 0$ and $N \to \infty$.

Proof of Theorem 3

1) From Definition 1,

$$
\begin{aligned}
\sup_{\boldsymbol{X}\in\mathcal{C}^p(\mathcal{T})} |GMFID_N(\boldsymbol{X};F_{\boldsymbol{\mathcal{Y}}_N},\beta) - GMFID(\boldsymbol{X};F_{\boldsymbol{\mathcal{Y}}},\beta)| &\leq \sup_{\boldsymbol{X}\in\mathcal{C}^p(\mathcal{T})} \{\int_{t\in\mathcal{T}} |D(\widehat{\boldsymbol{X}}^*(t);F_{\widetilde{\boldsymbol{\mathcal{Y}}}^*(t),N}) - D(\boldsymbol{X}^*(t);F_{\boldsymbol{\mathcal{Y}}^*(t)})|\mathrm{d}t \\
&+ \int_{t\in\mathcal{T}} |w_\beta(t;F_{\widetilde{\boldsymbol{\mathcal{Y}}}^*(t),N}) - w_\beta(t;F_{\boldsymbol{\mathcal{Y}}^*(t),N})|\mathrm{d}t\} \leq \sup_{\boldsymbol{X}\in\mathcal{C}^p(\mathcal{T})} \{\int_{t\in\mathcal{T}} |D(\widehat{\boldsymbol{X}}^*(t);F_{\widetilde{\boldsymbol{\mathcal{Y}}}^*(t),N}) - D(\boldsymbol{X}^*(t);F_{\widetilde{\boldsymbol{\mathcal{Y}}}^*(t),N})|\mathrm{d}t \\
&+ \int_{t\in\mathcal{T}} |D(\boldsymbol{X}^*(t);F_{\widetilde{\boldsymbol{\mathcal{Y}}}^*(t),N}) - D(\boldsymbol{X}^*(t);F_{\boldsymbol{\mathcal{Y}}^*(t)})|\mathrm{d}t + \int_{t\in\mathcal{T}} |w_\beta(t;F_{\widetilde{\boldsymbol{\mathcal{Y}}}^*(t),N}) - w_\beta(t;F_{\boldsymbol{\mathcal{Y}}^*(t)})|\mathrm{d}t\}.
\end{aligned}
$$

Then the proof is separated into three parts.

(i). From the strong law of large numbers, we have $\overline{\boldsymbol{Y}}(t_{i,k}) \xrightarrow{a.s.} \mathrm{E}\{\boldsymbol{\mathcal{Y}}(t_{i,k})\}$, and $\boldsymbol{Q}_{t_{i,k}} \xrightarrow{a.s.} \boldsymbol{\Sigma}_{t_{i,k}}$. The inverse and the power of a symmetric and positive definite matrix is continuous. From the continuous mapping theorem, $\widehat{\boldsymbol{\mathcal{Y}}}_N^*(t_{i,k}) \xrightarrow{a.s.} \boldsymbol{\mathcal{Y}}^*(t_{i,k})$.

(ii). On the space of curves in $\mathcal{C}^p(\mathcal{T})$ we define the uniform distance $\rho(\boldsymbol{X},\boldsymbol{Y}) = \|\boldsymbol{X} - \boldsymbol{Y}\|_\infty = \sup_{t\in\mathcal{T}} \|\boldsymbol{X}(t) - \boldsymbol{Y}(t)\|$, with $\|\cdot\|$ the Euclidean norm in $\mathbb{R}^p$. We first show that the function $\{\widetilde{\boldsymbol{\mathcal{Y}}}_i^* = (\widetilde{\mathcal{Y}}_i^{*(1)},\ldots,\widetilde{\mathcal{Y}}_i^{*(p)})^\top \in \mathbb{R}^p, i = 1,\ldots,N, t \in \mathcal{T}_N\} \xrightarrow{P} \boldsymbol{\mathcal{Y}}^*$. By Conditions (B.1)-(B.2), we have $|G(t_{i,j+1}) - G(t_{i,j})| = f(\epsilon_j)(t_{i,j+1} - t_{i,j})$, with ϵ_j in between $t_{i,j}$ and $t_{i,j+1}$. This is equivalent to $|t_{i,j+1} - t_{i,j}| = \frac{1}{f(\epsilon_j)(n_i-1)} \leq \frac{c}{n_i-1}$. For each $t \in \mathcal{T}_N$ there is precisely one interval $[t_{i,j}, t_{i,j+1})$ that contains t and since the interpolating process agrees with the observed curve on the time points $t_{i,1},\ldots,t_{i,n_i}$, it follows that for each curve $i = 1,\ldots,N$, $0 \leq \rho(\widetilde{\boldsymbol{\mathcal{Y}}}_i^*, \widehat{\boldsymbol{Y}}_i^*) \leq 2 \sup_{|s-t|\leq c/(n_i-1)} \|\widehat{\boldsymbol{Y}}_i^*(t) - \widehat{\boldsymbol{Y}}_i^*(s)\| \leq 2\omega_{\widehat{\boldsymbol{Y}}_n^*}(\frac{c}{n_i-1})$, with $\omega_{\widehat{\boldsymbol{Y}}^*}$ the modulus of continuity of $\widehat{\boldsymbol{Y}}^*$. Since $\mathcal{T}_N$ is compact and each function $\widehat{\boldsymbol{Y}}_i^*$ ($i = 1,\ldots,N$) is continuous, the function is also uniformly continuous with $\omega_{\widehat{\boldsymbol{Y}}_n^*}(\frac{c}{n_i-1}) \to 0$ when $n_i \to \infty$ and $N \to \infty$. Since the sample of curves is i.i.d. and by Theorems 3.2 and 7.5 in Billingsley (2013), we have the convergence of the random process $\widetilde{\boldsymbol{\mathcal{Y}}}_N^* \xrightarrow{P} \widehat{\boldsymbol{\mathcal{Y}}}_N^*$ as $\underline{T}_N \to \infty$ and $N \to \infty$. From Theorem 3.1 in Billingsley (2013), $\widetilde{\boldsymbol{\mathcal{Y}}}_N^* \xrightarrow{P} \boldsymbol{\mathcal{Y}}^*$.

(iii). Under the stated assumptions, Theorem 4.1 of Zuo & Serfling (2000b) yields the a.s. convergence of the β-trimmed regions $D_\beta(t, F_{\widetilde{\boldsymbol{\mathcal{Y}}}(t),N})$ to $D_\beta(t, F_{\boldsymbol{\mathcal{Y}}(t)})$, together with a nesting property such that for $0 < \epsilon < \beta$, $D_\beta(t, F_{\boldsymbol{\mathcal{Y}}(t)}) \subset D_\beta(t, F_{\widetilde{\boldsymbol{\mathcal{Y}}}(t),N}) \subset D_{\beta-\epsilon}(t, F_{\widetilde{\boldsymbol{\mathcal{Y}}}(t),N}) \subset D_{\beta\epsilon}(t, F_{\boldsymbol{\mathcal{Y}}(t)})$

for all $t \in \mathcal{T}$. By the dominated convergence, $\int_{t \in \mathcal{T}} |w_\beta(t; F_{\widetilde{\boldsymbol{y}}^*(t),N}) - w_\beta(t; F_{\boldsymbol{y}^*(t)})| \mathrm{d}t$ converges to 0 as $\underline{T}_N \to \infty$ and $N \to \infty$ when w is proportional to the volume range. When w is proportional to t, $\int_{t \in \mathcal{T}} |w_\beta(t; F_{\widetilde{\boldsymbol{y}}^*(t),N}) - w_\beta(t; F_{\boldsymbol{y}^*(t)})| \mathrm{d}t = \int_{t \in \mathcal{T}} |\widetilde{g}_{N,\delta}(t) - g(t)| \mathrm{d}t \leq \sup_{t \in \mathcal{T}} |\widetilde{G}_{n,\delta}(t) - G(t)| \xrightarrow{a.s.} 0$ from the Glivenko–Cantelli theorem.

2) The consistency of GMFED can be proven by the proof of consistency of ED in Narisetty & Nair (2016) under Conditions (A)-(E).

References

Abalo, K. & Kostreva, M. (2005). Berge equilibrium: some recent results from fixed-point theorems. *Applied Mathematics and Computation*, 169(1), 624–638.

Billingsley, P. (2013). *Convergence of Probability Measures*. John Wiley & Sons.

Claeskens, G., Hubert, M., Slaets, L., & Vakili, K. (2014). Multivariate functional halfspace depth. *Journal of the American Statistical Association*, 109(505), 411–423.

Etemadi, N. (2006). Convergence of weighted averages of random variables revisited. *Proceedings of the American Mathematical Society*, 134(9), 2739–2744.

Narisetty, N. N. & Nair, V. N. (2016). Extremal depth for functional data and applications. *Journal of the American Statistical Association*, 111(516), 1705–1714.

Zuo, Y. & Serfling, R. (2000a). General notions of statistical depth function. *Annals of Statistics*, 28(2), 461–482.

Zuo, Y. & Serfling, R. (2000b). Structural properties and convergence results for contours of sample statistical depth functions. *Annals of Statistics*, 28(2), 483–499.