SUPPLEMENTAL MATERIAL TO "FRÉCHET REGRESSION FOR RANDOM OBJECTS WITH EUCLIDEAN PREDICTORS"

Proofs of Theoretical Results

Alexander Petersen and Hans-Georg Müller

S.1. Propositions 1–3.

PROPOSITION 1. *The space* (Ω, d_W) *defined in Example 1 satisfies assumptions (P0)–(P2) and (U0)–(U2).*

PROOF. For any distribution $\omega \in \Omega$, let $Q(\omega)$ be the corresponding quantile function. Similarly, $Q^{-1}(h) \in \Omega$ is the distribution corresponding to $h \in Q(\Omega)$. Let $\langle \cdot, \cdot \rangle_{L^2}$, $\|\cdot\|_{L^2}$ and $d_{L^2}(\cdot, \cdot)$ be the L^2 inner product, norm and distance on $[0,1]$, respectively. Since $E\left(|s(X,x)| \|Q(Y)\|_{L^2}\right)$ is finite, the Riesz Representation Theorem implies the existence of an element $g_x \in L^2[0,1]$ such that

$$E(s(X,x)\langle Q(Y), h\rangle_{L^2}) = \langle g_x, h\rangle_{L^2}$$

for all $h \in Q(\Omega)$. Define $\hat{g}_x = n^{-1}\sum_{i=1}^n s_{in}(x) Q(Y_i)$. Then properties of the L^2 distance imply

$$M(\omega, x) = E\left(s(X,x) d_{L^2}(Q(Y), g_x)^2\right) + d_{L^2}(Q(w), g_x)^2,$$
$$M_n(\omega, x) = \frac{1}{n}\sum_{i=1}^n s_{in}(x) d_{L^2}(Q(Y_i), \hat{g}_x)^2 + d_{L^2}(Q(\omega), \hat{g}_x)^2,$$

yielding the solutions

$$m_\oplus(x) = Q^{-1}\left(\operatorname*{argmin}_{h \in Q(\Omega)} d_{L^2}(h, g_x)^2\right), \quad \hat{m}_\oplus(x) = Q^{-1}\left(\operatorname*{argmin}_{h \in Q(\Omega)} d_{L^2}(h, \hat{g}_x)^2\right),$$

which exist and are unique by convexity of $Q(\Omega)$ for any $x \in \mathcal{R}^p$, hence proving (P0) and (U0). Additionally, $m_\oplus(x)$ is characterized by

$$\langle g_x - Q(m_\oplus(x)), h - Q(m_\oplus(x))\rangle_{L^2} \leq 0$$

for all $h \in Q(\Omega)$. Consequently, we may take $C = D = 1$, $\beta = \alpha = 2$ and η and τ arbitrary in (P2) and (U2).

Lastly, we show that (U1) holds, which of course implies (P1). Let $(\mathcal{Q}, d_2)$ be the space of quantile functions endowed with the L^2 metric. For the remainder of this proof, for any $g \in L^2[0,1]$, $\omega \in \Omega$ and $\gamma > 0$, $B^2_\gamma(g)$ refers to the L^2 ball of radius γ centered at g, while $B_\gamma(\omega)$ refers to the d_W ball of radius γ centered at ω. By Theorem 2.7.5 of van der Vaart and Wellner (1996),

$$N(\epsilon, \Omega, d_W) \leq N(\epsilon/2, \mathcal{Q}, d_2) \leq e^{K\epsilon^{-1}},$$

where K is independent of ϵ. For $Q \in \mathcal{Q}$, let $\mathcal{C}_\epsilon(Q) = \{g_u : u \in U\}$, $U \subset \mathcal{R}$, be a collection of L^2 functions such that $|U| = N(\epsilon, B^2_1(Q) \cap \mathcal{Q}, d_2) \leq e^{K\epsilon^{-1}}$ and the balls $B^2_\epsilon(g_u)$ cover $B^2_1(Q) \cap \mathcal{Q}$. For $\delta > 0$, define $\tilde{g}_u = Q + \delta(g_u - Q)$ and $C_{\delta\epsilon}(Q) = \{\tilde{g}_u : u \in U\}$, so that the collection $B^2_{\delta\epsilon}(\tilde{g}_u)$, $u \in U$, forms a covering of $B^2_\delta(Q) \cap \mathcal{Q}$. Thus, we have shown that

$$\sup_{\omega \in \Omega} \log N(\delta\epsilon, B_\delta(\omega), d_W) = \sup_{Q \in \mathcal{Q}} \log N(\delta\epsilon, B^2_\delta(Q) \cap \mathcal{Q}, d_2) \leq K\epsilon^{-1}.$$

To finish, observe that $\sup_{\|x\|_E \leq B} \log N(\delta\epsilon, B_\delta(m_\oplus(x)), d_W) \leq K\epsilon^{-1}$, so for any $\delta > 0$ the integral in (U1) is bounded by

$$\int_0^1 \sqrt{1 + K\epsilon^{-1}}\, d\epsilon \leq 1 + 2\sqrt{K} < \infty.$$

□

PROPOSITION 2. *The space (Ω, d_F) defined in Example 2 satisfies assumptions (P0)–(P2) and (U0)–(U2).*

PROOF. Here, Y is an $r \times r$ correlation matrix. Denote the elements of Y as $Y(j,k)$, $1 \leq j,k \leq r$. Let $\langle\cdot,\cdot\rangle_F$, $\|\cdot\|_F$ and $d_F(\cdot,\cdot)$ be the Frobenius inner product, norm and distance, respectively. Let $B_{jk}(x) = E\left(s(X,x)Y(j,k)\right)$ and $\hat{B}_{jk}(x) = n^{-1}\sum_{i=1}^n s_{in}(x)Y_i(j,k)$. Then properties of the Frobenius distance imply that

$$M(\omega, x) = M(B(x), x) + d_F(\omega, B(x))^2,$$
$$M_n(\omega, x) = M_n(\hat{B}(x), x) + d_F(\omega, \hat{B}(x))^2,$$

yielding the solutions

$$m_\oplus(x) = \operatorname*{argmin}_{\omega\in\Omega} d_F(\omega, B(x))^2, \quad \hat{m}_\oplus(x) = \operatorname*{argmin}_{\omega\in\Omega} d_F(\omega, \hat{B}(x))^2,$$

which exist and are unique by the convexity of Ω for any $x \in \mathcal{R}^p$, hence proving (P0) and (U0). Additionally, $m_\oplus(x)$ is characterized by

$$\langle B(x) - m_\oplus(x), \omega - m_\oplus(x)\rangle_F \le 0$$

for all $\omega \in \Omega$. Consequently, we may take η and τ arbitrary, $C = D = 1$ and $\beta = \alpha = 2$ in (P2) and (U2).

Lastly, since Ω is a bounded subset of the larger finite-dimensional Euclidean space of $r \times r$ matrices, for any $\omega \in \Omega$,

$$N(\delta\epsilon, B_\delta(\omega), d_F) = N(\epsilon, B_1(\omega), d_F) \le K\epsilon^{-r^2}$$

by an argument similar to that in Proposition 1, where $K > 1$ depends on r only. Thus, the integral in (U1) is bounded by

$$\begin{aligned}\int_0^1 \sqrt{1 + \log K - r^2 \log\epsilon}\, d\epsilon &= 1 + \log K + r\int_0^1 \sqrt{-\log\epsilon}\, d\epsilon \\ &= 1 + \log K + r\int_1^\infty e^{-y}\sqrt{y}\, d\epsilon < \infty.\end{aligned}$$

using the substitution $y = -\log\epsilon$. Since this bound does not depend on δ, (U1) holds and thus (P1) as well.

□

Proposition 3. *The space (Ω, d) defined in Example 3 satisfies (P1) and (U1) provided the Riemannian metric is equivalent to the ambient Euclidean metric. Let $T_\omega\Omega$ be the tangent bundle at ω and Exp_ω and Log_ω be the exponential and logarithmic manifold maps at ω. For $u \in T_\omega\Omega$, define*

$$g_\omega(u) = M\left(\mathrm{Exp}_\omega(u), x\right), \quad h_\omega(u) = M_n\left(\mathrm{Exp}_\omega(u), x\right).$$

If (P0) holds and $g''_{m_\oplus(x)}(0)$ is positive definite, then (P2) holds. Similarly, if (U0) holds then

$$\inf_{\|x\|_E \le B} \lambda_{\min}(g''_{m_\oplus(x)}(0)) > 0$$

implies (U2), where $\lambda_{\min}(A)$ is the smallest eigenvalue of a square matrix A.

PROOF. Since Ω is bounded and of finite dimension, (U1) follows by an argument similar to the last part of the previous proof due to metric equivalency, whence (P1) also follows. If (P0) holds, let ε be the injectivity radius at $m_\oplus(x)$ and consider ω such that $d(\omega, m_\oplus(x)) < \varepsilon$. Taking $u_x = \mathrm{Log}_{m_\oplus(x)}(\omega)$,

$$M(\omega, x) - M(m_\oplus(x), x) = g_{m_\oplus(x)}(u) - g_{m_\oplus(x)}(0) = u_x^T g''_{m_\oplus(x)}(u_x^*)u_x$$

for some u_x^* between 0 and u_x. Since $u_x^T u_x = d^2(\omega, m_\oplus(x))$ and $g_{m_\oplus(x)}$ is continuous, the condition on $g''_{m_\oplus(x)}(0)$ implies (P2) with $\beta = 2$. Similar arguments using the other conditions show that $\alpha = 2$ in (U2) is permissible. □

S.2. Proofs of results in Section 3. Throughout, the symbol $\rightsquigarrow$ will denote weak convergence and the notation $l^\infty(\Omega)$ denotes the space of bounded functions on Ω. The ordinary Euclidean norm on $\mathcal{R}^p$ will be denoted by $\|\cdot\|_E$ and the Frobenius norm by $\|\cdot\|_F$. For simplicity of notation, when x is fixed, the dependence of objects such as M, $m_\oplus$, etc. on x will be dropped.

PROOF OF THEOREM 1. We first consider fixed $x \in \mathcal{R}^p$. By Corollary 3.2.3 in van der Vaart and Wellner (1996), convergence of $\sup_{\omega\in\Omega} |M_n(\omega) - M(\omega)|$ to zero in probability is sufficient. To do this, we show $M_n \rightsquigarrow M$ in $l^\infty(\Omega)$ and apply 1.3.6 of van der Vaart and Wellner (1996). This weak convergence is proved (see Theorem 1.5.4 of van der Vaart and Wellner (1996)) by showing that

i) $M_n(\omega) - M(\omega) = o_p(1)$ for all $\omega \in \Omega$ and
ii) M_n is asymptotically equicontinuous in probability, i.e. for all $\varepsilon, \eta > 0$, there exists $\delta > 0$ such that

$$\limsup_n P\left(\sup_{d(\omega_1,\omega_2)<\delta} |M_n(\omega_1) - M_n(\omega_2)| > \varepsilon\right) < \eta.$$

Begin with i). Set

$$s_i = \left[1 + (X_i - \mu)^T \Sigma^{-1}(x - \mu)\right] \tag{S.1}$$

and define

$$\tilde{M}_n(\omega) = n^{-1} \sum_{i=1}^n s_i d^2(Y_i, \omega).$$

Then, for all $\omega \in \Omega$, $E(\tilde{M}_n(\omega)) = M(\omega)$ and

$$\mathrm{Var}(\tilde{M}_n(\omega)) \le n^{-1}\mathrm{diam}^2(\Omega)E(s_i^2) \le 2n^{-1}\mathrm{diam}^2(\Omega)(1+(x-\mu)^T\Sigma^{-1}(x-\mu)),$$

so $\tilde{M}_n(\omega) - M(\omega) = o_p(1)$. Also, setting

$$W_{0n} := W_{0n}(x) = \bar{X}\Sigma^{-1}(x - \bar{X}) - \mu^T\Sigma^{-1}(x - \mu), \tag{S.2}$$
$$W_{1n} := W_{1n}(x) = \Sigma^{-1}(x - \mu) - \hat{\Sigma}^{-1}(x - \bar{X}),$$

we have $s_{in} - s_i = W_{0n} + W_{1n}^T X_i$. Then

$$M_n(\omega) - \tilde{M}_n(\omega) = \frac{W_{0n}}{n}\sum_{i=1}^n d^2(Y_i, \omega) + \frac{W_{1n}^T}{n}\sum_{i=1}^n X_i d^2(Y_i, \omega) = o_p(1)$$

for all $\omega \in \Omega$, since W_{0n} and $\|W_{1n}\|_E$ are both $O_p(n^{-1/2})$. Using the triangle inequality, we have proven i). Hence, for any $k \in \mathcal{N}$ and $\omega_1, \ldots, \omega_k \in \Omega$, we have $(M_n(\omega_1), \ldots, M_n(\omega_k)) \rightsquigarrow (M(\omega_1), \ldots, M(\omega_k))$.

Moving on to ii), for any γ_1, $\gamma_2 \in \Omega$,

$$\begin{aligned}
|M_n(\gamma_1) - M_n(\gamma_2)| &\leq \frac{1}{n}\sum_{i=1}^n |s_{in}||d(\omega_i, \gamma_1) - d(\omega_i, \gamma_2)||d(\omega_i, \gamma_1) + d(\omega_i, \gamma_2)| \\
&\leq 2\text{diam}(\Omega) d(\gamma_1, \gamma_2)\left(\frac{1}{n}\sum_{i=1}^n |s_i + W_{0n} + W_{1n}^T X_i|\right) \\
&= O_p(d(\gamma_1, \gamma_2)),
\end{aligned}$$

where the O_p term is independent of γ_1 and γ_2. Hence,

$$\sup_{d(\omega_1,\omega_2)<\delta} |M_n(\omega_1) - M_n(\omega_2)| = O_p(\delta),$$

which proves ii). This shows that $d(m_\oplus(x), \hat{m}_\oplus(x)) = o_p(1)$.

For the uniform result, consider the process $Z_n(x) = d(\hat{m}_\oplus(x), m_\oplus(x))$, so $Z_n(x) = o_p(1)$ for any $x \in \mathcal{R}^p$. By Theorem 1.5.4 in van der Vaart and Wellner (1996), it suffices to show that, for any $S > 0$ and as $\delta \to 0$,

$$\limsup_{n\to\infty} \quad P\left(\sup_{\substack{\|x-y\|_E<\delta \\ \|x\|_E, \|y\|_E \leq B}} |Z_n(x) - Z_n(y)| > 2S\right) \to 0.$$

Because $|Z_n(x) - Z_n(y)| \leq d(m_\oplus(x), m_\oplus(y)) + d(\hat{m}_\oplus(x), \hat{m}_\oplus(y))$, it suffices to show that $m_\oplus(\cdot)$ is uniformly continuous for $\|x\|_E \leq B$ and that, as $\delta \to 0$,

$$\limsup_{n\to\infty} P\left(\sup_{\substack{\|x-y\|_E<\delta \\ \|x\|_E, \|y\|_E \leq B}} d(\hat{m}_\oplus(x), \hat{m}_\oplus(y)) > S\right) \to 0. \tag{S.3}$$

Let $\delta > 0$ and x, $y \in \mathcal{R}^p$ with $\|x - y\|_E < \delta$. From the form of M, it is clear that $\sup_{\omega\in\Omega} |M(\omega, x) - M(\omega, y)| \to 0$ as $\delta \to 0$. Assumption (U0) then implies that $m_\oplus$ is continuous at x, and thus uniformly continuous over $\|x\|_E \leq B$. To show (S.3), let $\varepsilon > 0$ and suppose $d(\hat{m}_\oplus(x), \hat{m}_\oplus(y)) > \varepsilon$ with $\|x\|_E$, $\|y\|_E \leq B$. Then (U0) and the form of M_n imply that

$$\zeta \leq \sup_{\substack{\|x-y\|_E<\delta \\ \|x\|_E,\|y\|_E\leq B}} \sup_{\omega\in\Omega} |M_n(\omega, x) - M_n(\omega, y)| = O_p(\delta),$$

and the result follows when $\delta \to 0$. □

Proof of Theorem 2. Let $x \in \mathcal{R}^p$ being fixed and write $m_\oplus = m_\oplus(x)$. We follow the proof of Theorem 3.2.5 in van der Vaart and Wellner (1996) with a few modifications. A key component of this proof is the process $V_n(\omega) = M_n(\omega) - M(\omega)$. Let $D_i(\omega) = d^2(Y_i, \omega) - d^2(Y_i, m_\oplus)$ and s_i be as in (S.1). Then

$$|V_n(\omega) - V_n(m_\oplus)| \leq \left|\frac{1}{n}\sum_{i=1}^n (s_{in} - s_i) D_i(\omega)\right| + \left|\frac{1}{n}\sum_{i=1}^n \left(s_i D_i(\omega) - E(s_i D_i(\omega))\right)\right|. \tag{S.4}$$

This quantity needs to be controlled for small $d(\omega, m_\oplus)$. First, let W_{0n} and W_{1n} be as defined in (S.2). To control the first term on the right-hand side of (S.4), observe that

$$\sup_{d(\omega,m_\oplus)<\delta} \left|\frac{1}{n}\sum_{i=1}^n (s_{in} - s_i) D_i(\omega, x)\right| \leq \frac{2\mathrm{diam}(\Omega)\delta}{n} \sum_{i=1}^n |W_{0n}(x) + W_{1n}(x)^T X_i|,$$

so that the left hand side is $O_p(\delta n^{-1/2})$. Using this fact, we can define

$$B_R = \left\{ \sup_{d(\omega,m_\oplus)<\delta} \left|\frac{1}{n}\sum_{i=1}^n (s_{in} - s_i)\, D_i(\omega, x)\right| \leq R\delta n^{-1/2} \right\}$$

for $R > 0$, so that $P(B_R^c) \to 0$ as $R \to \infty$.

Next, to control the second term on the right-hand side of (S.4) uniformly over small $d(\omega, m_\oplus)$, define the functions $g_\omega : \mathcal{R}^p \times \Omega \to \mathcal{R}$ as

$$g_\omega(z, y) = \left[1 + (z - \mu)^T \Sigma^{-1} (x - \mu)\right] d^2(y, \omega)$$

and the function class

$$\mathcal{M}_\delta := \{g_\omega - g_{m_\oplus} : \ d(\omega, m_\oplus) < \delta\}.$$

An envelope function for $\mathcal{M}_\delta$ is $G_\delta(z) = 2\text{diam}(\Omega)\delta|1+(z-\mu)^T\Sigma^{-1}(x-\mu)^T|$, and $E(G_\delta(X)^2) = O(\delta^2)$. Define $J = J(\delta)$ to be the entropy integral given in (P1), so that $J = O(1)$ as $\delta \to 0$. Then, Theorems 2.7.11 and 2.14.2 of van der Vaart and Wellner (1996) and (P1) imply that, for small enough δ,

(S.5)
$$E\left(\sup_{d(\omega,m_\oplus)<\delta}\left|\frac{1}{n}\sum_{i=1}^{n}(s_iD_i(\omega,x) - E(s_iD_i(\omega,x)))\right|\right) \leq \frac{J\left[E(G_\delta(X)^2)\right]^{1/2}}{\sqrt{n}},$$

so that the left-hand side is $O(\delta n^{-1/2})$. Hence, combining (S.4), (S.5) and the definition of B_R, for small δ,

$$E\left(I_{B_R}\sup_{d(\omega,m_\oplus)<\delta}|V_n(\omega) - V_n(m_\oplus)|\right) \leq a\delta n^{-1/2},$$

for some $a > 0$.

To finish, set $r_n = n^{\frac{\beta}{4(\beta-1)}}$ and

$$S_{j,n}(x) = \{\omega : 2^{j-1} < r_n d(\omega, m_\oplus(x))^{\beta/2} \leq 2^j\}.$$

Choose $\eta > 0$ to satisfy (P2) and also small enough that (P1) holds for all $\delta < \eta$ and set $\tilde{\eta} := \eta^{\beta/2}$. For any integer L,

(S.6)
$$P\left(r_n d(\hat{m}_\oplus, m_\oplus)^{\beta/2} > 2^L\right) \leq P(B_R^c) + P(2d(\hat{m}_\oplus, m_\oplus) \geq \eta) + \sum_{\substack{j\geq L \\ 2^j \leq r_n\tilde{\eta}}} P\left(\left\{\sup_{\omega\in S_{j,n}}|V_n(\omega) - V_n(m_\oplus)| \geq C\frac{2^{2(j-1)}}{r_n^2}\right\} \cap B_R\right),$$

where $P(B_R^c) \to 0$ as discussed previously and the second term goes to zero by Lemma 1. For each j in the sum on the right-hand side of (S.6), we have $d(\omega, m_\oplus) \leq \left(\frac{2^j}{r_n}\right)^{2/\beta} \leq \eta$, so this sum is bounded by

$$4aC^{-1}\sum_{\substack{j\geq L \\ 2^j \leq r_n\tilde{\eta}}}\frac{2^{2j(1-\beta)/\beta}}{r_n^{2(1-\beta)/\beta}\sqrt{n}} \leq 4aC^{-1}\sum_{j\geq L}\left(\frac{1}{4^{(\beta-1)/\beta}}\right)^j.$$

Because $\beta > 1$, the last series converges and hence this probability can be made small by choosing L large. This proves the desired result that $d(\hat{m}_\oplus, m_\oplus) = O_p\left(r_n^{-2/\beta}\right) = O_p\left(n^{-\frac{1}{2(\beta-1)}}\right)$.

For the uniform result over $\|x\|_E \leq B$, use the fact that $W_{0n}(x)$ and $\|W_{1n}(x)\|_E$ are both $O_p(n^{-1/2})$, uniformly over $\|x\|_E \leq B$. Then

$$\sup_{\|x\|_E \leq B} \sup_{d(\omega, m_\oplus(x)) < \delta} \left| \frac{1}{n} \sum_{i=1}^{n} (s_{in}(x) - s_i(x)) D_i(\omega, x) \right| = O_p(\delta n^{-1/2}). \tag{S.7}$$

Then, define

$$A_R = \left\{ \sup_{\|x\|_E \leq B} \sup_{d(\omega, m_\oplus(x)) < \delta} \left| \frac{1}{n} \sum_{i=1}^{n} (s_{in}(x) - s_i(x)) D_i(\omega, x) \right| \leq R\delta n^{-1/2} \right\}$$

for $R > 0$, so $P(A_R^c) \to 0$. Using the definition of $s_i(x)$ in (S.1), we can bound the second term on the right-hand side of (S.4) by

$$\begin{aligned} \|\Sigma^{-1}(x-\mu)\|_E \sum_{j=1}^{p} \left| \frac{1}{n} \sum_{i=1}^{n} (X_{ij} - \mu_j) D_i(\omega, x) - E((X_{ij} - \mu_j) D_i(\omega, x)) \right| \\ + \left| \frac{1}{n} \sum_{i=1}^{n} [D_i(\omega, x) - E(D_i(\omega, x))] \right|. \end{aligned}$$

For $\delta > 0$, set $U_\delta = \{(x, \omega) : \|x\|_E \leq B, d(m_\oplus(x), \omega) < \delta\}$. Next, set

$$h^j_{x,\omega}(z, y) = \begin{cases} d^2(m_\oplus(x), y) - d^2(\omega, y), & j = 0, \\ (z_j - \mu_j) \left[d^2(m_\oplus(x), y) - d^2(\omega, y) \right], & j = 1, \ldots p, \end{cases}$$

and define the classes of functions $\mathcal{N}^j_\delta = \{h^j_{x,\omega} : (x, \omega) \in U_\delta\}$. Assumption (U2) can be used to show, for small $\|x_1 - x_2\|_E$, there is $L_B > 1$ such that

$$d(m_\oplus(x_1), m_\oplus(x_2)) \leq L_B \|x_1 - x_2\|_E^{2/\alpha}.$$

Then $\mathcal{N}^j_\delta$ are Lipschitz classes for small δ in the sense that

$$|h^j_{x_1,\omega_1}(z, y) - h^j_{x_2,\omega_2}(z, y)| \leq C_j \left[\|x_1 - x_2\|_E^{2/\alpha} + d(\omega_1, \omega_2) \right],$$

where $C_j = 2L_B \text{diam}(\Omega)$ if $j = 0$, and $C_j = 2L_B \text{diam}(\Omega)|z_j - \mu_j|$ otherwise. For $\epsilon > 0$, following van der Vaart and Wellner Theorem 2.7.11, the $\delta\epsilon$ bracketing numbers of these classes are all bounded by a multiple of

$$(\epsilon\delta)^{-m} \sup_{\|x\| \leq B} N(c\delta\epsilon, B_\delta(m_\oplus(x)), d),$$

where c and m depend on the dimension p and α only. Then, letting J be the integral in (U1),

$$\begin{aligned}\tilde{J}(\delta) &= \int_0^1 \sqrt{1+\log N_{[]}(\delta\epsilon, \mathcal{N}_\delta, \|\cdot\|_2)}\, d\epsilon = O\left(J + \int_0^1 \sqrt{-\log(\delta\epsilon)}\, d\epsilon\right) \\ &= O(-\log\delta)\end{aligned}$$

as $\delta \to 0$.

Now, $H_\delta^0(z) = 2\text{diam}(\Omega)\delta$ and $H^j\delta(z) = 2\text{diam}(\Omega)\delta|z_j - \mu_j|$ are envelopes for $\mathcal{N}_\delta^j$, and $E\left((H_\delta^j(X))^2\right)$ are all $O(\delta^2)$. Theorem 2.14.2 of van der Vaart and Wellner (1996) provides the bound

$$\begin{aligned}(\text{S.8}) \quad E\left(\sup_{\|x\|_E\le B}\sup_{d(\omega,m_\oplus(x))<\delta}\left|\frac{1}{n}\sum_{i=1}^n (s_i(x)D_i(\omega,x) - E(s_i(x)D_i(\omega,x)))\right|\right) \\ = O\left(\frac{\tilde{J}(\delta)\delta}{\sqrt{n}}\right) = O\left(\frac{\delta^\gamma}{\sqrt{n}}\right)\end{aligned}$$

for any $\gamma < 1$. By combining (S.4), (S.7) and (S.8), for small δ and any $\gamma < 1$,

$$E\left(1_{A_R}\sup_{\|x\|_E\le B}\sup_{d(\omega,m_\oplus(x))<\delta}|V_n(\omega) - V_n(m_\oplus(x))|\right) \le \frac{b\delta^\gamma}{\sqrt{n}}$$

for some constant $b = b(\gamma)$.

To finish, for any $\alpha' > \alpha$, set $\gamma = 1 + \alpha - \alpha'$ and $q_n = n^{\frac{\alpha'}{4(\alpha'-1)}}$. Following the pointwise rate argument, one can show that

$$\sup_{\|x\|_E\le B} d(\hat{m}_\oplus(x), m_\oplus(x))^{\alpha'/2} = O_p(q_n^{-1}),$$

so that

$$\sup_{\|x\|_E\le B} d(\hat{m}_\oplus(x), m_\oplus(x)) = O_p(q_n^{-2/\alpha'}) = O_p\left(n^{-\frac{1}{2(\alpha'-1)}}\right).$$

□

S.3. Proofs of results in Section 4. For completeness, we include the elementary results of auxiliary Lemma 1 and its proof, which are well-known (Fan and Gijbels 1996). The quantities of interest are $\mu_j = E\left(K_h(X-x)(X-x)^j\right)$, $\tau_j(y) = E\left(K_h(X-x)(X-x)^j|Y=y\right)$ and the estimators $\hat{\mu}_j = \frac{1}{n}\sum_{i=1}^n K_h(X_i - x)(X_i - x)^j$, for $j = 0, 1, 2$.

LEMMA 1. *Suppose (K0) and (L1) hold. Then,*

$$\mu_j = h^j \left[f(x) K_{1j} + h f'(x) K_{1(j+1)} + O(h^2) \right]$$

and $\hat{\mu}_j = \mu_j + O_p((h^{2j-1} n^{-1})^{1/2})$ *for* $j = 0, 1, 2$. *Additionally,*

$$\tau_j(y) = h^j \left[g_y(x) K_{1j} + h g_y'(x) K_{1(j+1)} + O(h^2) \right],$$

where the $O(h^2)$ *term is uniform over* $y \in \Omega$.

PROOF. The statements regarding μ_j and $\tau_j(y)$ follow from (K0) and (L1) using a second-order Taylor expansions of the densities f and g_y. Furthermore, $E(\hat{\mu}_j) = \mu_j$ is clear. Next,

$$E\left(K_h^2(X_i - x)(X_i - x)^{2j}\right) = h^{2j-1} \int K(u) u^{2j} f(x + hu)\, du = O(h^{2j-1}),$$

so $\mathrm{Var}(\hat{\mu}_j) = O(h^{2j-1} n^{-1})$, proving the result for the $\hat{\mu}_j$. □

PROOF OF THEOREM 3. First, we will show that $\mathrm{d}F_{Y|X}(x, y)/\mathrm{d}F_Y(y) = g_y(x)/f(x)$ for all x such that $f(x) > 0$. For any open set $U \subset \Omega$, set

$$a(x) = \int_U \frac{g_y(x)}{f(x)} \mathrm{d}F_Y(y), \quad b(x) = \int_U \mathrm{d}F_{Y|X}(x, y).$$

By assumption, both a and b are continuous. Then, for any $z \in \mathcal{R}$,

$$\begin{aligned}
\int_{-\infty}^{z} a(x) f(x)\, dx &= \int_U \left(\int_{-\infty}^{z} g_y(x)\, dx \right) \mathrm{d}F_Y(y) \\
&= \int_U \left(\int_{-\infty}^{z} \mathrm{d}F_{X|Y}(x, y) \right) \mathrm{d}F_Y(y) = \int_{(-\infty, z) \times U} \mathrm{d}F(x, y) \\
&= \int_{-\infty}^{z} \left(\int_U \mathrm{d}F_{Y|X}(x, y) \right) f(x)\, dx = \int_{-\infty}^{z} b(x) f(x)\, dx,
\end{aligned}$$

proving the claim.

Next, using Lemma 1

$$\int s(z, x, h) \mathrm{d}F_{X|Y}(z|y) = \frac{\mu_2 \tau_0(y) - \mu_1 \tau_1(y)}{\sigma_0^2} = \frac{g_y(x)}{f(x)} + O(h^2),$$

where the error term is uniform over $y \in \Omega$. Hence, using the previously established fact that $\mathrm{d}F_{Y|X}(x, y)/\mathrm{d}F_Y(y) = g_y(x)/f(x)$,

$$\begin{aligned}
\tilde{L}_n(\omega) &= \int d^2(y, \omega) s(z, x, h) \mathrm{d}F(z, y) = \int d^2(y, \omega) \frac{g_y(x)}{f(x)} \mathrm{d}F_Y(y) + O(h^2) \\
&= \int d^2(y, \omega) \mathrm{d}F_{Y|X}(x, y) + O(h^2) = M_{\oplus}(\omega, x) + O(h^2),
\end{aligned}$$

where the error term is now uniform over $\omega \in \Omega$. By (L0), we then have $d(m_\oplus(x), \tilde{l}_\oplus(x)) = o(1)$ as $h = h_n \to 0$.

Next, define $r_h = h^{-\frac{\beta_1}{\beta_1 - 1}}$ and set $S_{j,n} = \{\omega : 2^{j-1} < r_h d(\omega, m_\oplus(x))^{\beta_1/2} \leq 2^j\}$. Let I denote the indicator function. Then, for any $M > 0$, following similar arguments as the proof of Theorem 2 and using (L2), there exists $a > 0$ such that, for large n,

$$\begin{aligned} I\left(r_h d(\tilde{l}_\oplus(x), m_\oplus(x))^{\beta_1/2} > 2^M\right) &\leq a \sum_{j \geq M} \frac{2^{2j(1-\beta_1)/\beta}}{r_h^{2(1-\beta_1)/\beta_1} h^{-2}} \\ &\leq a \sum_{j \geq M} \left(\frac{1}{4^{(\beta_1 - 1)/\beta_1}}\right)^j, \end{aligned}$$

which converges since $\beta_1 > 1$. Thus, for some $M > 0$, we have

$$d(\tilde{l}_\oplus(x), m_\oplus(x)) \leq 2^{2M/\beta_1} h^{2/(\beta_1 - 1)}$$

for large n. □

LEMMA 2. *Suppose (K0) and (L0) hold, Ω is bounded and that $h \to 0$ and $nh \to \infty$. Then $d(\tilde{l}_\oplus(x), \hat{l}_\oplus(x)) = o_p(1)$.*

PROOF. We will show that $\tilde{L}_n - \hat{L}_n \rightsquigarrow 0$ in $l^\infty(\Omega)$. Together with (L0), this will prove the result.

To begin, write $s_i(x,h) = \sigma_0^{-2} K_h(X_i - x)\,[\mu_2 - \mu_1(X_i - x)]$. Then the difference $\hat{L}_n(\omega) - \tilde{L}_n(\omega)$ can be written as

$$\begin{aligned} \text{(S.9)} \qquad & \frac{1}{n}\sum_{i=1}^n \left[s_{in}(x,h) - s_i(x,h)\right] d^2(Y_i, \omega) \\ & + \frac{1}{n}\sum_{i=1}^n \left(s_i(x,h) d^2(Y_i,\omega) - E\left[s_i(x,h) d^2(Y_i,\omega)\right]\right). \end{aligned}$$

Observe that $s_{in}(x,h) - s_i(x,h) = W_{0n} K_h(X_i - x) + W_{1n} K_h(X_i - x)(X_i - x)$, where

$$\text{(S.10)} \qquad W_{0n} = \frac{\hat{\mu}_2}{\hat{\sigma}_0^2} - \frac{\mu_2}{\sigma_0^2}, \quad W_{1n} = \frac{\hat{\mu}_1}{\hat{\sigma}_0^2} - \frac{\mu_1}{\sigma_0^2}$$

Using the results of Lemma 1, it follows that $W_{0n} = O_p((nh)^{-1/2})$ and $W_{1n} = O_p((nh^3)^{-1/2})$. Since

$$\begin{aligned} E\left[K_h(X_i - x)(X_i - x)^j d^2(Y_i, \omega)\right] &= O(h^j) \\ E\left[K_h^2(X_i - x)(X_i - x)^{2j} d^4(Y_i, \omega)\right] &= O(h^{2j-1}) \end{aligned}$$

it follows that the first term in (S.9) is $O_p((nh)^{-1/2})$. One also finds that $E(s_i^2(x,h)) = O(h^{-1})$, so that the second term in (S.9) is also $O_p((nh)^{-1/2})$

So far, we have shown that $\tilde{L}_n(\omega) - \hat{L}_n(\omega) = o_p(1)$ for any $\omega \in \Omega$, since $nh \to \infty$. According to Theorem 1.5.4 in van der Vaart and Wellner (1996), the last thing we need to show is that, for any $\eta > 0$

$$\limsup_n P\left(\sup_{d(\omega_1,\omega_2)<\delta} |(\tilde{L}_n - \hat{L}_n)(\omega_1) - (\tilde{L}_n - \hat{L}_n)(\omega_2)| > \eta\right) \to 0 \quad \text{as } \delta \to 0.$$

Since $E(|s_i(x,h)|) = O(1)$ and $E(s_i^2(x,h)) = O(h^{-1})$, $n^{-1}\sum_{i=1}^n |s_{in}(x,h)| = O_p(1)$. Then, $|\hat{L}_n(\omega_1) - \hat{L}_n(\omega_2)| \leq 2\text{diam}(\Omega)d(\omega_1,\omega_2)n^{-1}\sum_{i=1}^n |s_{in}(x,h)| = O_p(d(\omega_1,\omega_2))$. Similarly, $|\tilde{L}_n(\omega_1) - \tilde{L}_n(\omega_2)| = O(d(\omega_1,\omega_2))$, which verifies the above.

□

Proof of Theorem 4. We adopt similar arguments as in the proof of Theorem 2, with some adjustments. Set $s_i(x,h) = K_h(X_i - x)\frac{\mu_0 - \mu_1(X_i - x)}{\sigma_0^2}$ and define $T_n(\omega) = \hat{L}_n(\omega) - \tilde{L}_n(\omega)$. Letting

$$D_i(\omega, x) = d^2(Y_i, \omega) - d^2(Y_i, \tilde{l}_\oplus(x)),$$

we have

(S.11)

$$|T_n(\omega) - T_n(\tilde{l}_\oplus(x))| \leq \left|\frac{1}{n}\sum_{i=1}^n [s_{in}(x,h) - s_i(x,h)]\, D_i(\omega,x)\right| + \left|\frac{1}{n}\sum_{i=1}^n (s_i(x,h)D_i - E\left[s_i(x,h)D_i(\omega,x)\right])\right|.$$

Since W_{0n} and W_{1n} from (S.10) are $O_p((nh)^{-1/2})$ and $O_p((nh^3)^{-1/2})$, respectively, and using the fact that $|D_i(\omega,x)| \leq 2\text{diam}(\Omega)d(\omega,\tilde{l}_\oplus(x))$,the first term on the right-hand side of (S.11) is $O_p(d(\omega,\tilde{l}_\oplus(x)))$, where the O_p term is independent of ω and $\tilde{l}_\oplus(x)$. Thus, we can define

$$B_R = \left\{\sup_{d(\omega,\tilde{l}_\oplus(x))<\delta} \left|\frac{1}{n}\sum_{i=1}^n [s_{in}(x,h) - s_i(x,h)]\, D_i(\omega,x)\right| \leq R\delta(nh)^{-1/2}\right\}$$

for $R > 0$, so that $P(B_R^c) \to 0$.

Next, to control the second term on the right-hand side of (S.11), define the functions $g_\omega : \mathcal{R} \times \Omega \to \mathcal{R}$ by

$$g_\omega(z,y) = \frac{1}{\sigma_0^2}K_h(z-x)[\mu_2 - \mu_1(z-x)]d^2(y,\omega)$$

and the corresponding function class

$$\mathcal{M}_{n\delta} = \{g_\omega - g_{\tilde{l}_\oplus(x)} : \ d(\omega, \tilde{l}_\oplus(x)) < \delta\}.$$

An envelope function for $\mathcal{M}_{n\delta}$ is

$$G_{n\delta}(z) = \frac{2\text{diam}(\Omega)\delta}{\sigma_0^2} K_h(z-x) \left|\mu_2 - \mu_1(z-x)\right|,$$

and $E(G_{n\delta}^2(X)) = O(\delta^2 h^{-1})$. Using this fact together with Theorems 2.7.11 and 2.14.2 of van der Vaart and Wellner (1996) and (P1), for small δ,

$$E\left(\sup_{d(\omega,\tilde{l}_\oplus(x))<\delta} \left|\frac{1}{n} s_i(x,h) D_i(\omega,x) - E\left[s_i(x,h) D_i(\omega,x)\right]\right|\right) = O(\delta(nh)^{-1/2}).$$

Combining this with (S.11) and the definition of B_R,

$$E\left(I_{B_R} \sup_{d(\omega,\tilde{l}_\oplus(x))<\delta} |T_n(\omega) - T_n(\tilde{l}_\oplus(x))|\right) \leq \frac{a\delta}{(nh)^{1/2}},$$

where I_{B_R} is the indicator function for the set B_R and a is a constant depending on R and the entropy integral in (P1).

To finish, set $t_n = (nh)^{\frac{\beta_2}{4(\beta_2-1)}}$ and define

$$S_{j,n}(x) = \{\omega : \ 2^{j-1} < t_n d(\omega, \tilde{l}_\oplus(x))^{\beta_2/2} \leq 2^j\}.$$

Choose η_2 satisfying (L2) and such that (P1) is satisfied for any $\delta < \eta_2$. Set $\tilde{\eta} := (\eta_2/2)^{\beta_2/2}$. For any integer M,

(S.12)

$$\begin{aligned} P\left(t_n d(\tilde{l}_\oplus(x), \hat{l}_\oplus(x))^{\beta/2} > 2^M\right) &\leq P(B_R^c) + P(2d(\tilde{l}_\oplus(x), \hat{l}_\oplus(x)) > \eta) \\ &+ \sum_{\substack{j \geq M \\ 2^j \leq t_n\tilde{\eta}}} P\left(\left\{\sup_{\omega \in S_{j,n}} |T_n(\omega) - T_n(\tilde{l}_\oplus(x))| \geq C\frac{2^{2(j-1)}}{t_n^2}\right\} \cap B_R\right), \end{aligned}$$

where the last term goes to zero for any $\eta > 0$ by Lemma 2. Since

$$d(\omega, \tilde{l}_\oplus(x)) < (2^j/t_n)^{2/\beta_2}$$

on $S_{j,n}(x)$, this implies that the sum on the right-hand side of (S.12) is bounded by

$$4aC^{-1} \sum_{\substack{j \geq M \\ 2^j \leq t_n\tilde{\eta}}} \frac{2^{2j(1-\beta_2)/\beta_2}}{t_n^{2(1-\beta_2)/\beta_2}\sqrt{nh}} \leq 4aC^{-1} \sum_{j \geq M} \left(\frac{1}{4^{(\beta_2-1)/\beta_2}}\right)^j,$$

which converges since $\beta_2 > 1$. Hence,

$$d(\hat{l}_{\oplus}(x), \tilde{l}_{\oplus}(x)) = O_p(t_n^{2/\beta_2}) = O_p\left[(nh)^{-\frac{1}{2(\beta_2-1)}}\right].$$

□

The proof of Corollary 1 is straightforward and is omitted.

S.4. Proofs of results in Section 5.

Proof of Theorem 5. Recall the notation introduced in Section 5. Observe that, when ω ranges over Ω, the object $E\langle Y, \omega\rangle$ is a continuous linear operator under the assumption $E\|Y\|_\Omega^2 < \infty$, so the existence and uniqueness of γ_0 follows by the Riesz representation theorem. The same is true for the operator $E\langle (X-\mu)Y, \alpha\rangle_p$, hence the existence and uniqueness of γ_1. Next

$$\begin{aligned} E(s(X,x)\langle Y,\omega\rangle) &= E\langle Y,\omega\rangle + E\left[(X-\mu)^T\Sigma^{-1}(x-\mu)\langle Y,\omega\rangle\right] \\ &= \langle\gamma_0,\omega\rangle + E\langle (X-\mu)Y, (x-\mu)^T\Sigma^{-1}\omega\rangle_p \\ &= \langle\beta_0,\omega\rangle + \langle (x-\mu)^T\Sigma^{-1}\gamma_1,\omega\rangle \\ &= \langle\beta_0 + \beta_1^T(x-\mu),\omega\rangle. \end{aligned} \tag{S.13}$$

Set $\tilde{\omega} = \beta_0 + \beta_1^T(x-\mu)$ as in (5.1) and observe that $E(s(X,x)) = 1$. Then, by expanding the square, we have

$$\begin{aligned} M(\omega,x) &= E\left(s(X,x)\|Y-\tilde{\omega}\|_\Omega^2 + 2s(X,x)\langle Y-\tilde{\omega}, \tilde{\omega}-\omega\rangle \right. \\ &\qquad \left. + s(X,x)\|\tilde{\omega}-\omega\|_\Omega^2\right) \\ &= M(\tilde{\omega},x) + 2\Big(E(s(X,x)\langle Y,\tilde{\omega}-\omega\rangle) - \langle\tilde{\omega},\tilde{\omega}-\omega\rangle\Big) + \|\tilde{\omega}-\omega\|_\Omega^2. \end{aligned}$$

Hence, the middle term vanishes using (S.13) and we must have $m_{\oplus}(x) = \tilde{\omega}$. As a weighted least squares problem, the empirical solution to (2.10) is clearly $\hat{m}_{\oplus}(x) = n^{-1}\sum_{i=1}^n s_{in}(x)Y_i$, which gives the proposed solution in (5.2). □

Proof of Theorem 6. First, let $q = p+1$ and define $\beta = (\beta_0, \beta_1^T)^T$ and $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1^T)^T$. By Theorem 1.8.4 in chapter 1.8 of van der Vaart and Wellner (1996), we only need to prove that, for all $\alpha \in \Omega^q$, $\sqrt{n}\langle\hat{\beta}-\beta,\alpha\rangle_q \rightsquigarrow \langle\mathcal{G},\alpha\rangle_q$ for the limiting process $\mathcal{G}$ and that $\sqrt{n}(\hat{\beta}-\beta)$ is asymptotically finite dimensional. The latter condition follows from the fact that $\bar{X}-\mu$ and

$\|\Sigma^{-1} - \hat{\Sigma}^{-1}\|_F$ are $O_p(n^{-1/2})$ and by the assumptions on the moments of $\|Y\|_\Omega$. We will now prove the first condition. This will require the definitions below, for any $m \times p$ matrix A and symmetric $p \times p$ matrix S:

$$\begin{aligned}\text{vec}(A) &= (A_{11}, \ldots, A_{m1}, A_{12}, \ldots, A_{m2}, \ldots, A_{1p}, \ldots, A_{mp})^T, \\ \text{vech}(S) &= (A_{11}, \ldots, A_{p1}, A_{22}, \ldots, A_{p2}, \ldots, A_{p,p-1}, A_{pp}).\end{aligned}$$

Let $\alpha \in \Omega^q$ be fixed. Define the $p \times p$ matrices $W_i = X_i X_i^T$ and η_i with elements $(\eta_i)_{jk} = \langle X_{ij} Y_i, \alpha_{k+1} \rangle$, and the vector $\xi_i \in \mathcal{R}^q$ with elements $\xi_{ij} = \langle Y_i, \alpha_j \rangle$. Also, define the vector $\rho \in \mathcal{R}^q$ with elements $\rho_j = \langle \gamma_0, \alpha_j \rangle$ and the $p \times p$ matrix τ with elements $\tau_{jk} = \langle \gamma_{1j}, \alpha_{k+1} \rangle + \mu_j \rho_{k+1}$. Let

$$Z_i = (X_i^T, \text{vech}(W_i)^T, \xi_i^T, \text{vec}(\eta_i)^T)^T. \tag{S.14}$$

Then, $Z_1, \ldots, Z_n$ are independently and identically distributed with expected value

$$E(Z_i) = \left(\mu^T, \text{vech}(\Sigma + \mu\mu^T)^T, \rho^T, \text{vec}(\tau)^T\right)^T.$$

$$\sqrt{n}\left[\bar{Z} - E(Z_1)\right] \rightsquigarrow \mathcal{N}(0, C_\alpha).$$

Next, for $a \in \mathcal{R}^p$, $c \in \mathcal{R}^q$, G a symmetric $p \times p$ matrix and H a $p \times p$ matrix, define the function

$$g(a, \text{vech}(G), c, \text{vec}(H)) = c_1 + \sum_{j=1}^{p}\sum_{k=1}^{p}\left[\left(G - aa^T\right)^{-1}\right]_{jk}(H_{jk} - a_j c_{k+1}).$$

Then

$$\begin{aligned}g(E(Z_1)) &= \langle \gamma_0, \alpha_1 \rangle + \sum_{j=1}^{p}\sum_{k=1}^{p}(\Sigma^{-1})_{jk}\langle \gamma_{1j}, \alpha_{k+1} \rangle \\ &= \langle \beta_0, \alpha_1 \rangle + \sum_{k=1}^{p}\langle \beta_{1k}, \alpha_{k+1} \rangle = \langle \beta, \alpha \rangle_q\end{aligned}$$

and, similarly, $g(\bar{Z}) = \langle \hat{\beta}, \alpha \rangle_q$. Let l_α be the gradient vector of g evaluated at $E(Z_1)$. The elements of l_α can be computed as follows. Let $\otimes$ denote the Kronecker product, $e_l \in \mathcal{R}^p$ be the vector of zeros with a single 1 in the lth entry, and J^{lm} be the $p \times p$ matrix of zeros with a single 1 in the (l, m)th entry. Set

$$\begin{aligned}\mathcal{A}^l &= \Sigma^{-1}(e_l^T \otimes \mu + \mu^T \otimes e_l)\Sigma^{-1}, \\ \mathcal{B}^{lm} &= -\Sigma^{-1}(J^{lm} + J^{ml} - J^{lm}J^{lm})\Sigma^{-1}.\end{aligned}$$

Let s_l be the lth column of Σ^{-1} and set $\alpha_{-1} = (\alpha_2, \ldots, \alpha_q)^T$. The vector l_α can be formed using the values

$$\begin{aligned}
\frac{\partial g}{\partial a_l}(E(Z_1)) &= \langle \mathcal{A}^l \gamma_1, \alpha_{-1} \rangle_p - \langle \gamma_0, \alpha_{-1}^T s_l \rangle, \quad 1 \le l \le p, \\
\frac{\partial g}{\partial B_{lm}}(E(Z_1)) &= \langle \mathcal{B}^{lm} \gamma_1, \alpha_{-1} \rangle_p, \quad 1 \le l \le m \le p, \\
\frac{\partial g}{\partial c_1}(E(Z_1)) &= 1, \\
\frac{\partial g}{\partial c_l}(E(Z_1)) &= -s_{l-1}^T \mu, \quad 2 \le l \le q, \\
\frac{\partial g}{\partial D_{lm}}(E(Z_1)) &= (\Sigma^{-1})_{lm}, \quad 1 \le l, m \le p.
\end{aligned} \tag{S.15}$$

Then, the δ-method yields

$$\sqrt{n}\langle \hat{\beta} - \beta, \alpha \rangle \rightsquigarrow N(0, l_\alpha^T C_\alpha l_\alpha).$$

□

Proof of Corollary 2. Again, set $q = p + 1$. The first display in the corollary follows since $\sup_{x \in V_B} \|\hat{m}_\oplus(x) - m_\oplus(x)\|_\Omega$ is bounded by

$$\|\hat{\beta}_0 - \beta_0\|_\Omega + (\|\mu\|_E + B)\|\hat{\beta}_1 - \beta_1\|_{\Omega^p} + \|\bar{X} - \mu\|_E \|\hat{\beta}_1\|_\Omega = O_p(n^{-1/2}).$$

For the second result, note that Lemmas 1.5.2, 1.5.3 and Theorem 1.5.4 of van der Vaart and Wellner (1996) can be generalized to the space $l_\Omega^\infty(V_B)$. Then, we need to show that $\mathcal{M}_n$ is asymptotically tight and that, for any finite collection $x_1, \ldots, x_J \subset \mathcal{R}^p$, $(\mathcal{M}_n(x_1), \ldots, \mathcal{M}_n(x_J))$ converges weakly to the corresponding marginals of $\mathcal{M}$.

For simplicity, take $x_1, x_2 \in \mathcal{R}^p$. Similar to the proof of Theorem 5, for fixed $\omega \in \Omega$, define $W_i = X_i X_i^T$, $\xi_i = \langle Y_i, \omega \rangle$ and $\eta_i \in \mathcal{R}^p$ with elements $\eta_{ij} = \langle X_{ij} Y_i, \omega \rangle$. Also, define $\rho = \langle \gamma_0, \omega \rangle$, $\tau \in \mathcal{R}^p$ with elements $\tau_j = \langle \gamma_{1j}, \omega \rangle + \mu_j \rho$, and set $Z_i = (X_i^T, \text{vech}(W_i)^T, \xi_i, \eta_i^T)^T$. Then $Z_1, \ldots, Z_n$ are independent with the same distribution and $E(Z_i) = (\mu^T, \text{vech}(\Sigma + \mu\mu^T)^T, \rho, \tau^T + \rho\mu^T)^T$. Letting $C_\omega = \text{Cov}(Z_i)$, we have

$$\sqrt{n}\left[\bar{Z} - E(Z_1)\right] \rightsquigarrow N(0, C_\omega).$$

For $a, c \in \mathcal{R}^p$, $b \in \mathcal{R}$ and G a $p \times p$ symmetric matrix, define

$$g_k(a, \text{vech}(G), b, c) = b + (x_k - a)^T (G - aa^T)^{-1}(c - ba), \quad k = 1, 2.$$

It is easy to verify that $\hat{m}_{\oplus}(x_k) = g_k(\bar{Z})$ and $m_{\oplus}(x_k) = g_k(E(Z_1))$. Define $r_{\omega,k}$ to be the gradient of g_k evaluated at $E(Z_1)$ and set $R_\omega = (r_{\omega,1}, r_{\omega,2})$. Then the bivariate delta method gives

$$(\mathcal{M}_n(x_1), \mathcal{M}_n(x_2))^T \rightsquigarrow N(0, R_\omega^T C_\omega R_\omega).$$

The process $\mathcal{M}$ is characterized by the distribution of its marginals, as given above.

For tightness, first let $\delta, \varepsilon > 0$ be given, define an orthonormal basis $\{e_j\}_{j=1}^{\infty}$ for Ω and let $\Pi_J(\omega) = \sum_{j=1}^{J} \langle \omega, e_j \rangle e_j$ for any integer J and $\omega \in \Omega$. By combining Theorem 5 and Lemma 1.8.1 of van der Vaart and Wellner (1996), there exists finite J_0 such that, with $\tilde{\mathcal{M}}_n(x) = \Pi_{J_0}(\mathcal{M}_n(x))$,

$$\limsup_n P\left(\|\mathcal{M}_n - \tilde{\mathcal{M}}_n\|_{V_B}^2 > \delta \right) < \varepsilon.$$

Note that $\tilde{\mathcal{M}}_n(x) - \tilde{\mathcal{M}}_n(y) = \sum_{k=1}^{p} \Pi_{J_0}(\hat{\beta}_{1k} - \beta_{1k})(x_k - y_k)$ so that, for any $\eta > 0$,

$$\lim_{\tau \to 0} \limsup_n P\left(\sup_{\substack{\|x-y\|_E < \tau \\ x,y \in V_B}} \|\tilde{\mathcal{M}}_n(x) - \tilde{\mathcal{M}}_n(y)\|_\Omega > \eta \right) \to 0$$

by again combining Theorem 5 with Lemma 1.8.1 of van der Vaart and Wellner (1996). This means that $\tilde{\mathcal{M}}_n$ is tight by Theorem 1.5.7 of van der Vaart and Wellner (1996), since $\tilde{\mathcal{M}}_n(x)$ takes values on the finite-dimensional Euclidean space spanned by the first J_0 basis functions $e_j \in \Omega$. For $A \subset l_\Omega^\infty(V_B)$, define

$$A^\delta = \{g \in l_\Omega^\infty(V_B) : \inf_{a \in A} \|a - g\|_{V_B} < \delta\}.$$

Then there exists a compact set $K \subset l_\Omega^\infty(V_B)$ such that

$$\liminf_n P(\tilde{\mathcal{M}}_n \in K^\delta) \geq 1 - \varepsilon$$

and, hence,

$$\begin{aligned} \liminf_n P(\mathcal{M}_n \in K^{2\delta}) \geq \liminf_n P(\tilde{\mathcal{M}}_n \in K^\delta) \\ - \limsup_n P(\|\mathcal{M}_n - \tilde{\mathcal{M}}_n\|_{V_B} > \delta) \geq 1 - 2\varepsilon, \end{aligned}$$

so $\mathcal{M}_n$ is asymptotically tight. □