SUPPLEMENT TO "UNIFORM CONVERGENCE OF LOCAL FRÉCHET REGRESSION, WITH APPLICATIONS TO LOCATING EXTREMA AND TIME WARPING FOR METRIC SPACE VALUED TRAJECTORIES": DETAILS ON THEORETICAL RESULTS

By Yaqing Chen and Hans-Georg Müller

S.1. Proofs of the uniform convergence of local Fréchet regression with fixed targets as in Section 2.

Proof of Theorem 1. For any $\varepsilon > 0$, taking $\beta_2' = \beta_2 + \varepsilon/2$, we first show (2.6).

We note that under (K0) and (R0), as $b \to 0$, it holds for $l = 0, 1, 2$ that

(S.1)
$$\rho_{l,b}(t) = \mathbb{E}\left[K_b(U-t)(U-t)^l\right] = b^l\left[f_U(t)\mathcal{K}_{1,l}(t,b) + bf_U'(t)\mathcal{K}_{1,l+1}(t,b) + O(b^2)\right],$$

where $\mathcal{K}_{k,l}(t,b) = \int_{\{(x-t)/b: x\in\mathcal{T}\}} K(x)^k x^l \mathrm{d}x$, for $k, l \in \mathbb{N}$, and the O terms are all uniform across $t \in \mathcal{T}$. These results are well known (Fan and Gijbels, 1996) and we therefore omit the proofs.

Let $\vartheta_m = (mb)^{-1/2}(-\log b)^{1/2}$. Under (K0) and (R0), it follows from (S.1) and similar arguments as given in the proof of Theorem B of Silverman (1978) that

(S.2)
$$\sup_{t\in\mathcal{T}}|\rho_{0,b}(t)| = O(1), \quad \sup_{t\in\mathcal{T}}|\rho_{1,b}^+(t)| = O(b), \quad \sup_{t\in\mathcal{T}}|\rho_{2,b}(t)| = O(b^2),$$
$$\sup_{t\in\mathcal{T}}|\widehat{\rho}_{l,m}(t) - \rho_{l,b}(t)| = O_P(\vartheta_m b^l),\ l = 0,1,2, \quad \sup_{t\in\mathcal{T}}|\widehat{\rho}_{1,m}^+(t)| = O(b) + O_P(\vartheta_m b),$$

where
$$\rho_{1,b}^+(t) = \mathbb{E}[K_b(U-t)|U-t|],$$

and
$$\widehat{\rho}_{1,m}^+(t) = m^{-1}\sum_{j=1}^m [K_b(U_j-t)|U_j-t|],$$

noting that $b^{-l}\widehat{\rho}_{l,m}(\cdot)$ and $b^{-1}\widehat{\rho}_{1,m}^+(\cdot)$ can all be viewed as kernel density estimators with kernels $K_l(x) = K(x)x^l$ and $K_1^+(x) = K(x)|x|$, respectively, as per Silverman (1978), and (S.2) implies $\sup_{t\in\mathcal{T}}|\widehat{\sigma}_m^2(t) - \sigma_b^2(t)| = O_P(\vartheta_m b^2)$.

Applying Taylor expansion yields

$$\sup_{t\in\mathcal{T}}\left|[\widehat{\sigma}^2_m(t)]^{-1}-[\sigma^2_b(t)]^{-1}\right|=O_P(\vartheta_m b^{-2}),$$

$$\sup_{t\in\mathcal{T}}\left|\frac{\widehat{\rho}_{2,m}(t)}{\widehat{\sigma}^2_m(t)}-\frac{\rho_{2,b}(t)}{\sigma^2_b(t)}\right|=O_P(\vartheta_m),\quad\text{and}\quad\sup_{t\in\mathcal{T}}\left|\frac{\widehat{\rho}_{1,m}(t)}{\widehat{\sigma}^2_m(t)}-\frac{\rho_{1,b}(t)}{\sigma^2_b(t)}\right|=O_P(\vartheta_m b^{-1}).$$

Hence,

(S.3)
$$\begin{aligned}&\sup_{t\in\mathcal{T}} m^{-1}\sum_{j=1}^m|\widehat{w}(U_j,t,b)-w(U_j,t,b)|\\&\quad\leqslant\sup_{t\in\mathcal{T}}|\widehat{\rho}_{0,m}(t)|\sup_{t\in\mathcal{T}}\left|\frac{\widehat{\rho}_{2,m}(t)}{\widehat{\sigma}^2_m(t)}-\frac{\rho_{2,b}(t)}{\sigma^2_b(t)}\right|+\sup_{t\in\mathcal{T}}|\widehat{\rho}^+_{1,m}(t)|\sup_{t\in\mathcal{T}}\left|\frac{\widehat{\rho}_{1,m}(t)}{\widehat{\sigma}^2_m(t)}-\frac{\rho_{1,b}(t)}{\sigma^2_b(t)}\right|\\&\quad=O_P(\vartheta_m).\end{aligned}$$

For any $R>0$, define a sequence of events

$$B_{R,m}=\left\{\sup_{t\in\mathcal{T}} m^{-1}\sum_{j=1}^m|\widehat{w}(U_j,t,b)-w(U_j,t,b)|\leqslant R\vartheta_m\right\}. \tag{S.4}$$

Given any $\upsilon>1$, set $a_m=\min\{(mb^2)^{\beta_2/[4(\beta_2-2+\upsilon)]},(mb^2(-\log b)^{-1})^{\beta_2/[4(\beta_2-1)]}\}$; for some $\eta\leqslant r_2$, set $\tilde{\eta}=\eta^{\beta_2/2}$, with r_2 and β_2 as per (R3). For any $\ell\in\mathbb{N}_+$, considering m large enough such that $\log_2(\tilde{\eta}a_m)>\ell$,

(S.5)
$$\begin{aligned}&P\left(a_m\sup_{t\in\mathcal{T}} d_{\mathcal{M}}\left(\widetilde{\nu}_b(t),\widehat{\nu}_m(t)\right)^{\beta_2/2}>2^\ell\right)\\&\leqslant P\left(\sup_{t\in\mathcal{T}} d_{\mathcal{M}}\left(\widetilde{\nu}_b(t),\widehat{\nu}_m(t)\right)>\eta/2\right)+P(B^c_{R,m})\\&\quad+\sum_{\substack{k>\ell\\2^k\leqslant\tilde{\eta}a_m}}P\left(\left\{2^{k-1}\leqslant a_m\sup_{t\in\mathcal{T}} d_{\mathcal{M}}\left(\widetilde{\nu}_b(t),\widehat{\nu}_m(t)\right)^{\beta_2/2}<2^k\right\}\cap B_{R,m}\right).\end{aligned}$$

Note that (S.3) implies $\lim_{R\to\infty}\limsup_{m\to\infty}P(B^c_{R,m})=0$. Regarding the first term on the right hand side of (S.5), we next show that

$$\sup_{t\in\mathcal{T}} d_{\mathcal{M}}\left(\widetilde{\nu}_b(t),\widehat{\nu}_m(t)\right)=o_P(1). \tag{S.6}$$

Given any $t\in\mathcal{T}$, $d_{\mathcal{M}}(\widetilde{\nu}_b(t),\widehat{\nu}_m(t))=o_P(1)$ follows from similar arguments as given in the proof of Lemma 2 in Section S.3 of the Supplementary Material of Petersen and Müller (2019). By Theorems 1.5.4, 1.5.7 and 1.3.6 of van der Vaart and Wellner (1996) and the total boundedness of $\mathcal{T}$, it suffices to show that for any $\epsilon > 0$, as $\delta \to 0$,

$$\limsup_{m\to\infty} P\left(\sup_{s,t\in\mathcal{T},\,|s-t|<\delta} \left|d_{\mathcal{M}}\left(\widetilde{\nu}_b(s),\widehat{\nu}_m(s)\right) - d_{\mathcal{M}}\left(\widetilde{\nu}_b(t),\widehat{\nu}_m(t)\right)\right| > 2\epsilon\right) \to 0.$$

In conjunction with (R1) and the fact that $|d_{\mathcal{M}}(\widetilde{\nu}_b(s),\widehat{\nu}_m(s)) - d_{\mathcal{M}}(\widetilde{\nu}_b(t),\widehat{\nu}_m(t))| \leqslant d_{\mathcal{M}}(\widetilde{\nu}_b(s),\widetilde{\nu}_b(t)) + d_{\mathcal{M}}(\widehat{\nu}_m(s),\widehat{\nu}_m(t))$, it suffices to show that

$$\text{(S.7)} \qquad \limsup_{b\to 0} \sup_{s,t\in\mathcal{T},\,|s-t|<\delta} \sup_{z\in\mathcal{M}} \left|\widetilde{L}_b(z,s) - \widetilde{L}_b(z,t)\right| \to 0, \quad \text{as } \delta \to 0,$$

and that for any $\epsilon > 0$,

(S.8)

$$\limsup_{m\to\infty} P\left(\sup_{s,t\in\mathcal{T},\,|s-t|<\delta} \sup_{z\in\mathcal{M}} \left|\widehat{L}_m(z,s) - \widehat{L}_m(z,t)\right| > \epsilon\right) \to 0, \quad \text{as } \delta \to 0.$$

Noting that by (K0) and (R0),

$$\mathbb{E}[w(U,t,b)] = 1 \quad \text{and} \quad \mathbb{E}[w(U,t,b)(U-t)] = 0,$$

(S.9)

$$\begin{aligned}
\widetilde{L}_b(z,t) &= \mathbb{E}\left[w(U,t,b)L(z,U)\right] \\
&= \mathbb{E}\left[w(U,t,b)\left(L(z,t) + (U-t)\frac{\partial L}{\partial t}(z,t)\right)\right] \\
&\quad + \mathbb{E}\left[w(U,t,b)\left(L(z,U) - L(z,t) - (U-t)\frac{\partial L}{\partial t}(z,t)\right)\right] \\
&= L(z,t) + \mathbb{E}\left[w(U,t,b)\left(L(z,U) - L(z,t) - (U-t)\frac{\partial L}{\partial t}(z,t)\right)\right].
\end{aligned}$$

Defining a function $\phi\colon \mathcal{M}\times\mathcal{T} \to \mathbb{R}$ as

$$\text{(S.10)} \qquad \phi(z,t) = \frac{\partial^2 L}{\partial t^2}(z,t), \quad z \in \mathcal{M},\ t \in \mathcal{T},$$

it follows from (S.9), (K0), and (R0) that

$$\sup_{z\in\mathcal{M},\,t\in\mathcal{T}} \left|\widetilde{L}_b(z,t) - L(z,t)\right| \leqslant \frac{1}{2} b^2 \sup_{t\in\mathcal{T}} \mathbb{E}\left[|w(U,t,b)|\left(\frac{U-t}{b}\right)^2\right] \sup_{t\in\mathcal{T},\,z\in\mathcal{M}} |\phi(z,t)|.$$

Note that $\sup_{t\in\mathcal{T}} \mathbb{E}[|w(U,t,b)|(U-t)^2 b^{-2}] = O(1)$. Furthermore, using similar arguments as given in the proof of Theorem 3 of Petersen and Müller (2019), we obtain

$$L(z,t) = \int d_{\mathcal{M}}^2(z',z)\mathrm{d}F_{V|U}(t,z') = \int d_{\mathcal{M}}^2(z',z)\frac{f_{U|V}(t,z')}{f_U(t)}\mathrm{d}F_V(z'), \tag{S.11}$$

where F_V and $F_{V|U}$ are the marginal and conditional distribution of V, the latter given U. In conjunction with (K0), (R0), and the dominated convergence theorem, (S.11) implies

$$\phi(z,t) = \int d_{\mathcal{M}}^2(z',z)\frac{\partial^2}{\partial t^2}\left[\frac{f_{U|V}(t,z')}{f_U(t)}\right]\mathrm{d}F_V(z'), \tag{S.12}$$

whence by (R0) and the boundedness of $\mathcal{M}$, we obtain $\sup_{z\in\mathcal{M},\, t\in\mathcal{T}} |\phi(z,t)| < \infty$. Thus,

$$\sup_{z\in\mathcal{M},\, t\in\mathcal{T}} \left|\widetilde{L}_b(z,t) - L(z,t)\right| = O(b^2), \quad \text{as } b \to 0. \tag{S.13}$$

Moreover, by (S.11) and (R0),

$$\begin{aligned} &\sup_{s,t\in\mathcal{T},\,|s-t|<\delta}\sup_{z\in\mathcal{M}} \left|\widetilde{L}_b(z,s) - \widetilde{L}_b(z,t)\right| \\ &\quad\leqslant \sup_{s,t\in\mathcal{T},\,|s-t|<\delta}\sup_{z\in\mathcal{M}} |L(z,s) - L(z,t)| + 2\sup_{z\in\mathcal{M},\,t\in\mathcal{T}}\left|\widetilde{L}_b(z,t) - L(z,t)\right| \\ &\quad= O(\delta) + O(b^2), \end{aligned} \tag{S.14}$$

whence (S.7) follows. For (S.8), let $\psi\colon \mathcal{M} \to \mathbb{R}$ be a function defined as

$$\psi(z) := \sup_{t\in\mathcal{T}}\left|m^{-1}\sum_{j=1}^m \left[w(U_j,t,b)d_{\mathcal{M}}^2(V_j,z)\right] - \mathbb{E}\left[w(U,t,b)d_{\mathcal{M}}^2(V,z)\right]\right|.$$

Then

$$\begin{aligned} \sup_{z\in\mathcal{M}}\left|\widehat{L}_m(z,s) - \widehat{L}_m(z,t)\right| &\leqslant 2\mathrm{diam}(\mathcal{M})^2\sup_{t\in\mathcal{T}} m^{-1}\sum_{j=1}^m |\widehat{w}(U_j,t,b) - w(U_j,t,b)| \\ &\quad + \sup_{z\in\mathcal{M}}\left|\widetilde{L}_b(z,s) - \widetilde{L}_b(z,t)\right| + 2\sup_{z\in\mathcal{M}}\psi(z). \end{aligned} \tag{S.15}$$

Regarding $\psi(z)$, since

$$
\begin{aligned}
& m^{-1} \sum_{j=1}^{m} \left[w(U_j,t,b) d^2_{\mathcal{M}}(V_j,z)\right] - \mathbb{E}\left[w(U,t,b) d^2_{\mathcal{M}}(V,z)\right] \\
&= \frac{\rho_{2,b}(t)}{\sigma_b^2(t)} \left\{ m^{-1} \sum_{j=1}^{m} \left[K_b(U_j-t) d^2_{\mathcal{M}}(V_j,z)\right] - \mathbb{E}\left[K_b(U-t) d^2_{\mathcal{M}}(V,z)\right] \right\} \\
&\quad - \frac{b\rho_{1,b}(t)}{\sigma_b^2(t)} \left\{ m^{-1} \sum_{j=1}^{m} \left[K_b(U_j-t)\left(\frac{U_j-t}{b}\right) d^2_{\mathcal{M}}(V_j,z)\right] \right. \\
&\qquad\qquad \left. - \mathbb{E}\left[K_b(U-t)\left(\frac{U-t}{b}\right) d^2_{\mathcal{M}}(V,z)\right] \right\},
\end{aligned}
$$

under (K0) and (R0), it follows from similar arguments as given in the proof of Proposition 4 of Mack and Silverman (1982) with kernels $K_l(x) = K(x)x^l$, for $l = 0, 1$, that $\psi(z) = o_P(1)$, for any given $z \in \mathcal{M}$. Furthermore, noting that

$$
\sup_{t\in\mathcal{T}} m^{-1} \sum_{j=1}^{m} |w(U_j,t,b)| \leqslant \sup_{t\in\mathcal{T}} |\widehat{\rho}_{0,m}(t)| \sup_{t\in\mathcal{T}} \left|\frac{\rho_{2,b}(t)}{\sigma_b^2(t)}\right| + \sup_{t\in\mathcal{T}} |\widehat{\rho}^{+}_{1,m}(t)| \sup_{t\in\mathcal{T}} \left|\frac{\rho_{1,b}(t)}{\sigma_b^2(t)}\right|,
$$

$$
\sup_{t\in\mathcal{T}} \mathbb{E}\,|w(U,t,b)| \leqslant \sup_{t\in\mathcal{T}} \frac{\rho_{0,b}(t)\rho_{2,b}(t)}{\sigma_b^2(t)} + \sup_{t\in\mathcal{T}} \frac{\rho^{+}_{1,b}(t)|\rho_{1,b}(t)|}{\sigma_b^2(t)},
$$

by (S.1) and (S.2),

$$
\sup_{t\in\mathcal{T}} m^{-1} \sum_{j=1}^{m} |w(U_j,t,b)| = O(1) + O_P(\vartheta_m), \quad \sup_{t\in\mathcal{T}} \mathbb{E}\,|w(U,t,b)| = O(1),
$$

and hence for any $z_1, z_2 \in \mathcal{M}$,

$$
\begin{aligned}
|\psi(z_1) - \psi(z_2)| &\leqslant 2\mathrm{diam}(\mathcal{M}) d_{\mathcal{M}}(z_1,z_2) \left(\sup_{t\in\mathcal{T}} m^{-1} \sum_{j=1}^{m} |w(U_j,t,b)| + \sup_{t\in\mathcal{T}} \mathbb{E}\,|w(U,t,b)| \right) \\
&= d_{\mathcal{M}}(z_1,z_2) \left[O(1) + O_P(\vartheta_m)\right].
\end{aligned}
$$

In conjunction with the total boundedness of $\mathcal{M}$, $\sup_{z\in\mathcal{M}} \psi(z) = o_P(1)$ follows (using Theorems 1.5.4, 1.5.7, and 1.3.6 of van der Vaart and Wellner, 1996). This implies (S.8) in conjunction with (S.15), (S.3) and (S.14). Thus, (S.6) follows.

We move on to the third term on the right hand side of (S.5). For $k \in \mathbb{N}_+$, define sets

$$
\mathcal{D}_{k,t} = \{z \in \mathcal{M} : 2^{k-1} \leqslant a_m d_{\mathcal{M}}(z, \widetilde{\nu}_b(t))^{\beta_2/2} < 2^k\}. \tag{S.16}
$$

We note that under (R3),

$$\liminf_{m\to\infty}\inf_{t\in\mathcal{T}}\inf_{z\in\mathcal{D}_{k,t}}\left[\widetilde{L}_b(z,t)-\widetilde{L}_b(\widetilde{\nu}_b(t),t)\right]\geqslant c_2 2^{2(k-1)}a_m^{-2}.$$

Defining functions $J_t(\cdot)=\widehat{L}_m(\cdot,t)-\widetilde{L}_b(\cdot,t)$ on $\mathcal{M}$, applying Markov's inequality, the third term on the right hand side of (S.5) can be bounded (from above) by

(S.17)
$$\begin{aligned}&\sum_{\substack{k>\ell\\ 2^k\leqslant\widetilde{\eta}a_m}} P\left(\left\{\sup_{t\in\mathcal{T}}\sup_{z\in\mathcal{D}_{k,t}}|J_t(z)-J_t(\widetilde{\nu}_b(t))|\geqslant c_2 2^{2(k-1)}a_m^{-2}\right\}\cap B_{R,m}\right)\\ &\quad\leqslant\sum_{\substack{k>\ell\\ 2^k\leqslant\widetilde{\eta}a_m}} c_2^{-1}2^{-2(k-1)}a_m^2\mathbb{E}\left(\mathbb{I}\left(B_{R,m}\right)\sup_{t\in\mathcal{T}}\sup_{z\in\mathcal{D}_{k,t}}|J_t(z)-J_t(\widetilde{\nu}_b(t))|\right),\end{aligned}$$

where $\mathbb{I}(E)$ is the indicator function of an event E. For any $z\in\mathcal{M}$, defining

(S.18)
$$\begin{aligned}J_t^{(1)}(z)&=m^{-1}\sum_{j=1}^m\left[\widehat{w}(U_j,t,b)-w(U_j,t,b)\right]d_{\mathcal{M}}^2(V_j,z),\\ J_t^{(2)}(z)&=m^{-1}\sum_{j=1}^m\left[w(U_j,t,b)d_{\mathcal{M}}^2(V_j,z)\right]-\mathbb{E}\left[w(U,t,b)d_{\mathcal{M}}^2(V,z)\right],\end{aligned}$$

we have $J_t=J_t^{(1)}+J_t^{(2)}$.

For $J_t^{(1)}$, note that

$$\left|J_t^{(1)}(z)-J_t^{(1)}(\widetilde{\nu}_b(t))\right|\leqslant 2\mathrm{diam}(\mathcal{M})d_{\mathcal{M}}(z,\widetilde{\nu}_b(t))m^{-1}\sum_{j=1}^m|\widehat{w}(U_j,t,b)-w(U_j,t,b)|,$$

for all $z\in\mathcal{M}$ and $t\in\mathcal{T}$. Hence, given $\delta>0$, it holds on $B_{R,m}$ that

(S.19)
$$\sup_{t\in\mathcal{T}}\sup_{d_{\mathcal{M}}(z,\widetilde{\nu}_b(t))<\delta}\left|J_t^{(1)}(z)-J_t^{(1)}(\widetilde{\nu}_b(t))\right|\leqslant 2\mathrm{diam}(\mathcal{M})R\delta\vartheta_m.$$

For $J_t^{(2)}$, given any $z\in\mathcal{M}$, $t\in\mathcal{T}$ and $\delta>0$, define functions $g_{t,z}\colon\mathcal{T}\times\mathcal{M}\to\mathbb{R}$ by

$$g_{t,z}(s,z')=w(s,t,b)\left[d_{\mathcal{M}}^2(z',z)-d_{\mathcal{M}}^2(z',\widetilde{\nu}_b(t))\right],\quad s\in\mathcal{T},z'\in\mathcal{M},$$

and a function class

$$\mathcal{G}_{b,\delta}=\left\{g_{t,z}:d_{\mathcal{M}}(z,\widetilde{\nu}_b(t))<\delta,\,t\in\mathcal{T}\right\}.$$

For any $t_1, t_2 \in \mathcal{T}$ and $z_l \in B_\delta(\widetilde{\nu}_b(t_l))$ for $l = 1, 2$,

$$\begin{aligned}&\left|g_{t_1,z_1}(s,z') - g_{t_2,z_2}(s,z')\right| \\ &\quad \leqslant 2\mathrm{diam}(\mathcal{M})\left(d_{\mathcal{M}}(z_1,z_2) + d_{\mathcal{M}}(\widetilde{\nu}_b(t_1),\widetilde{\nu}_b(t_2))\right)\sup_{t\in\mathcal{T}}|w(s,t,b)| \\ &\qquad + 2\mathrm{diam}(\mathcal{M})\delta\sup_{s\in\mathcal{T}}\left|w(s,t_1,b) - w(s,t_2,b)\right|.\end{aligned}$$

By (S.7) and (R1),

$$\limsup_{b\to 0}\sup_{s,t\in\mathcal{T},\,|s-t|<\delta} d_{\mathcal{M}}(\widetilde{\nu}_b(s),\widetilde{\nu}_b(t)) \to 0, \quad \text{as } \delta \to 0. \tag{S.20}$$

For small $|t_1 - t_2|$ such that $d_{\mathcal{M}}(\widetilde{\nu}_b(t_1),\widetilde{\nu}_b(t_2)) < r_2$, by (R3), it holds that as $b \to 0$,

$$\begin{aligned}2c_2 d_{\mathcal{M}}(\widetilde{\nu}_b(t_1),\widetilde{\nu}_b(t_2))^{\beta_2} &\leqslant 2\sup_{z\in\mathcal{M}}\left|\widetilde{L}_b(z,t_1) - \widetilde{L}_b(z,t_2)\right| \\ &\leqslant 2\mathrm{diam}(\mathcal{M})^2\sup_{s\in\mathcal{T}}|w(s,t_1,b) - w(s,t_2,b)|.\end{aligned}$$

Noting that

$$\sup_{s\in\mathcal{T}}|w(s,t_1,b) - w(s,t_2,b)| = |t_1 - t_2| O(b^{-2})$$

by (K0) and (R0), there exists a constant $C > 0$ such that for small $|t_1 - t_2|$,

$$\left|g_{t_1,z_1}(s,z') - g_{t_2,z_2}(s,z')\right| \leqslant C\left[d_{\mathcal{M}}(z_1,z_2) + d_{\mathcal{T}}(t_1,t_2)\right] D_{b,\delta}(s),$$

where $d_{\mathcal{T}} \colon \mathcal{T} \times \mathcal{T} \to [0,+\infty)$ is a metric on $\mathcal{T}$ defined as

$$d_{\mathcal{T}}(t_1,t_2) = \max\left\{b^{-2}|t_1 - t_2|, (b^{-2}|t_1 - t_2|)^{1/\beta_2}\right\}, \quad t_1, t_2 \in \mathcal{T},$$

with a function

$$D_{b,\delta}(s,z) = \sup_{t\in\mathcal{T}}|w(s,t,b)| + \delta, \quad s \in \mathcal{T}.$$

It is not difficult to verify that $d_{\mathcal{T}}$ is indeed a metric.

An envelope function $G_{b,\delta} \colon \mathcal{T} \times \mathcal{M} \to \mathbb{R}$ for the function class $\mathcal{G}_{b,\delta}$ is

$$G_{b,\delta}(s,z) = 2\mathrm{diam}(\mathcal{M})\delta\sup_{t\in\mathcal{T}}|w(s,t,b)|, \quad s \in \mathcal{T}.$$

Denoting the joint distribution of (U, V) by $\mathcal{F}$, the $\mathcal{L}^2_{\mathcal{F}}$ norm $\|\cdot\|_{\mathcal{F}}$ is given by $\|g\|_{\mathcal{F}} = [\mathbb{E}(g(U,V)^2)]^{1/2}$, for any function $g \colon \mathcal{T} \times \mathcal{M} \to \mathbb{R}$. The envelope function entails $\|G_{b,\delta}\|_{\mathcal{F}} = O(\delta b^{-1})$, by (K0) and (R0). Furthermore, by Theorem 2.7.11 of van der Vaart and Wellner (1996), for $\epsilon > 0$, the $\epsilon\|G_{b,\delta}\|_{\mathcal{F}}$ bracketing number of the function class $\mathcal{G}_{b,\delta}$ can be bounded as follows,

$$\begin{aligned} & N_{[]}\left(\epsilon\|G_{b,\delta}\|_{\mathcal{F}}, \mathcal{G}_{b,\delta}, \|\cdot\|_{\mathcal{F}}\right) \\ & = N_{[]}\left(2\frac{\epsilon\|G_{b,\delta}\|_{\mathcal{F}}}{2\|D_{b,\delta}\|_{\mathcal{F}}}\|D_{b,\delta}\|_{\mathcal{F}}, \mathcal{G}_{b,\delta}, \|\cdot\|_{\mathcal{F}}\right) \\ & \leqslant N\left(\frac{\epsilon\|G_{b,\delta}\|_{\mathcal{F}}}{2\|D_{b,\delta}\|_{\mathcal{F}}}, \{(t,z) : z \in B_\delta(\tilde{\nu}_b(t)),\, t \in \mathcal{T}\}, d_{\mathcal{T}\times\mathcal{M}}\right) \\ & \leqslant N\left(\frac{\epsilon\|G_{b,\delta}\|_{\mathcal{F}}}{4\|D_{b,\delta}\|_{\mathcal{F}}}, \mathcal{T}, d_{\mathcal{T}}\right) \cdot \sup_{t\in\mathcal{T}} N\left(\frac{\epsilon\|G_{b,\delta}\|_{\mathcal{F}}}{4\|D_{b,\delta}\|_{\mathcal{F}}}, B_\delta(\tilde{\nu}_b(t)), d_{\mathcal{M}}\right), \end{aligned}$$

where $d_{\mathcal{T}\times\mathcal{M}}((t_1, z_1), (t_2, z_2)) = d_{\mathcal{T}}(t_1, t_2) + d_{\mathcal{M}}(z_1, z_2)$, for any $t_1, t_2 \in \mathcal{T}$ and $z_1, z_2 \in \mathcal{M}$. Therefore,

(S.21)
$$N_{[]}\left(\epsilon\|G_{b,\delta}\|_{\mathcal{F}}, \mathcal{G}_{b,\delta}, \|\cdot\|_{\mathcal{F}}\right) \leqslant C_1(\epsilon\delta b^2)^{-C_0} \sup_{t\in\mathcal{T}} N\left(C_2\epsilon\delta, B_\delta(\tilde{\nu}_b(t)), d_{\mathcal{M}}\right),$$

where $C_0, C_1, C_2 > 0$ are constants only depending on β_2, noting that $\|G_{b,\delta}\|_{\mathcal{F}}/\|D_{b,\delta}\|_{\mathcal{F}} \sim \delta$. In conjunction with (2.5), to be shown later, for b sufficiently small, there exists a constant $C_3 > C_2$ such that $B_\delta(\tilde{\nu}_b(t)) \subset B_{C_3\delta}(\nu(t))$, for any $\delta > 0$ and $t \in \mathcal{T}$. Choose η in (S.5) such that (R2) holds for all $r \leqslant C_3\eta$. Observing that

$$\begin{aligned} & \int_0^1 \sup_{t\in\mathcal{T}} \sqrt{1 + \log N\left(C_2\epsilon\delta, B_{C_3\delta}(\nu(t)), d_{\mathcal{M}}\right)}\mathrm{d}\epsilon \\ & = \frac{C_3}{C_2}\int_0^{C_2/C_3} \sup_{t\in\mathcal{T}} \sqrt{1 + \log N\left(\epsilon C_3\delta, B_{C_3\delta}(\nu(t)), d_{\mathcal{M}}\right)}\mathrm{d}\epsilon \\ & \leqslant \frac{C_3}{C_2}\int_0^1 \sup_{t\in\mathcal{T}} \sqrt{1 + \log N\left(\epsilon C_3\delta, B_{C_3\delta}(\nu(t)), d_{\mathcal{M}}\right)}\mathrm{d}\epsilon, \end{aligned}$$

(S.21) implies for any $\delta \leqslant \eta$,

$$\begin{aligned} & \int_0^1 \sqrt{1 + \log N_{[]}\left(\epsilon\|G_{b,\delta}\|_{\mathcal{F}}, \mathcal{G}_{b,\delta}, \|\cdot\|_{\mathcal{F}}\right)}\mathrm{d}\epsilon \\ & \leqslant \int_0^1 \sup_{t\in\mathcal{T}} \sqrt{1 + \log N\left(C_2\epsilon\delta, B_\delta(\tilde{\nu}_b(t)), d_{\mathcal{M}}\right)}\mathrm{d}\epsilon + \int_0^1 \sqrt{-C_0\log(\epsilon\delta b^2) + \log C_1}\mathrm{d}\epsilon \\ & = O\left(\mathcal{I} + \int_0^1 \sqrt{-\log(\delta b) - \log\epsilon}\,\mathrm{d}\epsilon\right) = O\left(\sqrt{-\log(\delta b)}\right), \end{aligned}$$

with $\mathcal{I}$ being the integral in (R2). By Theorem 2.14.2 of van der Vaart and Wellner (1996),

(S.22)
$$\begin{aligned}\mathbb{E}\left(\sup_{t\in\mathcal{T}}\sup_{d_{\mathcal{M}}(z,\tilde{\nu}_b(t))<\delta}\left|J_t^{(2)}(z)-J_t^{(2)}(\tilde{\nu}_b(t))\right|\right) &= O\left(\delta b^{-1}\sqrt{-\log(\delta b)}m^{-1/2}\right)\\ &= O\left(\delta^{2-\upsilon}(mb^2)^{-1/2}+\delta(mb^2)^{-1/2}\sqrt{-\log b}\right).\end{aligned}$$

Combining (S.19) and (S.22), it holds that

$$\begin{aligned}\mathbb{E}\left(\mathbb{I}\left(B_{R,m}\right)\sup_{t\in\mathcal{T}}\sup_{d_{\mathcal{M}}(z,\tilde{\nu}_b(t))<\delta}\left|J_t(z)-J_t(\tilde{\nu}_b(t))\right|\right)\\ \leqslant C(mb^2)^{-1/2}\left(\delta^{2-\upsilon}+\delta\sqrt{-\log b}\right),\end{aligned} \tag{S.23}$$

where $C>0$ is a constant depending on R and the entropy integral in (R2). Note that on $\mathcal{D}_{k,t}$, it holds that $d_{\mathcal{M}}(z,\tilde{\nu}_b(t))<(2^k a_m^{-1})^{2/\beta_2}$. Hence, (S.17) can be bounded by

$$\begin{aligned}&C\sum_{\substack{k>\ell\\ 2^k\leqslant\tilde{\eta}a_m}}2^{-2(k-1)}a_m^2(mb^2)^{-1/2}\left[(2^k a_m^{-1})^{2(2-\upsilon)/\beta_2}+(2^k a_m^{-1})^{2/\beta_2}\sqrt{-\log b}\right]\\ &\quad\leqslant 4Ca_m^{2(\beta_2-2+\upsilon)/\beta_2}(mb^2)^{-1/2}\sum_{k>\ell}2^{-2k(\beta_2-2+\upsilon)/\beta_2}\\ &\quad\quad+4Ca_m^{2(\beta_2-1)/\beta_2}(mb^2)^{-1/2}\sqrt{-\log b}\sum_{k>\ell}2^{-2k(\beta_2-1)/\beta_2},\end{aligned}$$

which converges to 0 as $\ell\to\infty$, since $\beta_2,\upsilon>1$. Thus,

$$\sup_{t\in\mathcal{T}}d_{\mathcal{M}}\left(\tilde{\nu}_b(t),\hat{\nu}_m(t)\right)=O_P\left(a_m^{-2/\beta_2}\right),$$

and (2.6) follows.

Next we establish (2.5). By (S.13) and (R1), $d_{\mathcal{M}}(\nu(t),\tilde{\nu}_b(t))=o(1)$, as $b\to 0$, for any $t\in\mathcal{T}$. By (2.4) and the compactness of $\mathcal{T}$, the conditional Fréchet mean trajectory ν is $d_{\mathcal{M}}$-continuous at any $t\in\mathcal{T}$ and hence uniformly $d_{\mathcal{M}}$-continuous on $\mathcal{T}$. In conjunction with (S.20), $\sup_{t\in\mathcal{T}}d_{\mathcal{M}}(\nu(t),\tilde{\nu}_b(t))=o(1)$ follows. Let $g_{z,t}\colon\mathcal{T}\to\mathbb{R}$ be a function defined as $g_{z,t}(s)=L(z,s)-L(\nu(t),s)$, for $s\in\mathcal{T}$. For any $\delta>0$, (S.9) and (S.12), in conjunction with (K0), (R0), and the boundedness of $\mathcal{M}$,

$$\begin{aligned}
&\sup_{t\in\mathcal{T}} \sup_{d_{\mathcal{M}}(z,\nu(t))<\delta} \left|(\widetilde{L}_b - L)(z,t) - (\widetilde{L}_b - L)(\nu(t),t)\right| \\
&\quad = \sup_{t\in\mathcal{T}} \sup_{d_{\mathcal{M}}(z,\nu(t))<\delta} \left|\mathbb{E}\left[w(U,t,b)\left(g_{z,t}(U) - g_{z,t}(t) - (U-t)g'_{z,t}(t)\right)\right]\right| \\
&\quad \leqslant \frac{1}{2}b^2 \sup_{t\in\mathcal{T}} \mathbb{E}\left[|w(U,t,b)|\left(\frac{U-t}{b}\right)^2\right] \cdot \sup_{t\in\mathcal{T}} \sup_{d_{\mathcal{M}}(z,\nu(t))<\delta} |\phi(z,t) - \phi(\nu(t),t)| \\
&\quad = O(b^2\delta),
\end{aligned}$$

with $\phi(z,t)$ defined as per (S.10). Set $q_b = b^{-\beta_1/(\beta_1-1)}$. Using similar arguments as given in the proof of (2.6), there exists a constant $C > 0$ such that for small b,

$$\mathbb{I}\left(q_b \sup_{t\in\mathcal{T}} d_{\mathcal{M}}\left(\nu(t), \widetilde{\nu}_b(t)\right)^{\beta_1/2} > 2^{\ell}\right) \leqslant C \sum_{k>\ell} \frac{b^2(2^k q_b^{-1})^{2/\beta_1}}{2^{2(k-1)} q_b^{-2}} = 4C \sum_{k>\ell} 2^{-2k(\beta_1-1)/\beta_1},$$

which converges to 0 as $\ell \to \infty$, and hence (2.5) follows.

Lastly, we note that for any $\gamma \in (0, 0.5)$, if $b \sim m^{-\gamma}$, then

$$\frac{(mb^2(-\log b)^{-1})^{-1/[2(\beta_2-1)]}}{(mb^2)^{-1/[2(\beta_2'-1)]}} \sim \frac{(\log m)^{1/[2(\beta_2-1)]}}{m^{(0.5-\gamma)\left[(\beta_2-1)^{-1}-(\beta_2'-1)^{-1}\right]}} \to 0, \quad \text{as } m \to \infty.$$

With $b \sim m^{-(\beta_1-1)/(2\beta_1+4\beta_2'-6)}$, it holds that $b^{2/(\beta_1-1)} \sim (mb^2)^{-1/[2(\beta_2'-1)]} \sim m^{-1/(\beta_1+2\beta_2'-3)}$, whence (2.7) follows, completing the proof. □

S.2. Proofs of the uniform convergence of local Fréchet regression with random targets as in Section 3.

Proof of Theorem 2. Given any fixed $\varepsilon > 0$, define

$$a_m = \min\left\{(mb^2)^{\beta_2/[4(\beta_2-1+\varepsilon/2)]}, [mb^2(-\log b)^{-1}]^{\beta_2/[4(\beta_2-1)]}\right\}.$$

We show for the bias and stochastic parts respectively that

$$\sup_{\omega_1\in\Omega_1} \sup_{t\in\mathcal{T}} d_{\mathcal{M}}\left(Y_{\omega_1}(t), \widetilde{Y}_{\omega_1,b}(t)\right) = O\left(b^{2/(\beta_1-1)}\right), \tag{S.24}$$

(S.25)

$$\limsup_{m\to\infty} \sup_{\omega_1\in\Omega_1} P_{\Omega_2}\left(a_m \sup_{t\in\mathcal{T}} d_{\mathcal{M}}\left(\widetilde{Y}_{\omega_1,b}(t), \widehat{Y}_{\omega_1,m}(t)\right)^{\beta_2/2} > C\right) \to 0, \quad \text{as } C \to \infty.$$

Observing that

$$\sup_{t\in\mathcal{T}} d_{\mathcal{M}}\left(Y(t), \widetilde{Y}(t)\right) \leqslant \sup_{\omega_1\in\Omega_1}\sup_{t\in\mathcal{T}} d_{\mathcal{M}}\left(Y_{\omega_1}(t), \widetilde{Y}_{\omega_1,b}(t)\right),$$

$$\limsup_{m\to\infty} P\left(\sup_{t\in\mathcal{T}} d_{\mathcal{M}}\left(\widetilde{Y}(t), \widehat{Y}(t)\right) > C a_m^{-2/\beta_2}\right)$$
$$\leqslant \limsup_{m\to\infty}\sup_{\omega_1\in\Omega_1} P_{\Omega_2}\left(a_m \sup_{t\in\mathcal{T}} d_{\mathcal{M}}\left(\widetilde{Y}_{\omega_1,b}(t), \widehat{Y}_{\omega_1,m}(t)\right)^{\beta_2/2} > C\right),$$

(3.8) follows, which implies (3.9) if $b \sim m^{-(\beta_1-1)/(2\beta_1+4\beta_2-6+2\varepsilon)}$.

For (S.24), we note that for any given $\omega_1 \in \Omega_1$, and $t \in \mathcal{T}$, $d_{\mathcal{M}}(Y_{\omega_1}(t), \widetilde{Y}_{\omega_1,b}(t)) = o(1)$, as $b \to 0$, by Theorem 1. We note that by (K0) and (U0),

$$\sup_{t\in\mathcal{T}} \mathbb{E}_{\Omega_2}\left[|v(T,t,b)|(T-t)^2 b^{-2}\right] = O(1), \quad \text{as } n\to\infty. \tag{S.26}$$

Defining

$$\phi_{\omega_1}(z,t) = \frac{\partial^2 M_{\omega_1}}{\partial t^2}(z,t), \tag{S.27}$$

it holds following similar arguments as given in the proof of (S.12) that

$$\phi_{\omega_1}(z,t) = \int d_{\mathcal{M}}^2(z',z)\frac{\partial^2}{\partial t^2}\left[\frac{f_{T|Z_{\omega_1}}(t,z')}{f_T(t)}\right] \mathrm{d}F_{Z_{\omega_1}}(z'). \tag{S.28}$$

In conjunction with (U0) and the boundedness of $\mathcal{M}$,

$$\sup_{\omega_1\in\Omega_1,\, t\in\mathcal{T},\, z\in\mathcal{M}} |\phi_{\omega_1}(z,t)| < \infty, \tag{S.29}$$

whence we obtain

$$\sup_{\omega_1\in\Omega_1,\, t\in\mathcal{T}} \left|\widetilde{M}_{\omega_1,b}(z,t) - M_{\omega_1}(z,t)\right|$$
$$\leqslant \frac{1}{2}b^2 \sup_{t\in\mathcal{T}} \mathbb{E}_{\Omega_2}\left[|v(T,t,b)|\left(\frac{T-t}{b}\right)^2\right] \sup_{\omega_1\in\Omega_1,\, t\in\mathcal{T},\, z\in\mathcal{M}} |\phi_{\omega_1}(z,t)|$$
$$= O(b^2).$$

This implies $\sup_{\omega_1\in\Omega_1,\, t\in\mathcal{T}} d_{\mathcal{M}}(Y_{\omega_1}(t), \widetilde{Y}_{\omega_1,b}(t)) = o(1)$ under (U1). Furthermore, by (S.26) and (S.28), there exists a constant $C > 0$ such that for m

large enough,
(S.30)
$$\sup_{\omega_1\in\Omega_1,\, t\in\mathcal{T}} \sup_{d_{\mathcal{M}}(z,Y_{\omega_1}(t))<\delta} \left|(\widetilde{M}_{\omega_1,b}-M_{\omega_1})(z,t)-(\widetilde{M}_{\omega_1,b}-M_{\omega_1})(Y_{\omega_1}(t),t)\right|$$
$$\leqslant \frac{1}{2}b^2 \sup_{t\in\mathcal{T}} \mathbb{E}_{\Omega_2}\left[|v(T,t,b)|\left(\frac{T-t}{b}\right)^2\right]\cdot \sup_{\substack{\omega_1\in\Omega_1\\ t\in\mathcal{T}}} \sup_{d_{\mathcal{M}}(z,Y_{\omega_1}(t))<\delta} |\phi_{\omega_1}(z,t)-\phi_{\omega_1}(Y_{\omega_1}(t),t)|$$
$$\leqslant Cb^2\delta.$$

Using similar arguments as given in the proof of (2.5), with $q_b = b^{-\beta_1/(\beta_1-1)}$, there exists a constant $C>0$ such that for large m,

$$\mathbb{I}\left(\sup_{\omega_1\in\Omega_1}\sup_{t\in\mathcal{T}} d_{\mathcal{M}}\left(Y_{\omega_1}(t),\widetilde{Y}_{\omega_1,b}(t)\right)^{\beta_1/2} > 2^\ell q_b^{-1}\right)$$
$$\leqslant C\sum_{k>\ell}\frac{b^2(2^k q_b^{-1})^{2/\beta_1}}{2^{2(k-1)}q_b^{-2}} = 4C\sum_{k>\ell}2^{-2k(\beta_1-1)/\beta_1},$$

which converges to zero as $\ell\to\infty$, whence (S.24) follows.

Next, we establish (S.25). Let η be the minimum integer not less than $\log_2(a_m \mathrm{diam}(\mathcal{M})^{\beta_2/2}+1)$, and for any $R>0$, define sets

$$B_{R,m} = \left\{\sup_{t\in\mathcal{T}} m^{-1}\sum_{j=1}^{m}|\widehat{v}(T_j,t,b)-v(T_j,t,b)| \leqslant R\vartheta_m\right\}. \tag{S.31}$$

For any $\ell\in\mathbb{N}_+$, considering m large enough such that $\eta>\ell$,
(S.32)
$$\sup_{\omega_1\in\Omega_1} P_{\Omega_2}\left(a_m\sup_{t\in\mathcal{T}} d_{\mathcal{M}}\left(\widetilde{Y}_{\omega_1,b}(t),\widehat{Y}_{\omega_1,m}(t)\right)^{\beta_2/2} > 2^\ell\right) \leqslant P_{\Omega_2}(B^c_{R,m})$$
$$+\sum_{\ell<k\leqslant\eta}\sup_{\omega_1\in\Omega_1} P_{\Omega_2}\left(\left\{2^{k-1}\leqslant a_m\sup_{t\in\mathcal{T}} d_{\mathcal{M}}\left(\widetilde{Y}_{\omega_1,b}(t),\widehat{Y}_{\omega_1,m}(t)\right)^{\beta_2/2} < 2^k\right\}\cap B_{R,m}\right).$$

Under (K0) and (U0), it follows from similar arguments as given in the proof of Theorem B of Silverman (1978) that there exists $R>0$ with $P_{\Omega_2}(B^c_{R,m}) = 0$, for m large enough. Regarding the convergence of the second term on the right hand side of (S.32), using similar arguments as given in the proof of (2.6), it holds for

$$J^{(2)}_{t,\omega_1}(z) = m^{-1}\sum_{j=1}^{m}[v(T_j,t,b)d^2_{\mathcal{M}}(Z_{\omega_1 j},z)] - \mathbb{E}_{\Omega_2}[v(T,t,b)d^2_{\mathcal{M}}(Z_{\omega_1},z)]$$

that

$$\mathbb{E}\left(\sup_{t\in\mathcal{T}}\sup_{d_{\mathcal{M}}(z,\widetilde{Y}_{\omega_1,b}(t))<\delta}\left|J^{(2)}_{t,\omega_1}(z)-J^{(2)}_{t,\omega_1}(\widetilde{Y}_{\omega_1,b}(t))\right|\right)$$
$$=O\left(\delta^{2-v}(mb^2)^{-1/2}+\delta(mb^2)^{-1/2}\sqrt{-\log b}\right),$$

where the O term is uniform over $\omega_1\in\Omega_1$, by (S.24), (U1) and Theorem 2.14.2 of van der Vaart and Wellner (1996). Under (K0) and (U0)–(U3), the second term on the right hand side of (S.32) can be bounded by

$$C\sum_{\ell<k\leqslant\eta}2^{-2(k-1)}a_m^2(mb^2)^{-1/2}\left[(2^ka_m^{-1})^{2(1-\varepsilon/2)/\beta_2}+(2^ka_m^{-1})^{2/\beta_2}\sqrt{-\log b}\right]$$
$$\leqslant 4Ca_m^{2(\beta_2-1+\varepsilon/2)/\beta_2}(mb^2)^{-1/2}\sum_{k>\ell}2^{-2k(\beta_2-1+\varepsilon/2)/\beta_2}$$
$$+4Ca_m^{2(\beta_2-1)/\beta_2}(mb^2)^{-1/2}\sqrt{-\log b}\sum_{k>\ell}2^{-2k(\beta_2-1)/\beta_2}$$
$$\leqslant 4C\sum_{k>\ell}2^{-2k(\beta_2-1+\varepsilon/2)/\beta_2}+4C\sum_{k>\ell}2^{-2k(\beta_2-1)/\beta_2},$$

which converges to zero as $\ell\to\infty$, whence (S.25) follows. □

S.3. Proofs of results in Section 4.

Proof of Corollary 1. By (4.1), (4.2), and (D1),

$$\Lambda(\widehat{t}_{\min})-\Lambda(t_{\min})\leqslant\Lambda(\widehat{t}_{\min})-\Lambda(t_{\min})+\widehat{\Lambda}(t_{\min})-\widehat{\Lambda}(\widehat{t}_{\min})$$
$$\leqslant 2\sup_{t\in\mathcal{T}}|\widehat{\Lambda}(t)-\Lambda(t)|\leqslant 2C_1\sup_{t\in\mathcal{T}}d_{\mathcal{M}}(\widehat{\nu}_m(t),\nu(t))^{\alpha_1}.$$

This implies $|\widehat{t}_{\min}-t_{\min}|=o_P(1)$ in conjunction with (D2) and Theorem 1, whence (4.3) follows under (D3). □

S.4. Proofs of results in Section 5. We first present two auxiliary results (Lemmas S.1 and S.2), where Lemma S.1 is needed for the proof of Lemma S.2, and Lemma S.2 implies that the coefficient vector $\theta_{g_{i'i}}$ is the unique minimizer of $\mathscr{C}_\mu$ given in (5.10) under certain constraints. This will be used to derive the rate of convergence for the M-estimator $\widehat{\theta}_{g_{i'i}}$ of the coefficient vector in Theorem 3.

Lemma S.1. *For any $g,g^*\in\mathcal{W}$ such that $d_{\mathcal{M}}(\mu(g(t)),\mu(g^*(t)))=0$, for all $t\in\mathcal{T}$, where μ satisfies (W1)–(W2), it holds that*

$$g(t)=g^*(t),\quad \textit{for all } t\in\mathcal{T}.$$

PROOF. Suppose there exists $x_0 \in (0,\tau)$ such that $g(x_0) \neq g^*(x_0)$. Without loss of generality, we assume $g(x_0) < g^*(x_0)$. Let $t_0 = g(x_0)$. We define a sequence $\{t_k\}_{k=1}^{\infty}$ iteratively by

$$t_k = g^*(g^{-1}(t_{k-1})), \quad \text{for } k = 1, 2, \ldots \tag{S.33}$$

Then it can be shown by induction that $\{t_k\}_{k=1}^{\infty}$ is a strictly increasing sequence, whence there exists $t^* \in \mathcal{T}$ such that $t_k \uparrow t^*$ as $k \to \infty$, since $\{t_k\}_{k=1}^{\infty} \subset \mathcal{T} = [0,\tau]$. Due to the continuity of g and g^*, taking $k \to \infty$ on both sides of (S.33) provides $t^* = g^*(g^{-1}(t^*))$. Let $x^* = g^{-1}(t^*)$, then $t^* = g(x^*) = g^*(x^*)$. Furthermore, for all $k = 1, 2, \ldots,$

$$d_{\mathcal{M}}\left(\mu(t_k), \mu(t_{k-1})\right) = d_{\mathcal{M}}\left(\mu(g^*(g^{-1}(t_{k-1}))), \mu(g(g^{-1}(t_{k-1}))\right) = 0, \tag{S.34}$$

since $d_{\mathcal{M}}(\mu(g(t)), \mu(g^*(t))) = 0$, for all $t \in \mathcal{T}$. By (W2), there exists $s_0 \in (t_0, t_1)$ such that

$$d_{\mathcal{M}}\left(\mu(s_0), \mu(t_0)\right) = d_{\mathcal{M}}\left(\mu(s_0), \mu(t_1)\right) > 0. \tag{S.35}$$

Similarly, we can iteratively define another sequence $s_k = g^*(g^{-1}(s_{k-1}))$, for $k = 1, 2, \ldots,$ for which it also holds that $s_k \uparrow t^*$ as $k \to \infty$ and $d_{\mathcal{M}}(\mu(s_k), \mu(s_{k-1})) = 0$, for all $k = 1, 2, \ldots$ By (S.34), $d_{\mathcal{M}}(\mu(t_0), \mu(t^*)) = d_{\mathcal{M}}(\mu(t_k), \mu(t^*))$ for all $k = 1, 2, \ldots$ Taking $k \to \infty$ yields $d_{\mathcal{M}}(\mu(t_0), \mu(t^*)) = \lim_{k\to\infty} d_{\mathcal{M}}(\mu(t_k), \mu(t^*)) = 0$, by (W1). Similarly, it can be verified that $d_{\mathcal{M}}(\mu(s_0), \mu(t^*)) = 0$, whence we obtain $d_{\mathcal{M}}(\mu(s_0), \mu(t_0)) = 0$, which contradicts (S.35). □

LEMMA S.2. *Suppose (W1)–(W3) hold. For any $i, i' = 1, \ldots, n$, such that $i \neq i'$, the coefficient vector $\theta_{g_{i'i}}$ corresponding to the pairwise warping function $g_{i'i}$ is the unique minimizer of the following constrained optimization problem*

$$\min_{\theta \in \mathbb{R}^{p+1}} \mathscr{C}_{\mu}(\theta; h_{i'}, h_i), \tag{S.36}$$

$$\textit{subject to } I_0(\theta) = \theta_{p+1} - \tau = 0,\ I_k(\theta) = \theta_{k-1} - \theta_k + \xi \leqslant 0,\ k = 1, 2, \ldots, p+1,$$

where $\mathscr{C}_{\mu}$ is as per (5.10), $\theta = (\theta_1, \ldots, \theta_{p+1})^\top \in \mathbb{R}^{p+1}$, $\theta_0 = 0$, and $\xi \in (0, cC^{-1}\tau/(p+1))$ is a constant, with c and C as per (W3).

PROOF. Considering the fact that $\mathscr{C}_{\mu}(\theta; h_{i'}, h_i) \geqslant 0$, for all $\theta \in \mathbb{R}^{p+1}$, and that $\mathscr{C}_{\mu}(\theta_{g_{i'i}}; h_i, h_{i'}) = 0$, since $h_{i'}^{-1}(\theta_{g_{i'i}}^\top A(t))) = h_{i'}^{-1}(h_{i'}(h_i^{-1}(t))) = h_i^{-1}(t)$, $\theta_{g_{i'i}}$ is a constrained minimizer of the optimization problem in (S.36) in conjunction with (W3); it suffices to show the uniqueness. Suppose θ_* is a constrained minimizer of $\mathscr{C}_{\mu}(\cdot; h_{i'}, h_i)$. The Lagrangian function corresponding to (S.36) is $\mathscr{L}_\mu(\theta; h_{i'}, h_i) = \mathscr{C}_\mu(\theta; h_{i'}, h_i) + \sum_{k=0}^{p+1} \zeta_k I_k(\theta)$, where $\zeta_0 \in \mathbb{R}$, and $\zeta_k \geqslant 0$ for $k = 1, \ldots, p+1$. By the Karush–Kuhn–Tucker condition (Karush, 1939; Kuhn and Tucker, 1951), there exist $\zeta_k^* \geqslant 0$, $k = 0, 1, \ldots, p+1$, such that $\nabla \mathscr{C}_\mu(\theta_*; h_{i'}, h_i) + \sum_{k=0}^{p+1} \zeta_k^* \nabla I_k(\theta_*) = 0$ and $\zeta_k^* I_k(\theta_*) = 0$ for $k = 1, \ldots, p+1$. By (W3), it holds that $\nabla \mathscr{C}_\mu(\theta_*; h_{i'}, h_i) = 0$, which, in conjunction with (W2) and (W3), implies

$$d_{\mathcal{M}}\left(\mu(h_{i'}^{-1}[\theta_*^\top A(t)]), \mu(h_i^{-1}(t))\right) = 0,$$

almost everywhere on $\mathcal{T}$ and hence for all $t \in \mathcal{T}$ by (W1) and the continuity of h_i and $h_{i'}$. Applying Lemma S.1 yields

$$h_{i'}^{-1}[\theta_*^\top A(t)] = h_i^{-1}(t), \quad \text{for all } t \in \mathcal{T},$$

and hence

$$\theta_*^\top A(t) = h_{i'} \circ h_i^{-1}(t) = \theta_{g_{i'i}}^\top A(t), \quad \text{for all } t \in \mathcal{T}.$$

For any $k = 0, 1, \ldots, p$ and $t \in [t_k, t_{k+1})$,

$$\theta_*^\top A(t) = (\theta_*)_{k+1} \frac{t - t_k}{t_{k+1} - t_k} - (\theta_*)_k \frac{t - t_{k+1}}{t_{k+1} - t_k},$$
$$\theta_{g_{i'i}}^\top A(t) = (\theta_{g_{i'i}})_{k+1} \frac{t - t_k}{t_{k+1} - t_k} - (\theta_{g_{i'i}})_k \frac{t - t_{k+1}}{t_{k+1} - t_k}.$$

If there exists $k_0 \in \{0, 1, \ldots, p\}$ such that $(\theta_{g_{i'i}})_{k_0} \neq (\theta_*)_{k_0}$, then $(\theta_{g_{i'i}})_k \neq (\theta_*)_k$, for all $k = k_0, \ldots, p+1$, which is contradictory to $(\theta_{g_{i'i}})_{p+1} = (\theta_*)_{p+1} = \tau$. Thus, $\theta_* = \theta_{g_{i'i}}$. □

PROOF OF THEOREM 3. For any $i, i' = 1, \ldots, n$ such that $i \neq i'$, a Taylor expansion yields

$$\begin{aligned}
&\mathscr{C}_\mu(\theta; h_{i'}, h_i) - \mathscr{C}_\mu(\theta_{g_{i'i}}; h_{i'}, h_i) \\
&\quad = \frac{1}{2}(\theta - \theta_{g_{i'i}})^\top \frac{\partial^2 \mathscr{C}_\mu}{\partial\theta\partial\theta^\top}(\theta_{g_{i'i}}; h_{i'}, h_i)(\theta - \theta_{g_{i'i}}) + o\left(\|\theta - \theta_{g_{i'i}}\|^2\right) \\
&\quad = \frac{1}{2}\int_{\mathcal{T}} \left[\frac{\partial^2 d_\mu^2}{\partial s^2}(s, h_i^{-1}(t)) \frac{1}{h_{i'}'(s)^2}\right]_{s = h_{i'}^{-1}(\theta_{g_{i'i}}^\top A(t))} \left[(\theta - \theta_{g_{i'i}})^\top A(t)\right]^2 \mathrm{d}t + o\left(\|\theta - \theta_{g_{i'i}}\|^2\right) \\
&\quad \geqslant \frac{1}{2} C^{-2} \left[\inf_{s=t \in \mathcal{T}} \frac{\partial^2 d_\mu^2}{\partial s^2}(s, t)\right] \int_{\mathcal{T}} \left[(\theta - \theta_{g_{i'i}})^\top A(t)\right]^2 \mathrm{d}t + o\left(\|\theta - \theta_{g_{i'i}}\|^2\right),
\end{aligned}$$

as $\|\theta - \theta_{g_{i'i}}\| \to 0$, where C is as per (W3), and

$$\inf_{s=t \in \mathcal{T}} (\partial^2 d_\mu / \partial s^2)(s, t) = \inf_{s=t \in \mathcal{T}} 2[(\partial d_\mu / \partial s)(s, t)]^2 > 0$$

by (W2). Noting that $\int_{\mathcal{T}} A(t)A(t)^\top \mathrm{d}t$ is positive definite, there exist $\delta_0 > 0$ such that for all $\theta \in B_{\delta_0}(\theta_{g_{i'i}})$,

$$\mathscr{C}_\mu(\theta; h_{i'}, h_i) - \mathscr{C}_\mu(\theta_{g_{i'i}}; h_{i'}, h_i) \geqslant \frac{1}{4} C^{-2} \left[\inf_{s=t \in \mathcal{T}} \frac{\partial^2 d_\mu^2}{\partial s^2}(s,t) \right] \lambda_{\min}^A \|\theta - \theta_{g_{i'i}}\|^2,$$

where $\lambda_{\min}^A > 0$ is the smallest eigenvalue of $\int_{\mathcal{T}} A(t)A(t)^\top \mathrm{d}t$, and $B_\delta(\theta_{g_{i'i}})$ is a ball of radius δ centered at $\theta_{g_{i'i}}$. Furthermore, by Lemma S.2 and the compactness of the feasible region

$$\Theta_\xi := \{\theta \in \mathbb{R}^{p+1} : \theta_k - \theta_{k-1} \geqslant \xi,\ k = 1, \ldots, p+1,\ \theta_{p+1} = \tau\} \subset \Theta$$

of the optimization problem in (S.36), it holds for any $\delta > 0$ that

$$\inf_{\Theta_\xi \cap B_\delta(\theta_{g_{i'i}})^c} \mathscr{C}_\mu(\theta; h_{i'}, h_i) - \mathscr{C}_\mu(\theta_{g_{i'i}}; h_{i'}, h_i) > 0.$$

Observing that $\|\theta - \theta_{g_{i'i}}\|^2 \leqslant p\tau^2$, let

$$C_0 = \min\left\{ \frac{1}{4} C^{-2} \inf_{s=t\in\mathcal{T}} \frac{\partial^2 d_\mu^2}{\partial s^2}(s,t) \lambda_{\min}^A, (p\tau^2)^{-1} \inf_{\Theta_\xi \cap B_{\delta_0}(\theta_{g_{i'i}})^c} \left[\mathscr{C}_\mu(\theta; h_{i'}, h_i) - \mathscr{C}_\mu(\theta_{g_{i'i}}; h_{i'}, h_i) \right] \right\},$$

where we note that $C_0 > 0$ and

$$\mathscr{C}_\mu(\theta; h_{i'}, h_i) - \mathscr{C}_\mu(\theta_{g_{i'i}}; h_{i'}, h_i) \geqslant C_0 \|\theta - \theta_{g_{i'i}}\|^2, \quad \text{for all } \theta \in \Theta_\xi.$$

By (5.7), $\widehat{\theta}_{g_{i'i}}$ minimizes $\mathscr{C}_{\widehat{Y},\lambda}(\theta; \widehat{Y}_{i'}, \widehat{Y}_i)$ subject to the constraint $\theta \in \Theta_\xi$ for some $\xi \in (0, cC^{-1}\tau/(p+1))$, whence we obtain

(S.37)

$$\begin{aligned} &\|\widehat{\theta}_{g_{i'i}} - \theta_{g_{i'i}}\| \\ &\leqslant C_0^{-1/2} \left[\mathscr{C}_\mu(\widehat{\theta}_{g_{i'i}}; h_{i'}, h_i) - \mathscr{C}_\mu(\theta_{g_{i'i}}; h_{i'}, h_i) + \mathscr{C}_{\widehat{Y},\lambda}(\theta_{g_{i'i}}; \widehat{Y}_{i'}, \widehat{Y}_i) - \mathscr{C}_{\widehat{Y},\lambda}(\widehat{\theta}_{g_{i'i}}; \widehat{Y}_{i'}, \widehat{Y}_i) \right]^{1/2} \\ &\leqslant \sqrt{2} C_0^{-1/2} \sup_{\theta \in \Theta} \left| \mathscr{C}_{\widehat{Y},\lambda}(\theta; \widehat{Y}_{i'}, \widehat{Y}_i) - \mathscr{C}_\mu(\theta; h_{i'}, h_i) \right|^{1/2}. \end{aligned}$$

Furthermore, noting that

$$\begin{aligned} &\left| \mathscr{C}_{\widehat{Y},\lambda}(\theta; \widehat{Y}_{i'}, \widehat{Y}_i) - \mathscr{C}_\mu(\theta; h_{i'}, h_i) \right| \\ &\leqslant \int_{\mathcal{T}} \left| d_{\mathcal{M}}^2 \left(\widehat{Y}_i(t), \widehat{Y}_{i'}(\theta^\top A(t)) \right) - d_{\mathcal{M}}^2 \left(\mu(h_i^{-1}(t)), \mu(h_{i'}^{-1}[\theta^\top A(t)]) \right) \right| \mathrm{d}t + \lambda \int_{\mathcal{T}} \left(\theta^\top A(t) - t \right)^2 \mathrm{d}t \\ &\leqslant 2\mathrm{diam}(\mathcal{M}) \int_{\mathcal{T}} \left| d_{\mathcal{M}} \left(\widehat{Y}_i(t), \widehat{Y}_{i'}(\theta^\top A(t)) \right) - d_{\mathcal{M}} \left(Y_i(t), Y_{i'}(\theta^\top A(t)) \right) \right| \mathrm{d}t + \frac{\tau^3}{3}\lambda \\ &\leqslant 2\mathrm{diam}(\mathcal{M}) \int_{\mathcal{T}} \left[d_{\mathcal{M}} \left(\widehat{Y}_i(t), Y_i(t) \right) + d_{\mathcal{M}} \left(\widehat{Y}_{i'}(\theta^\top A(t)), Y_{i'}(\theta^\top A(t)) \right) \right] \mathrm{d}t + \frac{\tau^3}{3}\lambda, \end{aligned}$$

(S.37) can be bounded as

$$\begin{aligned}
&\|\widehat{\theta}_{g_{i'i}} - \theta_{g_{i'i}}\| \\
&\quad \leqslant \left(\frac{4\mathrm{diam}(\mathcal{M})\tau}{C_0}\right)^{1/2} \left[\sup_{t\in\mathcal{T}} d_{\mathcal{M}}\left(\widehat{Y}_i(t), Y_i(t)\right)^{1/2} + \sup_{t\in\mathcal{T}} d_{\mathcal{M}}\left(\widehat{Y}_{i'}(t), Y_{i'}(t)\right)^{1/2}\right] \\
&\qquad + \left(\frac{2\tau^3}{3C_0}\right)^{1/2} \lambda^{1/2},
\end{aligned}$$

whence (5.11) follows by Theorem 2, and hence (5.12) follows by observing that

$$\sup_{t\in\mathcal{T}} |\widehat{g}_{i'i}(t) - g_{i'i}(t)| = \sup_{t\in\mathcal{T}} |(\widehat{\theta}_{g_{i'i}} - \theta_{g_{i'i}})^\top A(t)| \leqslant \|\widehat{\theta}_{g_{i'i}} - \theta_{g_{i'i}}\|,$$

since $\sup_{t\in\mathcal{T}} |A_k(t)| \leqslant 1$, for all $k = 1, \dots, p+1$. □

Proof of Corollary 2. With $b_i \sim m_i^{-(1-\varepsilon')(\beta_1-1)/(2\beta_1+4\beta_2-6+2\varepsilon)}$, (5.13) follows from (3.8) in Theorem 2. We only need to show (5.14).

Given any fixed $\varepsilon > 0$ and $\varepsilon' \in (0,1)$, define $\gamma = \varepsilon'\beta_2/[4(\beta_2 - 1 + \varepsilon/2)]$, and

$$a_{m_i} = \min\left\{(m_i b_i^2)^{\beta_2/[4(\beta_2-1+\varepsilon/2)]}, [m_i b_i^2(-\log b_i)^{-1}]^{\beta_2/[4(\beta_2-1)]}\right\}.$$

We show that for the bias part,

(S.38)
$$\sup_{1\leqslant i\leqslant n} \sup_{\omega_1\in\Omega_1} \sup_{t\in\mathcal{T}} d_{\mathcal{M}}\left(Y_{\omega_1}(t), \widetilde{Y}_{\omega_1,b_i}(t)\right) = O\left(\sup_{1\leqslant i\leqslant n} b_i^{2/(\beta_1-1)}\right) = O\left(b^{2/(\beta_1-1)}\right),$$

and for the stochastic part,

(S.39)
$$\limsup_{n\to\infty} \sum_{i=1}^{n} \sup_{\omega_1\in\Omega_1} P_{\Omega_2}\left(a_{m_i} m_i^{-\gamma} \sup_{t\in\mathcal{T}} d_{\mathcal{M}}\left(\widetilde{Y}_{\omega_1,b_i}(t), \widehat{Y}_{\omega_1,m_i}(t)\right)^{\beta_2/2} > C\right) \to 0, \quad \text{as } C\to\infty.$$

For each $i = 1, \dots, n$, and $t \in \mathcal{T}$, define $\widetilde{Y}_i(t)\colon \Omega_1 \to \mathcal{M}$ as

$$\widetilde{Y}_i(t)(\omega_1) = \widetilde{Y}_{\omega_1,b_i}(t).$$

Observing that

$$n^{-1}\sum_{i=1}^{n}\sup_{t\in\mathcal{T}} d_{\mathcal{M}}\left(Y_i(t),\widetilde{Y}_i(t)\right)^{\alpha} \leqslant \left[\sup_{1\leqslant i\leqslant n}\sup_{\omega_1\in\Omega_1}\sup_{t\in\mathcal{T}} d_{\mathcal{M}}\left(Y_{\omega_1}(t),\widetilde{Y}_{\omega_1,b_i}(t)\right)\right]^{\alpha},$$

$$\limsup_{n\to\infty} P\left(n^{-1}\sum_{i=1}^{n}\sup_{t\in\mathcal{T}} d_{\mathcal{M}}\left(\widetilde{Y}_i(t),\widehat{Y}_i(t)\right)^{\alpha} > C\sup_{1\leqslant i\leqslant n}(a_{m_i}m_i^{-\gamma})^{-2\alpha/\beta_2}\right)$$
$$\leqslant \limsup_{n\to\infty}\sum_{i=1}^{n} P\left(\sup_{t\in\mathcal{T}} d_{\mathcal{M}}\left(\widetilde{Y}_i(t),\widehat{Y}_i(t)\right) > C(a_{m_i}m_i^{-\gamma})^{-2/\beta_2}\right)$$
$$\leqslant \limsup_{n\to\infty}\sum_{i=1}^{n}\sup_{\omega_1\in\Omega_1} P_{\Omega_2}\left(a_{m_i}m_i^{-\gamma}\sup_{t\in\mathcal{T}} d_{\mathcal{M}}\left(\widetilde{Y}_{\omega_1,b_i}(t),\widehat{Y}_{\omega_1,m_i}(t)\right)^{\beta_2/2} > C\right),$$

(5.14) follows if $b_i \sim m_i^{-(1-\varepsilon')(\beta_1-1)/(2\beta_1+4\beta_2-6+2\varepsilon)}$.

For (S.38), we first show

$$\sup_{1\leqslant i\leqslant n}\sup_{\omega_1\in\Omega_1,\, t\in\mathcal{T}} d_{\mathcal{M}}(Y_{\omega_1}(t),\widetilde{Y}_{\omega_1,b_i}(t)) = o(1).$$

By the Cauchy criterion for uniform convergence, it suffices to show

(S.40)
$$\sup_{\omega_1\in\Omega_1,\, t\in\mathcal{T}}\left|\sup_{1\leqslant i\leqslant n} d_{\mathcal{M}}\left(Y_{\omega_1}(t),\widetilde{Y}_{\omega_1,b_i}(t)\right) - \sup_{1\leqslant i'\leqslant n'} d_{\mathcal{M}}\left(Y_{\omega_1}(t),\widetilde{Y}_{\omega_1,b_{i'}}(t)\right)\right| \to 0,$$

as $n, n' \to \infty$. We note that by (K0), (U0), and (W4),

(S.41)
$$\sup_{1\leqslant i\leqslant n}\sup_{t\in\mathcal{T}}\mathbb{E}_{\Omega_2}\left[|v(T,t,b_i)|(T-t)^2 b_i^{-2}\right] = O(1), \quad \text{as } n\to\infty,$$

whence in conjunction with (S.29) we obtain

$$\sup_{1\leqslant i\leqslant n}\sup_{\substack{\omega_1\in\Omega_1,\, t\in\mathcal{T}\\ z\in\mathcal{M}}}\left|\widetilde{M}_{\omega_1,b_i}(z,t) - M_{\omega_1}(z,t)\right|$$
$$\leqslant \frac{1}{2}\sup_{1\leqslant i\leqslant n} b_i^2 \sup_{1\leqslant i\leqslant n}\sup_{t\in\mathcal{T}}\mathbb{E}_{\Omega_2}\left[|v(T,t,b_i)|\left(\frac{T-t}{b_i}\right)^2\right]\sup_{\substack{\omega_1\in\Omega_1,\, t\in\mathcal{T}\\ z\in\mathcal{M}}}|\phi_{\omega_1}(z,t)|$$
$$= O\left(b(n)^2\right),$$

where ϕ_{ω_1} is defined as per (S.27). Hence,

$$\sup_{1\leqslant i\leqslant n,\ 1\leqslant i'\leqslant n'}\sup_{\substack{\omega_1\in\Omega_1,\, t\in\mathcal{T}\\ z\in\mathcal{M}}}\left|\widetilde{M}_{\omega_1,b_i}(z,t) - \widetilde{M}_{\omega_1,b_{i'}}(z,t)\right|$$
$$\leqslant \sup_{1\leqslant i\leqslant n}\sup_{\substack{\omega_1\in\Omega_1,\, t\in\mathcal{T}\\ z\in\mathcal{M}}}\left|\widetilde{M}_{\omega_1,b_i}(z,t) - M_{\omega_1}(z,t)\right| + \sup_{1\leqslant i'\leqslant n'}\sup_{\substack{\omega_1\in\Omega_1,\, t\in\mathcal{T}\\ z\in\mathcal{M}}}\left|\widetilde{M}_{\omega_1,b_{i'}}(z,t) - M_{\omega_1}(z,t)\right|$$
$$= O\left(b(n)^2\right) + O\left(b(n')^2\right),$$

as $n, n' \to \infty$. Observing that

$$\begin{aligned}
&\sup_{\omega_1\in\Omega_1,\, t\in\mathcal{T}} \left| \sup_{1\leqslant i\leqslant n} d_{\mathcal{M}}\left(Y_{\omega_1}(t), \widetilde{Y}_{\omega_1,b_i}(t)\right) - \sup_{1\leqslant i'\leqslant n'} d_{\mathcal{M}}\left(Y_{\omega_1}(t), \widetilde{Y}_{\omega_1,b_{i'}}(t)\right)\right| \\
&\quad\leqslant \sup_{\omega_1\in\Omega_1,\, t\in\mathcal{T}} \sup_{1\leqslant i\leqslant n,\ 1\leqslant i'\leqslant n'} \left| d_{\mathcal{M}}\left(Y_{\omega_1}(t), \widetilde{Y}_{\omega_1,b_i}(t)\right) - d_{\mathcal{M}}\left(Y_{\omega_1}(t), \widetilde{Y}_{\omega_1,b_{i'}}(t)\right)\right| \\
&\quad\leqslant \sup_{\omega_1\in\Omega_1,\, t\in\mathcal{T}} \sup_{1\leqslant i\leqslant n,\ 1\leqslant i'\leqslant n'} d_{\mathcal{M}}\left(\widetilde{Y}_{\omega_1,b_i}(t), \widetilde{Y}_{\omega_1,b_{i'}}(t)\right),
\end{aligned}$$

(S.40) follows in conjunction with (U1).

Using similar arguments as given in the proof of (2.5), with $q_{b_i} = b_i^{-\beta_1/(\beta_1-1)}$, by (S.30), there exists a constant $C > 0$ such that for large n,

$$\begin{aligned}
&\mathbb{I}\left(\sup_{1\leqslant i\leqslant n} \sup_{\omega_1\in\Omega_1} \sup_{t\in\mathcal{T}} d_{\mathcal{M}}\left(Y_{\omega_1}(t), \widetilde{Y}_{\omega_1,b_i}(t)\right)^{\beta_1/2} > 2^{\ell} \sup_{1\leqslant i\leqslant n} q_{b_i}^{-1}\right) \\
&\quad\leqslant \sup_{1\leqslant i\leqslant n} \mathbb{I}\left(\sup_{\omega_1\in\Omega_1} \sup_{t\in\mathcal{T}} d_{\mathcal{M}}\left(Y_{\omega_1}(t), \widetilde{Y}_{\omega_1,b_i}(t)\right)^{\beta_1/2} > 2^{\ell} \sup_{1\leqslant i\leqslant n} q_{b_i}^{-1}\right) \\
&\quad\leqslant \sup_{1\leqslant i\leqslant n} \mathbb{I}\left(q_{b_i} \sup_{\omega_1\in\Omega_1} \sup_{t\in\mathcal{T}} d_{\mathcal{M}}\left(Y_{\omega_1}(t), \widetilde{Y}_{\omega_1,b_i}(t)\right)^{\beta_1/2} > 2^{\ell}\right) \\
&\quad\leqslant \sup_{1\leqslant i\leqslant n} C \sum_{k>\ell} \frac{b_i^2 (2^k q_{b_i}^{-1})^{2/\beta_1}}{2^{2(k-1)} q_{b_i}^{-2}} = 4C \sum_{k>\ell} 2^{-2k(\beta_1-1)/\beta_1},
\end{aligned}$$

which converges to zero as $\ell \to \infty$, whence (S.38) follows.

Furthermore, by (W4), replacing a_m with $a_{m_i} m_i^{-\gamma}$ in the proof of (S.25) yields

$$\begin{aligned}
&\limsup_{n\to\infty} \sum_{i=1}^{n} \sup_{\omega_1\in\Omega_1} P_{\Omega_2}\left(a_{m_i} m_i^{-\gamma} \sup_{t\in\mathcal{T}} d_{\mathcal{M}}\left(\widetilde{Y}_{\omega_1,b_i}(t), \widehat{Y}_{\omega_1,m_i}(t)\right)^{\beta_2/2} > 2^{\ell}\right) \\
&\quad\leqslant 4C \sum_{k>\ell} 2^{-2k(\beta_2-1+\varepsilon/2)/\beta_2} \limsup_{n\to\infty} \sum_{i=1}^{n} m_i^{-2\gamma(\beta_2-1+\varepsilon/2)/\beta_2} \\
&\qquad + 4C \sum_{k>\ell} 2^{-2k(\beta_2-1)/\beta_2} \limsup_{n\to\infty} \sum_{i=1}^{n} m_i^{-2\gamma(\beta_2-1)/\beta_2} \\
&\quad\leqslant 4C \sum_{k>\ell} 2^{-2k(\beta_2-1+\varepsilon/2)/\beta_2} \limsup_{n\to\infty} n m^{-2\gamma(\beta_2-1+\varepsilon/2)/\beta_2} \\
&\qquad + 4C \sum_{k>\ell} 2^{-2k(\beta_2-1)/\beta_2} \limsup_{n\to\infty} n m^{-2\gamma(\beta_2-1)/\beta_2},
\end{aligned}$$

which converges to zero as $\ell \to \infty$, whence (S.39) follows. □

Proof of Corollary 3. By (5.9) and Theorem 3,

$$\sup_{t\in\mathcal{T}}\left|\widehat{h}_i^{-1}(t)-h_i^{-1}(t)\right| \leqslant \frac{1}{n}\sum_{i'=1}^{n}\sup_{t\in\mathcal{T}}|\widehat{g}_{i'i}(t)-g_{i'i}(t)| + \sup_{t\in\mathcal{T}}\left|\frac{1}{n}\sum_{i'=1}^{n} g_{i'i}(t)-h_i^{-1}(t)\right|.$$

By Theorem 2.7.5 of van der Vaart and Wellner (1996),

$$\begin{aligned}\sup_{t\in\mathcal{T}}\left|\frac{1}{n}\sum_{i'=1}^{n} g_{i'i}(t)-h_i^{-1}(t)\right| &= \sup_{t\in\mathcal{T}}\left|\frac{1}{n}\sum_{i'=1}^{n} h_{i'}(t)-t\right| \\ &= \sup_{t\in\mathcal{T}}\left|\frac{1}{n}\sum_{i'=1}^{n} h_{i'}(t)-\mathbb{E}(h_{i'}(t))\right| = O_P\left(n^{-1/2}\right).\end{aligned}$$

Observing that

$$\begin{aligned}&\frac{1}{n}\sum_{i'=1}^{n}\sup_{t\in\mathcal{T}}|\widehat{g}_{i'i}(t)-g_{i'i}(t)| \leqslant \frac{1}{n}\sum_{i'=1}^{n}\|\widehat{\theta}_{g_{i'i}}-\theta_{g_{i'i}}\| \\ &\quad\leqslant \text{const.}\left[\sup_{t\in\mathcal{T}} d_{\mathcal{M}}\left(\widehat{Y}_i(t),Y_i(t)\right)^{1/2} + n^{-1}\sum_{i'=1}^{n}\sup_{t\in\mathcal{T}} d_{\mathcal{M}}\left(\widehat{Y}_{i'}(t),Y_{i'}(t)\right)^{1/2}+\lambda^{1/2}\right],\end{aligned}$$

it follows from Corollary 2 that

$$\sup_{t\in\mathcal{T}}\left|\widehat{h}_i^{-1}(t)-h_i^{-1}(t)\right| = O\left(\lambda^{1/2}\right)+O_P\left(m^{-(1-\varepsilon')/[2(\beta_1+2\beta_2-3+\varepsilon)]}\right)+O_P\left(n^{-1/2}\right).$$

By (W3),

$$\begin{aligned}\sup_{t\in\mathcal{T}}\left|\widehat{h}_i(t)-h_i(t)\right| &= \sup_{t\in\mathcal{T}}\left|\widehat{h}_i(\widehat{h}_i^{-1}(t))-h_i(\widehat{h}_i^{-1}(t))\right| = \sup_{t\in\mathcal{T}}\left|t-h_i(\widehat{h}_i^{-1}(t))\right| \\ &\leqslant C\sup_{t\in\mathcal{T}}\left|h_i^{-1}(t)-h_i^{-1}(h_i(\widehat{h}_i^{-1}(t)))\right| = C\sup_{t\in\mathcal{T}}\left|h_i^{-1}(t)-\widehat{h}_i^{-1}(t)\right|,\end{aligned}$$

whence (5.15) follows. Furthermore, observing that

$$\begin{aligned}d_{\mathcal{M}}\left(\widehat{Y}_i(\widehat{h}_i(t)),Y_i(h_i(t))\right) &\leqslant d_{\mathcal{M}}\left(\widehat{Y}_i(\widehat{h}_i(t)),Y_i(\widehat{h}_i(t))\right)+d_{\mathcal{M}}\left(\mu(h_i^{-1}(\widehat{h}_i(t))),\mu(t)\right) \\ &\leqslant \sup_{t\in\mathcal{T}} d_{\mathcal{M}}\left(\widehat{Y}_i(t),Y_i(t)\right)+C_\mu\sup_{t\in\mathcal{T}}\left|h_i^{-1}(\widehat{h}_i(t))-t\right| \\ &= \sup_{t\in\mathcal{T}} d_{\mathcal{M}}\left(\widehat{Y}_i(t),Y_i(t)\right)+C_\mu\sup_{t\in\mathcal{T}}\left|h_i^{-1}(t)-\widehat{h}_i^{-1}(t)\right|,\end{aligned}$$

(5.16) also follows, which completes the proof. □

S.5. Simulation studies. In this section, we compare the performance of the proposed warping method for metric valued functional data for different choices of the penalty parameter λ and the number of knots p. Here, the time domain is $\mathcal{T} = [0,1]$, and the metric space $(\mathcal{M}, d_W)$ considered is the Wasserstein space of continuous probability measures on $[0,1]$ with finite second moments endowed with the $\mathcal{L}^2$-Wasserstein distance as in Example 1. With sample size $n = 30$, two cases were implemented with fixed trajectories μ in (5.2) as follows.

Case 1: $\mu(t) = \mathrm{Beta}(\xi_t, \gamma_t)$, where $\xi_t = 1.1 + 10(t - 0.4)^2$, and $\gamma_t = 2.6 + 1.5\sin(2\pi t - \pi)$, for $t \in \mathcal{T}$.

Case 2: $\mu(t) = N(\xi_t, \gamma_t^2)$ truncated on $[0,1]$, where $\xi_t = 0.1 + 0.8t$, and $\gamma_t = 0.6 + 0.2\sin(10\pi t)$, for $t \in \mathcal{T}$. Specifically, the corresponding distribution function is

$$F_{\mu(t)}(x) = \frac{\Phi((x - \xi_t)/\gamma_t) - \Phi(-\xi_t/\gamma_t)}{\Phi((1 - \xi_t)/\gamma_t) - \Phi(-\xi_t/\gamma_t)} \mathbf{1}_{[0,1]}(x) + \mathbf{1}_{(1,+\infty)}(x),\ x \in \mathbb{R},$$

where Φ is the distribution function of a standard Gaussian distribution.

We consider a family of perturbation/distortion functions $\{\mathscr{T}_a : a \in \mathbb{Z}\backslash\{0\}\}$, where $\mathscr{T}_a(x) = x - |a\pi|^{-1}\sin(a\pi x)$, for $x \in \mathbb{R}$. The warping functions h_i were generated through the distortion functions $\mathscr{T}_a$; specifically, $h_i = \mathscr{T}_{a_{i1}} \circ \mathscr{T}_{a_{i2}}$, where a_{il} are independent and identically distributed for $l = 1, 2$ and $i = 1, \ldots, n$, such that

$$P(a_{il} = -k) = P(a_{il} = k) = P(V_2 = k)/[2(1 - P(V_2 = 0))],$$

for any $k \in \mathbb{N}_+$, with $V_2 \sim \mathrm{Poisson}(2)$. We note that this generation mechanism ensures $h_i \in \mathcal{W}$ and $\mathbb{E}[h_i(t)] = t$, for any $t \in \mathcal{T}$. With μ and h_i, the sample trajectories Y_i were computed as per (5.2).

Set the number of discrete observations per trajectory $m_i = 30$, for all $i = 1, \ldots, n$. We sampled $T_{ij} \sim \mathrm{Uniform}(\mathcal{T})$ independently, for $j = 1, \ldots, m_i$, and $i = 1, \ldots, n$. Given a measurable function $g \colon \mathbb{R} \to \mathbb{R}$, a push-forward measure $g\#z$ is defined as $g\#z(E) = z(\{x : g(x) \in E\})$, for any distribution $z \in \mathcal{M}$ and set $E \subset \mathbb{R}$. The observed distributions Z_{ij} were generated by adding perturbations to the trajectory evaluated at T_{ij}, $Y(T_{ij})$, through push-forward measures; specifically $Z_{ij} = \mathscr{T}_{u_{ij}}\#(Y_i(T_{ij}))$, where u_{ij} are independent and identically distributed following $\mathrm{Uniform}\{\pm 4\pi, \pm 5\pi, \ldots, \pm 8\pi\}$, and are also independent of the observed times T_{ij}, $j = 1, \ldots, m_i$, and $i = 1, \ldots, n$.

We applied the proposed pairwise warping method to the simulated data with Epanechnikov kernel and bandwidths b_i chosen by cross-validation in the presmoothing step as per (5.6), where the local Fréchet regression was implemented using the R package `frechet` (Chen et al., 2020). We assessed the results through mean integrated squared errors (MISEs) for the estimated time-synchronized trajectories $\widehat{Y}_i(\widehat{h}_i(\cdot))$ and the estimated warping functions $\widehat{h}_i$ as per (5.6) and (5.9); specifically,

$$\text{(S.42)} \qquad \begin{aligned} \text{TMISE} &= \frac{1}{n}\sum_{i=1}^{n}\int_{\mathcal{T}} d_{\mathcal{M}}^2\left(\widehat{Y}_i(\widehat{h}_i(t)), \mu(t)\right) \mathrm{d}t, \\ \text{WMISE} &= \frac{1}{n}\sum_{i=1}^{n}\int_{\mathcal{T}} \left(\widehat{h}_i(t) - h_i(t)\right)^2 \mathrm{d}t. \end{aligned}$$

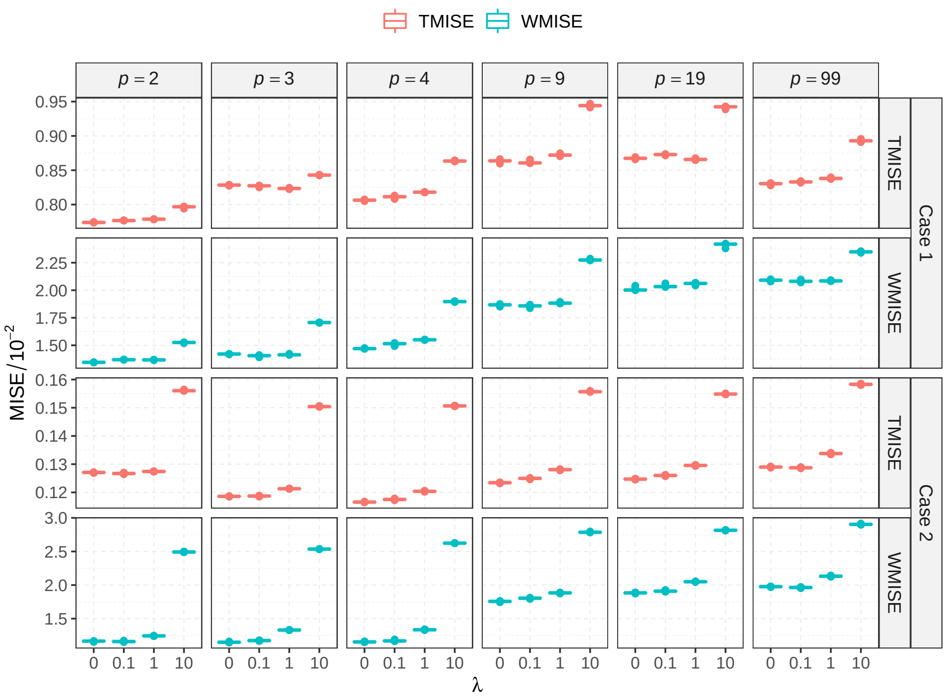

Fig S.1: Summary of TMISE (red) and WMISE (blue) as per (S.42) out of 1000 Monte Carlo runs for Case 1 (top two rows) and Case 2 (bottom two rows).

Since too many knots will result in shape distortion of the estimated warping function (Ramsay and Li, 1998), 1000 Monte Carlo runs were conducted for $p \in \{2, 3, 4, 9, 19, 99\}$, and $\lambda \in \{0\} \cup \{10^l : l = -1, 0, 1\}$. Results in terms of TMISE and WMISE for Case 1 and Case 2 are summarized in the boxplots in

Figure S.1, the top two rows show the results for Case 1 and the bottom two rows for Case 2. For both cases, for any given value of the number of knots p, the proposed estimators perform almost equally well in terms of TMISE and WMISE with small values (no more than 1) of the penalty parameter λ, and the performance turns worse as λ increases from 1 to 10. Furthermore, across different choices of the number of knots p, the estimators achieve the minimum estimation errors with small $p \in \{2, 3, 4\}$. Thus, the simulations indicate that the proposed method is not sensitive to the choice of p and λ when p and λ are relatively small, which is in agreement with findings in the literature (Ramsay and Li, 1998; Tang and Müller, 2008).

References

Chen, Y., Gajardo, A., Fan, J., Zhong, Q., Dubey, P., Han, K., Bhattacharjee, S. and Müller, H.-G. (2020). frechet: Statistical Analysis for Random Objects and Non-Euclidean Data R package version 0.1.0, available at `https://CRAN.R-project.org/package=frechet`.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*. Chapman & Hall, London.

Karush, W. (1939). Minima of functions of several variables with inequalities as side constraints, Master's thesis, Department of Mathematics, University of Chicago.

Kuhn, H. W. and Tucker, A. W. (1951). Nonlinear programming. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability* (J. Neyman, ed.) 481–492. University of California Press, Berkeley, CA.

Mack, Y. P. and Silverman, B. W. (1982). Weak and strong uniform consistency of kernel regression estimates. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **61** 405–415.

Petersen, A. and Müller, H.-G. (2019). Fréchet regression for random objects with Euclidean predictors. *The Annals of Statistics* **47** 691–719.

Ramsay, J. O. and Li, X. (1998). Curve registration. *Journal of the Royal Statistical Society: Series B* **60** 351–363.

Silverman, B. W. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *The Annals of Statistics* **6** 177–184.

Tang, R. and Müller, H.-G. (2008). Pairwise curve synchronization for functional data. *Biometrika* **95** 875–889.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York.