SUPPLEMENTARY MATERIAL TO 'EFFICIENT FUNCTIONAL ESTIMATION AND THE SUPER-ORACLE PHENOMENON'

By Thomas B. Berrett[1,a] and Richard J. Samworth[2,b]

[1]Department of Statistics, University of Warwick, [a]tom.berrett@warwick.ac.uk

[2]Statistical Laboratory, University of Cambridge, [b]r.samworth@statslab.cam.ac.uk

This is the supplementary material to Berrett and Samworth (2023), hereafter referred to as the main text.

Throughout our proofs we will use the notation

$$u_{x,s} := \frac{k_X}{mV_d h_{x,f}^{-1}(s)^d} \quad \text{and} \quad v_{x,t} := \frac{k_Y}{nV_d h_{x,g}^{-1}(t)^d}.$$

for $x \in \mathcal{X}$ and $s, t \in (0,1)$. Moreover, since many of our error terms will depend on k_X, k_Y, f, g and ϕ (as well as q, in Theorem 5), we adopt the convention, without further comment, that all of these error bounds hold uniformly over the relevant sets as claimed in the statements of the results. In addition, when we write $a \lesssim b$, we mean that there exists $C > 0$, depending only on the parameters d, ϑ and ξ of the problem, such that $a \leq Cb$. It will be convenient throughout to assume that $m, n \geq 3$.

S1.1. *Proof of Proposition 1.*

Proof of Proposition 1. First, we have that $\mu_\alpha(f) \leq 1$ and $\|f\|_\infty \leq C_{d,a,b}$, and it remains to bound the function $M_{f,\beta}(\cdot)$ for each $\beta > 0$. Writing $g(r) := C_{d,a,b} r^{a-1}(1-r)^{b-1}\mathbb{1}_{\{r \leq 1\}}$, so that $f(x) = g(\|x\|)$ we may see by induction that

$$\sup_{r \in (0,2/3)} r^{-(a-t-1)}|g^{(t)}(r)| < \infty \quad \text{and} \quad \sup_{r \in (1/3,1)} r^{-(b-t-1)}|g^{(t)}(r)| < \infty$$

for any $t \in \mathbb{N}$. Moreover, for any $t \in \mathbb{N}$ and multi-index $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}_0^d$ with $|\boldsymbol{\alpha}| = t$, we have that

$$\sup_{x \in B_0(2/3)} \|x\|^{t-1}\big|\partial^{\boldsymbol{\alpha}}\|x\|\big| < \infty \quad \text{and} \quad \sup_{x \in B_0(1)\setminus B_0(1/3)} \big|\partial^{\boldsymbol{\alpha}}\|x\|\big| < \infty.$$

Using these facts we have that

$$\big|\partial^{\boldsymbol{\alpha}} f(x)\big| \lesssim \frac{f(x)}{\|x\|^t(1-\|x\|)^t}$$

for any $t \in \mathbb{N}$. Now, writing $\underline{\beta} := \lceil \beta \rceil - 1$ and fixing $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}_0^d$ with $|\boldsymbol{\alpha}| = \underline{\beta}$, if $y, z \in B_x\big(\|x\|(1-\|x\|)/8\big)$ then we have for any some w on the line segment between x and y that

$$\begin{aligned}\big|\partial^{\boldsymbol{\alpha}} f(z) - \partial^{\boldsymbol{\alpha}} f(y)\big| &\leq d^{1/2}\|z-y\| \Big\| \Big\| f^{(\underline{\beta}+1)}(w) \Big\| \Big\| \\ &\lesssim \|z-y\|\|w\|^{a-1-(\underline{\beta}+1)}(1-\|w\|)^{b-1-(\underline{\beta}+1)} \\ &\lesssim f(y)\|z-y\|^{\beta-\underline{\beta}} \frac{\|z-y\|^{\underline{\beta}+1-\beta}}{\|x\|^{\underline{\beta}+1}(1-\|x\|)^{\underline{\beta}+1}} \lesssim \frac{f(y)\|z-y\|^{\beta-\underline{\beta}}}{\|x\|^{\beta}(1-\|x\|)^{\beta}}.\end{aligned}$$

It follows that $M_{f,\beta}(x) \lesssim 1/\{\|x\|(1-\|x\|)\}$. Therefore, for any $\lambda \in \big(0, b/(b+d-1)\big)$, we have

$$\begin{aligned}\int_{B_0(1)} f(x)\biggl\{\frac{M_{f,\beta}(x)^d}{f(x)}\biggr\}^{\lambda} dx \\ &\lesssim \int_{B_0(1)} \|x\|^{a-1}(1-\|x\|)^{b-1}\biggl\{\frac{1}{\|x\|^{a+d-1}(1-\|x\|)^{b+d-1}}\biggr\}^{\lambda} dx \\ &= dV_d \int_0^1 r^{a+d-2-\lambda(a+d-1)}(1-r)^{b-1-\lambda(b+d-1)}\,dr < \infty,\end{aligned}$$

as claimed. □

S1.2. *Proof of Proposition 6 on asymptotic bias.* The following general result on the bias of the naive estimator $\widetilde{T}_{m,n}$ yields Proposition 6 as an immediate consequence.

PROPOSITION S1. *Fix $d \in \mathbb{N}$, $\vartheta = (\alpha, \beta, \lambda_1, \lambda_2, C) \in \Theta$ and $\xi = (\kappa_1, \kappa_2, \beta^*, L) \in \Xi$. Let $k_X^{\mathrm{L}} \leq k_X^{\mathrm{U}}, k_Y^{\mathrm{L}} \leq k_Y^{\mathrm{U}}$ be deterministic sequences of positive integers such that $k_X^{\mathrm{L}}/\log m \to \infty$, $k_Y^{\mathrm{L}}/\log n \to \infty$, $k_X^{\mathrm{U}} = O(m^{1-\epsilon})$ and $k_Y^{\mathrm{U}} = O(n^{1-\epsilon})$ for some $\epsilon > 0$. Suppose that $\zeta < 1$. Then for each $i_1, i_2 \in \big[\lceil d/2 \rceil - 1\big]$ and $j_1, j_2 \in \mathbb{N}_0$ with $j_1 + j_2 \leq \lceil (\beta^*-1)/2 \rceil$, we can find $\lambda_{i_1 i_2 j_1 j_2} \equiv \lambda_{i_1 i_2 j_1 j_2}(d, f, g, \phi)$, with the properties that $\lambda_{0,0,0,0} = T(f,g)$,*

$$\sup_{\phi \in \Phi(\xi)} \sup_{(f,g) \in \mathcal{F}_{d,\vartheta}} |\lambda_{i_1 i_2 j_1 j_2}| < \infty,$$

and that, for every $\epsilon > 0$,

$$\begin{aligned}\sup_{\phi \in \Phi(\xi)} \sup_{(f,g) \in \mathcal{F}_{d,\vartheta}} \biggl| \mathbb{E}_{f,g}(\widetilde{T}_{m,n}) - \sum_{i_1,i_2=0}^{\lceil d/2 \rceil - 1} \sum_{j_1,j_2=0}^{\infty} \mathbb{1}_{\{j_1+j_2 \leq \lceil (\beta^*-1)/2 \rceil\}} \frac{\lambda_{i_1 i_2 j_1 j_2}}{k_X^{j_1} k_Y^{j_2}} \Bigl(\frac{k_X}{m}\Bigr)^{\frac{2i_1}{d}} \Bigl(\frac{k_Y}{n}\Bigr)^{\frac{2i_2}{d}} \biggr| \\ = O\biggl(\max\biggl\{ k_X^{-\beta^*/2}, \Bigl(\frac{k_X}{m}\Bigr)^{\frac{2\wedge\beta}{d}\beta^*}, \Bigl(\frac{k_X}{m}\Bigr)^{\beta/d}, \Bigl(\frac{k_X}{m}\Bigr)^{\lambda_1(1-\zeta)-\epsilon}, k_Y^{-\beta^*/2}, \\ \Bigl(\frac{k_Y}{n}\Bigr)^{\frac{2\wedge\beta}{d}\beta^*}, \Bigl(\frac{k_Y}{n}\Bigr)^{\beta/d}, \Bigl(\frac{k_Y}{n}\Bigr)^{\lambda_2(1-\zeta)-\epsilon}, 1/m, 1/n \biggr\}\biggr),\end{aligned}$$

as $m, n \to \infty$, uniformly for $k_X \in \{k_X^{\mathrm{L}}, \ldots, k_X^{\mathrm{U}}\}$ and $k_Y \in \{k_Y^{\mathrm{L}}, \ldots, k_Y^{\mathrm{U}}\}$.

PROOF. Define

$$a_{m,X}^{\pm} := 0 \vee \frac{k_X}{m}\biggl(1 \pm \frac{3\log^{1/2} m}{k_X^{1/2}}\biggr) \wedge 1, \quad a_{n,Y}^{\pm} := 0 \vee \frac{k_Y}{n}\biggl(1 \pm \frac{3\log^{1/2} n}{k_Y^{1/2}}\biggr) \wedge 1,$$

let $\mathcal{I}_{m,X} := [a_{m,X}^-, a_{m,X}^+]$, $\mathcal{I}_{n,Y} := [a_{n,Y}^-, a_{n,Y}^+]$, and set

$$\mathcal{X}_{m,n} := \biggl\{ x \in \mathcal{X} : \frac{f(x)}{M_\beta(x)^d} \geq \frac{k_X \log m}{m}, \frac{g(x)}{M_\beta(x)^d} \geq \frac{k_Y \log n}{n} \biggr\}.$$

To begin our bias calculation, we recall the definitions of $\widehat{f}_{(k_X),i}$ and $\widehat{g}_{(k_Y),i}$ from (13). Observe that, conditionally on X_1, we have $h_{X_1,f}(\|X_j - X_1\|) \sim U[0,1]$ for $j \in \{2, \ldots, n\}$, and it follows that

$$(\widehat{f}_{(k_X),1}, \widehat{g}_{(k_Y),1}) \mid X_1 \overset{d}{=} \biggl(\frac{k_X}{mV_d h_{X_1,f}^{-1}(B_1)^d}, \frac{k_Y}{nV_d h_{X_1,g}^{-1}(B_2)^d}\biggr) \biggm| X_1,$$

where $B_1 \sim \mathrm{Beta}(k_X, m-k_X)$ and $B_2 \sim \mathrm{Beta}(k_Y, n+1-k_Y)$ are independent. Moreover, we may write, for example,

$$\frac{u_{x,s}}{f(x)} - 1 = \frac{k_X}{ms} - 1 + \frac{s}{V_d f(x) h_{x,f}^{-1}(s)^d} - 1 + \Bigl(\frac{k_X}{ms} - 1\Bigr)\Bigl(\frac{s}{V_d f(x) h_{x,f}^{-1}(s)^d} - 1\Bigr), \tag{S1}$$

and use Lemma S4 to expand $V_d f(x) h_{x,f}^{-1}(s)^d/s$ in powers of $s^{2/d}$. Since the $\mathrm{Beta}(k, n-k)$ distribution concentrates around its mean at rate $k^{-1/2}$ in an approximately symmetric way, we will also see later that for every $a \in \mathbb{R}$, we have an asymptotic expansion of the form

$$\Bigl(\frac{n}{k}\Bigr)^a \int_0^1 s^a \Bigl(\frac{k}{ns} - 1\Bigr)^j \mathrm{B}_{k,n-k}(s)\, ds = c_1 k^{-\lceil j/2 \rceil} + c_2 k^{-\lceil j/2\rceil - 1} + \ldots + O(1/n),$$

provided that $k = k_n \to \infty$ and $k/n \to 0$ as $n \to \infty$. These facts mean that for remainder terms $R_1, \ldots, R_4$ to be bounded below and functions $c_{i_1 i_2 j_1 j_2}(x)$ to be specified later we may write

$$\begin{aligned}
\mathbb{E}\widetilde{T}_{m,n} &= \int_{\mathcal{X}} f(x) \int_0^1 \int_0^1 \phi(u_{x,s}, v_{x,t}) \mathrm{B}_{k_X, m-k_X}(s) \mathrm{B}_{k_Y, n+1-k_Y}(t)\, ds\, dt\, dx \\
&= \int_{\mathcal{X}_{m,n}} f(x) \int_{\mathcal{I}_{m,X}} \int_{\mathcal{I}_{n,Y}} \phi(u_{x,s}, v_{x,t}) \mathrm{B}_{k_X, m-k_X}(s) \mathrm{B}_{k_Y, n+1-k_Y}(t)\, ds\, dt\, dx + R_1 \\
&= \sum_{\ell_1,\ell_2=0}^{\infty} \frac{\mathbb{1}_{\{\ell_1+\ell_2 \le \beta^*-1\}}}{\ell_1!\ell_2!} \int_{\mathcal{X}_{m,n}} f(x) \int_{\mathcal{I}_{m,X}} \int_{\mathcal{I}_{n,Y}} \Bigl(\frac{u_{x,s}}{f(x)} - 1\Bigr)^{\ell_1} \Bigl(\frac{v_{x,t}}{g(x)} - 1\Bigr)^{\ell_2} f(x)^{\ell_1} g(x)^{\ell_2} \\
&\qquad \times \phi_{\ell_1 \ell_2}\bigl(f(x), g(x)\bigr) \mathrm{B}_{k_X, m-k_X}(s) \mathrm{B}_{k_Y, n+1-k_Y}(t)\, ds\, dt\, dx + R_1 + R_2 \qquad \text{(S2)} \\
&= \sum_{i_1,i_2=0}^{\lceil d/2\rceil - 1} \sum_{j_1,j_2=0}^{\infty} \mathbb{1}_{\{j_1+j_2 \le \beta^*-1\}} \int_{\mathcal{X}_{m,n}} f(x) c_{i_1 i_2 j_1 j_2}(x) \int_{\mathcal{I}_{m,X}} \int_{\mathcal{I}_{n,Y}} \Bigl(\frac{k_X}{ms} - 1\Bigr)^{j_1} \Bigl(\frac{k_Y}{nt} - 1\Bigr)^{j_2} \\
&\qquad \times s^{\frac{2i_1}{d}} t^{\frac{2i_2}{d}} \mathrm{B}_{k_X, m-k_X}(s) \mathrm{B}_{k_Y, n+1-k_Y}(t)\, ds\, dt\, dx + R_1 + R_2 + R_3 \qquad \text{(S3)} \\
&= \sum_{i_1,i_2=0}^{\lceil d/2\rceil - 1} \sum_{j_1,j_2=0}^{\infty} \mathbb{1}_{\{j_1+j_2 \le \lceil (\beta^*-1)/2 \rceil\}} \frac{\lambda_{i_1 i_2 j_1 j_2}}{k_X^{j_1} k_Y^{j_2}} \Bigl(\frac{k_X}{m}\Bigr)^{\frac{2i_1}{d}} \Bigl(\frac{k_Y}{n}\Bigr)^{\frac{2i_2}{d}} + R_1 + R_2 + R_3 + R_4.
\end{aligned}$$

It now remains to bound each of the remainder terms.

To bound R_1: Since we are assuming that $\zeta < 1$, we may apply Lemma S9 to see that

$$\begin{aligned}
&\int_{\mathcal{X}} f(x) \int_0^1 \int_0^1 \bigl(1 - \mathbb{1}_{\{s \in \mathcal{I}_{m,X}\}} \mathbb{1}_{\{t \in \mathcal{I}_{n,Y}\}}\bigr) \phi(u_{x,s}, v_{x,t}) \\
&\qquad \times \mathrm{B}_{k_X, m-k_X}(s) \mathrm{B}_{k_Y, n+1-k_Y}(t)\, ds\, dt\, dx = o(m^{-4} + n^{-4}). \qquad \text{(S4)}
\end{aligned}$$

When $s \in \mathcal{I}_{m,X}$ and $k_X^{\mathrm{L}} \ge 36 \log m$, we have by Lemma S8 that $u_{x,s} \le \frac{Ck_X}{ma_{m,X}^-} \le 2C$, and, similarly, when $t \in \mathcal{I}_{n,Y}$ and $k_Y^{\mathrm{L}} \ge 36 \log n$, we have $v_{x,t} \le 2C$. Thus, when $s \in \mathcal{I}_{m,X}, t \in \mathcal{I}_{n,Y}$ and $\min(k_X^{\mathrm{L}}/\log m, k_Y^{\mathrm{L}}/\log n) \ge 36$, we may use the fact that $|u_{x,s}^{\ell_1} v_{x,t}^{\ell_2} \phi_{\ell_1,\ell_2}(u_{x,s}, v_{x,t})| \le L(2C)^{2L+|\kappa_1|+|\kappa_2|} u_{x,s}^{\kappa_1} v_{x,t}^{\kappa_2}$ for all $\ell_1, \ell_2 \in \mathbb{N}_0$ such that $\ell_1 + \ell_2 \le \beta^* - 1$.

In the following we consider the decomposition $\mathcal{X}_{m,n}^c = \mathcal{X}_{m,f}^c \cup \mathcal{X}_{n,g}^c$, where $\mathcal{X}_{m,f} := \{x : f(x)M_\beta(x)^{-d} \geq k_X \log m/m\}$ and $\mathcal{X}_{n,g} := \{x : g(x)M_\beta(x)^{-d} \geq k_Y \log n/n\}$. Using Lemma S7 and Lemma S8 we have that

$$\int_{\mathcal{X}_{m,f}^c} f(x) \int_{\mathcal{I}_{m,X}} \int_{\mathcal{I}_{n,Y}} \phi(u_{x,s}, v_{x,t}) \mathrm{B}_{k_X,m-k_X}(s) \mathrm{B}_{k_Y,n+1-k_Y}(t)\, ds\, dt\, dx$$

$$\lesssim \int_{\mathcal{X}_{m,f}^c} f(x)^{1-\kappa_1^-} g(x)^{-\kappa_2^-} M_\beta(x)^{d(\kappa_1^-+\kappa_2^-)} (1+\|x\|)^{d(\kappa_1^-+\kappa_2^-)}\, dx$$

$$\leq \inf_{a>0} \Big(\frac{k_X \log m}{m}\Big)^a \int_{\mathcal{X}} f(x) \frac{M_\beta(x)^{d(a+\kappa_1^-+\kappa_2^-)}}{f(x)^{a+\kappa_1^-} g(x)^{\kappa_2^-}} (1+\|x\|)^{d(\kappa_1^-+\kappa_2^-)}\, dx$$

$$= O\bigg(\Big(\frac{k_X}{m}\Big)^{\lambda_1(1-\zeta)-\epsilon}\bigg) \tag{S5}$$

for every $\epsilon > 0$. With a similar bound over $\mathcal{X}_{n,g}^c$ we conclude that

$$\int_{\mathcal{X}_{m,n}^c} f(x) \int_{\mathcal{I}_{m,X}} \int_{\mathcal{I}_{n,Y}} \phi(u_{x,s}, v_{x,t}) \mathrm{B}_{k_X,m-k_X}(s) \mathrm{B}_{k_Y,n+1-k_Y}(t)\, ds\, dt\, dx$$

$$= O\bigg(\max\Big\{\Big(\frac{k_X}{m}\Big)^{\lambda_1(1-\zeta)-\epsilon}, \Big(\frac{k_Y}{n}\Big)^{\lambda_2(1-\zeta)-\epsilon}\Big\}\bigg) \tag{S6}$$

for every $\epsilon > 0$. From (S4) and (S6), we deduce that

$$R_1 = O\bigg(\max\Big\{\Big(\frac{k_X}{m}\Big)^{\lambda_1(1-\zeta)-\epsilon}, \Big(\frac{k_Y}{n}\Big)^{\lambda_2(1-\zeta)-\epsilon}, \frac{1}{m^4}, \frac{1}{n^4}\Big\}\bigg). \tag{S7}$$

To bound R_2: We first observe that, by (S1) and Lemma S4, we have that

$$\epsilon_{m,n} := \sup_{x \in \mathcal{X}_{m,f}} \sup_{s \in \mathcal{I}_{m,X}} \Big|\frac{u_{x,s}}{f(x)} - 1\Big| \bigvee \sup_{x \in \mathcal{X}_{n,g}} \sup_{t \in \mathcal{I}_{n,Y}} \Big|\frac{v_{x,t}}{g(x)} - 1\Big| = o(1). \tag{S8}$$

Now, for $t \in [0,1]$ we have that $h(t) := t - \log(1+t) \geq t^2/4$. Thus, letting $B \sim \mathrm{Beta}(k, n-k)$, whenever $\frac{3\alpha^{1/2}\log^{1/2} n}{k^{1/2}} \leq 1$ and $\frac{k^{1/2}+3\alpha^{1/2}\log^{1/2} n}{n^{1/2}} \leq 2^{1/2}-1$ we may integrate the Beta tail bound in Lemma S6 to see that

$$\int_0^1 \Big|\frac{ns}{k} - 1\Big|^\alpha \mathrm{B}_{k,n-k}(s)\, ds = \alpha \int_0^{n/k} y^{\alpha-1} \mathbb{P}\Big(\Big|B - \frac{k}{n}\Big| \geq \frac{ky}{n}\Big)\, dy$$

$$\leq 2\alpha k^{-\alpha/2} \int_0^{k^{-1/2}n} u^{\alpha-1} \Big\{\exp\Big(-kh\Big(\frac{n^{1/2}k^{-1/2}u}{n^{1/2}+k^{1/2}+u}\Big)\Big) + \exp\Big(-nh\Big(\frac{u}{n^{1/2}+k^{1/2}+u}\Big)\Big)\Big\}\, du$$

$$\leq 4\alpha k^{-\alpha/2} \int_0^{3\alpha^{1/2}\log^{1/2} n} u^{\alpha-1} e^{-u^2/8}\, du + \frac{4n^\alpha}{k^\alpha} \exp\Big(-\frac{9\alpha \log n}{8}\Big)$$

$$\leq \frac{2^{3(\alpha-1)/2}\alpha\Gamma(\alpha/2)}{k^{\alpha/2}} + \frac{4}{k^\alpha}. \tag{S9}$$

Next, by Lemma S7, we have for any $\tau \geq 0$ that

$$\Big(\frac{k_X}{m}\Big)^\tau \int_{\mathcal{X}_{m,n}} f(x)^{1-\kappa_1^-} g(x)^{-\kappa_2^-} \Big\{\frac{M_\beta(x)^d}{f(x)}\Big\}^\tau\, dx$$

$$\leq \inf_{a>0}\Big(\frac{k_X}{m}\Big)^{\tau-a}\int_{\mathcal{X}} f(x)\Big\{\frac{M_\beta(x)^d}{f(x)}\Big\}^{\tau+\kappa_1^- -a}\Big\{\frac{M_\beta(x)^d}{g(x)}\Big\}^{\kappa_2^-}\,dx$$

$$= O\biggl(\max\biggl\{\Big(\frac{k_X}{m}\Big)^{\tau},\Big(\frac{k_X}{m}\Big)^{\lambda_1(1-\zeta)-\epsilon}\biggr\}\biggr) \tag{S10}$$

for all $\epsilon > 0$. Analogously,

$$\int_{\mathcal{X}_{m,n}} f(x)^{1-\kappa_1^-} g(x)^{-\kappa_2^-}\Big\{\frac{M_\beta(x)^d}{g(x)}\Big\}^{\tau}\,dx = O\biggl(\max\biggl\{1,\Big(\frac{k_Y}{n}\Big)^{\lambda_2(1-\zeta)-\tau-\epsilon}\biggr\}\biggr) \tag{S11}$$

for any $\tau \geq 0$ and $\epsilon > 0$. Now, since $\phi \in \Phi$ and by (S1), (S8), (S9), (S10), (S11) and Lemmas S3(ii) and S4 we have, when m,n are sufficiently large that $\epsilon_{m,n} < 1/2$,

$$|R_2| \lesssim L\int_{\mathcal{X}_{m,n}} f(x)^{1+\kappa_1} g(x)^{\kappa_2}\int_{\mathcal{I}_{m,X}}\int_{\mathcal{I}_{n,Y}}\biggl\{\Big|\frac{u_{x,s}}{f(x)}-1\Big|^{\beta^*}+\Big|\frac{v_{x,t}}{g(x)}-1\Big|^{\beta^*}\biggr\}$$
$$\times \mathrm{B}_{k_X,m-k_X}(s)\mathrm{B}_{k_Y,n+1-k_Y}(t)\,ds\,dt\,dx$$
$$\lesssim \int_{\mathcal{X}_{m,n}} f(x)^{1+\kappa_1} g(x)^{\kappa_2}\int_{\mathcal{I}_{m,X}}\int_{\mathcal{I}_{n,Y}} \mathrm{B}_{k_X,m-k_X}(s)\mathrm{B}_{k_Y,n+1-k_Y}(t)\biggl[\Big|\frac{k_X}{ms}-1\Big|^{\beta^*}$$
$$+\Big|\frac{k_Y}{nt}-1\Big|^{\beta^*}+\Big\{\frac{sM_\beta(x)^d}{f(x)}\Big\}^{\frac{2\wedge\beta}{d}\beta^*}+\Big\{\frac{tM_\beta(x)^d}{g(x)}\Big\}^{\frac{2\wedge\beta}{d}\beta^*}\biggr]\,ds\,dt\,dx$$

$$= O\biggl(\max\biggl\{k_X^{-\beta^*/2},k_Y^{-\beta^*/2},\Big(\frac{k_X}{m}\Big)^{\frac{2\wedge\beta}{d}\beta^*},\Big(\frac{k_X}{m}\Big)^{\lambda_1(1-\zeta)-\epsilon},\Big(\frac{k_Y}{n}\Big)^{\frac{2\wedge\beta}{d}\beta^*}\Big(\frac{k_Y}{n}\Big)^{\lambda_2(1-\zeta)-\epsilon}\biggr\}\biggr). \tag{S12}$$

To bound R_3: By (S1) and Lemma S4, when $\ell_1 > 0$ we have expansions of the form

$$\biggl|\Big(\frac{u_{x,s}}{f(x)}-1\Big)^{\ell_1}-\sum_{i=0}^{\lceil d/2\rceil-1}\sum_{j=0}^{\ell_1} b_{i,j}(x)s^{2i/d}\Big(\frac{k}{ms}-1\Big)^{j}\biggr| \lesssim \Big\{\frac{sM_\beta(x)^d}{f(x)}\Big\}^{\frac{\beta}{d}\wedge 1},$$

with $|b_{i,j}(x)| \lesssim \{M_\beta(x)^d/f(x)\}^{2i/d}$ and $b_{0,0}=0$. A similar expansion can also be written for $(v_{x,t}/g(x)-1)^{\ell_2}$. Using these two expansions it can be seen that $c_{i_1 i_2 j_1 j_2}$ can be chosen in (S2) with $|c_{i_1 i_2 j_1 j_2}(x)| \lesssim f(x)^{\kappa_1}g(x)^{\kappa_2}\{M_\beta(x)^d/f(x)\}^{2i_1/d}\{M_\beta(x)^d/g(x)\}^{2i_2/d}$, with $c_{0,0,0,0}(x) = \phi(f(x),g(x))$, and, using (S10) and (S11), with

$$|R_3| \lesssim \int_{\mathcal{X}_{m,n}} f(x)^{1+\kappa_1} g(x)^{\kappa_2}\biggl\{\Big(\frac{k_X M_\beta(x)^d}{mf(x)}\Big)^{\frac{\beta}{d}\wedge 1}+\Big(\frac{k_Y M_\beta(x)^d}{ng(x)}\Big)^{\frac{\beta}{d}\wedge 1}\biggr\}\,dx$$
$$= O\biggl(\max\biggl\{\Big(\frac{k_X}{m}\Big)^{\frac{\beta}{d}\wedge 1},\Big(\frac{k_X}{m}\Big)^{\lambda_1(1-\zeta)-\epsilon},\Big(\frac{k_Y}{n}\Big)^{\frac{\beta}{d}\wedge 1},\Big(\frac{k_Y}{n}\Big)^{\lambda_2(1-\zeta)-\epsilon}\biggr\}\biggr).$$

To bound R_4: Whenever $a \in \mathbb{R}$ is fixed, we have an asymptotic series of the form

$$\frac{\Gamma(m+a)}{k_X^a\Gamma(m)}\int_0^1 s^a \mathrm{B}_{k_X,m-k_X}(s)\,ds = \frac{\Gamma(k_X+a)}{\Gamma(k_X)k_X^a} = 1 + c_1/k_X + c_2/k_X^2 + \dots. \tag{S13}$$

On the other hand, arguing similarly to (S9), for fixed $j \in \mathbb{N}$ we have the bound

$$\Big(\frac{m}{k_X}\Big)^a\int_0^1 s^a\Big|\frac{k_X}{ms}-1\Big|^j \mathrm{B}_{k_X,m-k_X}(s)\,ds$$

$$\leq \frac{2^{j-1} m^a \Gamma(k_X + a - j)\Gamma(m)}{k_X^a \Gamma(k_X)\Gamma(m+a-j)} \biggl\{ \biggl| \frac{k_X + a - j}{m + a - j} - \frac{k_X}{m} \biggr|^j + \int_0^1 \biggl| s - \frac{k_X + a - j}{m + a - j} \biggr|^j \mathrm{B}_{k_X + a - j, m - k_X}(s)\, ds \biggr\}$$

$$= O(k_X^{-j/2}). \tag{S14}$$

Moreover, by Lemma S6, letting $B \sim \mathrm{Beta}(k_X + a - j, m - k_X)$ we have that

$$\Bigl(\frac{m}{k_X}\Bigr)^a \int_{[0,1]\setminus \mathcal{I}_{m,X}} s^a \biggl| \frac{k_X}{ms} - 1 \biggr|^j \mathrm{B}_{k_X, m-k_X}(s)\, ds$$

$$\lesssim \int_{[0,1]\setminus \mathcal{I}_{m,X}} \biggl| \frac{ms}{k_X} - 1 \biggr|^j \mathrm{B}_{k_X + a - j, m - k_X}(s)\, ds$$

$$\leq \mathbb{P}\biggl(\biggl| \frac{mB}{k_X} - 1 \biggr| \geq \frac{3 \log^{1/2} m}{k_X^{1/2}} \biggr) + \Bigl(\frac{m}{k_X}\Bigr)^j \mathbb{P}\biggl(\biggl| \frac{mB}{k_X} - 1 \biggr| \geq 1 \biggr) = o(m^{-4}). \tag{S15}$$

With the similar expression in terms of k_Y and n, we now conclude from (S13), (S14) and (S15) that we have an asymptotic expansion of the form

$$\int_{\mathcal{I}_{m,X}} \int_{\mathcal{I}_{n,Y}} s^{\frac{2i_1}{d}} t^{\frac{2i_2}{d}} \Bigl(\frac{k_X}{ms} - 1 \Bigr)^{j_1} \Bigl(\frac{k_Y}{nt} - 1 \Bigr)^{j_2} \mathrm{B}_{k_X, m-k_X}(s) \mathrm{B}_{k_Y, n+1-k_Y}(t)\, ds\, dt$$

$$= \Bigl(\frac{k_X}{m}\Bigr)^{\frac{2i_1}{d}} \Bigl(\frac{k_Y}{n}\Bigr)^{\frac{2i_2}{d}} \biggl\{ \sum_{r=\lceil j_1/2 \rceil}^{\infty} c_r k_X^{-r} + O(1/m) \biggr\} \biggl\{ \sum_{r=\lceil j_2/2 \rceil}^{\infty} d_r k_Y^{-r} + O(1/n) \biggr\}. \tag{S16}$$

Now for fixed $i_1, i_2 \in [\lceil d/2 \rceil - 1]$ with $\frac{\kappa_1^- + 2i_1/d}{\lambda_1} + \frac{\kappa_2^- + 2i_2/d}{\lambda_2} \geq 1$, we have by Lemma S7 that

$$\int_{\mathcal{X}_{m,n}} f(x) |c_{i_1 i_2 j_1 j_2}(x)|\, dx \lesssim \int_{\mathcal{X}_{m,n}} f(x) \frac{M_\beta(x)^{2i_1 + 2i_2}}{f(x)^{\kappa_1^- + 2i_1/d} g(x)^{\kappa_2^- + 2i_2/d}}\, dx$$

$$\leq \min \biggl\{ \inf_{a>0} \Bigl(\frac{k_X}{m}\Bigr)^{-a} \int_{\mathcal{X}} f(x) \Bigl\{ \frac{M_\beta(x)^d}{f(x)} \Bigr\}^{\kappa_1^- + \frac{2i_1}{d} - a} \Bigl\{ \frac{M_\beta(x)^d}{g(x)} \Bigr\}^{\kappa_2^- + \frac{2i_2}{d}}\, dx, \inf_{a>0} \Bigl(\frac{k_Y}{n}\Bigr)^{-a} \int_{\mathcal{X}} f(x) \Bigl\{ \frac{M_\beta(x)^d}{f(x)} \Bigr\}^{\kappa_1^- + \frac{2i_1}{d}} \Bigl\{ \frac{M_\beta(x)^d}{g(x)} \Bigr\}^{\kappa_2^- + \frac{2i_2}{d} - a}\, dx \biggr\}$$

$$= O\biggl(\min \biggl\{ \Bigl(\frac{k_X}{m}\Bigr)^{\lambda_1 (1 - \zeta - \frac{2i_2}{d\lambda_2}) - 2i_1/d - \epsilon}, \Bigl(\frac{k_Y}{n}\Bigr)^{\lambda_2 (1 - \zeta - \frac{2i_1}{d\lambda_1}) - 2i_2/d - \epsilon} \biggr\} \biggr)$$

$$= O\biggl(\Bigl(\frac{k_X}{m}\Bigr)^{-\frac{2i_1}{d}} \Bigl(\frac{k_Y}{n}\Bigr)^{-\frac{2i_2}{d}} \max \biggl\{ \Bigl(\frac{k_X}{m}\Bigr)^{\lambda_1(1-\zeta) - \epsilon}, \Bigl(\frac{k_Y}{n}\Bigr)^{\lambda_2(1-\zeta)-\epsilon} \biggr\} \biggr), \tag{S17}$$

for all $\epsilon > 0$, where the final inequality can be established by considering the cases $(\frac{k_X}{m})^{\lambda_1} \geq (\frac{k_Y}{n})^{\lambda_2}$ and $(\frac{k_X}{m})^{\lambda_1} < (\frac{k_Y}{n})^{\lambda_2}$ separately. For such i_1, i_2 we set $\lambda_{i_1 i_2 j_1 j_2} = 0$ for all j_1, j_2. When, instead, $\frac{\kappa_1^- + 2i_1/d}{\lambda_1} + \frac{\kappa_2^- + 2i_2/d}{\lambda_2} < 1$, we again consider these two cases separately, use the decomposition $\mathcal{X}_{m,n}^c = (\mathcal{X}_{m,f}^c \cap \mathcal{X}_{n,g}^c) \cup (\mathcal{X}_{m,f}^c \cap \mathcal{X}_{n,g}) \cup (\mathcal{X}_{m,f} \cap \mathcal{X}_{n,g}^c)$ and apply Lemma S7 to write

$$\int_{\mathcal{X}_{m,n}^c} f(x) |c_{i_1 i_2 j_1 j_2}(x)|\, dx \lesssim \int_{\mathcal{X}_{m,f}^c \cap \mathcal{X}_{n,g}^c} f(x) \frac{M_\beta(x)^{2i_1 + 2i_2}}{f(x)^{\kappa_1^- + 2i_1/d} g(x)^{\kappa_2^- + 2i_2/d}}\, dx\, dx + O\biggl(\Bigl(\frac{k_X}{m}\Bigr)^{-\frac{2i_1}{d}} \Bigl(\frac{k_Y}{n}\Bigr)^{-\frac{2i_2}{d}} \biggl\{ \Bigl(\frac{k_X}{m}\Bigr)^{\lambda_1(1-\zeta)-\epsilon} \vee \Bigl(\frac{k_Y}{n}\Bigr)^{\lambda_2(1-\zeta)-\epsilon} \biggr\} \biggr)$$

$$\leq \min\biggl\{\inf_{a>0}\Bigl(\frac{k_X \log m}{m}\Bigr)^a \int_{\mathcal{X}} f(x)\Bigl\{\frac{M_\beta(x)^d}{f(x)}\Bigr\}^{\kappa_1^- + \frac{2i_1}{d}+a}\Bigl\{\frac{M_\beta(x)^d}{g(x)}\Bigr\}^{\kappa_2^- + \frac{2i_2}{d}}\,dx,$$

$$\inf_{a>0}\Bigl(\frac{k_Y \log n}{n}\Bigr)^a \int_{\mathcal{X}} f(x)\Bigl\{\frac{M_\beta(x)^d}{f(x)}\Bigr\}^{\kappa_1^- + \frac{2i_1}{d}}\Bigl\{\frac{M_\beta(x)^d}{g(x)}\Bigr\}^{\kappa_2^- + \frac{2i_2}{d}+a}\,dx\biggr\}$$

$$+ O\biggl(\Bigl(\frac{k_X}{m}\Bigr)^{-\frac{2i_1}{d}}\Bigl(\frac{k_Y}{n}\Bigr)^{-\frac{2i_2}{d}}\biggl\{\Bigl(\frac{k_X}{m}\Bigr)^{\lambda_1(1-\zeta)-\epsilon}\vee\Bigl(\frac{k_Y}{n}\Bigr)^{\lambda_2(1-\zeta)-\epsilon}\biggr\}\biggr)$$

$$= O\biggl(\Bigl(\frac{k_X}{m}\Bigr)^{-\frac{2i_1}{d}}\Bigl(\frac{k_Y}{n}\Bigr)^{-\frac{2i_2}{d}}\biggl\{\Bigl(\frac{k_X}{m}\Bigr)^{\lambda_1(1-\zeta)-\epsilon}\vee\Bigl(\frac{k_Y}{n}\Bigr)^{\lambda_2(1-\zeta)-\epsilon}\biggr\}\biggr) \tag{S18}$$

for all $\epsilon > 0$. It follows from (S16), (S17) and (S18) that

$$R_4 = O\biggl(\max\biggl\{k_X^{-\lceil(\beta^*-1)/2\rceil-1}, k_Y^{-\lceil(\beta^*-1)/2\rceil-1}, \Bigl(\frac{k_X}{m}\Bigr)^{\lambda_1(1-\zeta)-\epsilon}, \Bigl(\frac{k_Y}{n}\Bigr)^{\lambda_2(1-\zeta)-\epsilon}, \frac{1}{m}, \frac{1}{n}\biggr\}\biggr),$$

and this concludes the proof. □

S1.3. *Proof of Proposition 8 on improved bias bounds.*

PROOF OF PROPOSITION 8. From (S2) in the proof of Proposition S1, we may write

$$\mathbb{E}\widetilde{T}_{m,n}$$
$$= \sum_{\ell_1,\ell_2=0}^{\infty} \frac{\mathbb{1}_{\{\ell_1+\ell_2\leq\beta^*-1\}}}{\ell_1!\ell_2!}\int_{\mathcal{X}_{m,n}} f(x)\int_{\mathcal{I}_{m,X}}\int_{\mathcal{I}_{n,Y}}\Bigl(\frac{u_{x,s}}{f(x)}-1\Bigr)^{\ell_1}\Bigl(\frac{v_{x,t}}{g(x)}-1\Bigr)^{\ell_2} f(x)^{\ell_1}g(x)^{\ell_2}$$
$$\times\, \phi_{\ell_1\ell_2}\bigl(f(x),g(x)\bigr)\mathrm{B}_{k_X,m-k_X}(s)\mathrm{B}_{k_Y,n+1-k_Y}(t)\,ds\,dt\,dx + R_1 + R_2$$

where R_1 and R_2 satisfy the bounds (S7) and (S12) respectively. We may expand $(u_{x,s}/f(x)-1)^{\ell_1}$ using (S1), and also expand and $(v_{x,t}/g(x)-1)^{\ell_2}$ analogously. Any term including $s/\{V_d f(x)h_{x,f}^{-1}(s)^d\}-1$ and $t/\{V_d g(x)h_{x,g}^{-1}(t)^d\}-1$ to a combined power greater than one can be bounded by

$$O\biggl(\max\biggl\{\Bigl(\frac{k_X}{m}\Bigr)^{2\beta/d}, \Bigl(\frac{k_X}{m}\Bigr)^{\lambda_1(1-\zeta)-\epsilon}, \Bigl(\frac{k_Y}{n}\Bigr)^{2\beta/d}, \Bigl(\frac{k_Y}{n}\Bigr)^{\lambda_2(1-\zeta)-\epsilon}\biggr\}\biggr)$$

by Lemma S4, so can absorbed into the error term in (15). The key difference with the proof of Proposition S1 is that for $s \in \mathcal{I}_{m,X}$ we now write

$$\int_{\mathcal{X}_{m,n}} f(x)^2\phi_{10}\bigl(f(x),g(x)\bigr)\bigl\{s - V_d f(x)h_{x,f}^{-1}(s)^d\bigr\}\,dx$$
$$= \int_{\mathcal{X}_{m,n}}\int_{\mathcal{X}_{m,n}} f(x)^2\phi_{10}\bigl(f(x),g(x)\bigr)\mathbb{1}_{\{\|x-y\|\leq h_{x,f}^{-1}(s)\}}\{f(y)-f(x)\}\,dy\,dx + \tilde{R}_1(s), \tag{S19}$$

where

$$\int_{\mathcal{I}_{m,X}} \frac{1}{s}\mathrm{B}_{k_X,m-k_X}(s)|\tilde{R}_1(s)|\,ds = O\biggl(\Bigl(\frac{k_X}{m}\Bigr)^{\lambda_1(1-\zeta)-\epsilon}\biggr).$$

Moreover, by Fubini's theorem,

$$\int_{\mathcal{X}_{m,n}}\int_{\mathcal{X}_{m,n}} f(x)^2\phi_{10}\bigl(f(x),g(x)\bigr)\mathbb{1}_{\{\|x-y\|\leq h_{x,f}^{-1}(s)\}}\{f(y)-f(x)\}\,dy\,dx$$

$$= \frac{1}{2}\int_{\mathcal{X}_{m,n}}\int_{\mathcal{X}_{m,n}}\biggl[f(x)^2\phi_{10}\bigl(f(x),g(x)\bigr)\mathbb{1}_{\{\|x-y\|\leq h_{x,f}^{-1}(s)\}}\{f(y)-f(x)\} + f(y)^2\phi_{10}\bigl(f(y),g(y)\bigr)\mathbb{1}_{\{\|x-y\|\leq h_{y,f}^{-1}(s)\}}\{f(x)-f(y)\}\biggr]\,dy\,dx$$

$$= \frac{1}{2}\int_{\mathcal{X}_{m,n}}\int_{\mathcal{X}_{m,n}}\biggl[\mathbb{1}_{\{\|x-y\|\leq h_{x,f}^{-1}(s)\}}\{f(y)-f(x)\}\bigl\{f(x)^2\phi_{10}\bigl(f(x),g(x)\bigr) - f(y)^2\phi_{10}\bigl(f(y),g(y)\bigr)\bigr\}$$

(S20)

$$+ \{f(y)-f(x)\}f(y)^2\phi_{10}\bigl(f(y),g(y)\bigr)\bigl(\mathbb{1}_{\{\|x-y\|\leq h_{y,f}^{-1}(s)\}} - \mathbb{1}_{\{\|x-y\|\leq h_{x,f}^{-1}(s)\}}\bigr)\biggr]\,dy\,dx.$$

Using Lemmas S3(i), S5 and S4, and arguing as around (S77), for $s \in \mathcal{I}_{m,X}$, $x, y \in \mathcal{X}_{m,n}$ with $\|x-y\| \leq h_{x,f}^{-1}(s)$ and m, n sufficiently large, we have

$$\biggl|\int_{\mathcal{X}_{m,n}}\int_{\mathcal{X}_{m,n}}\mathbb{1}_{\{\|x-y\|\leq h_{x,f}^{-1}(s)\}}\{f(y)-f(x)\}\bigl\{f(x)^2\phi_{10}\bigl(f(x),g(x)\bigr) - f(y)^2\phi_{10}\bigl(f(y),g(y)\bigr)\bigr\}\,dy\,dx\biggr|$$

$$\lesssim \int_{\mathcal{X}_{m,n}}\int_{\mathcal{X}_{m,n}}\mathbb{1}_{\{\|x-y\|\leq h_{x,f}^{-1}(s)\}}f(x)^{1+\kappa_1}g(x)^{\kappa_2}|f(y)-f(x)|\biggl\{\biggl|\frac{f(y)}{f(x)}-1\biggr|+\biggl|\frac{g(y)}{g(x)}-1\biggr|\biggr\}\,dy\,dx$$

$$\lesssim \int_{\mathcal{X}_{m,n}}\int_{\mathcal{X}_{m,n}}\mathbb{1}_{\{\|x-y\|\leq h_{x,f}^{-1}(s)\}}f(x)^{2+\kappa_1}g(x)^{\kappa_2}\{h_{x,f}^{-1}(s)M_\beta(x)\}^{2\beta}\,dy\,dx$$

$$\lesssim \int_{\mathcal{X}_{m,n}} f(x)^{2+\kappa_1}g(x)^{\kappa_2}h_{x,f}^{-1}(s)^d\{h_{x,f}^{-1}(s)M_\beta(x)\}^{2\beta}\,dx$$

(S21)

$$\lesssim s\int_{\mathcal{X}_{m,n}} f(x)^{1+\kappa_1}g(x)^{\kappa_2}\biggl\{\frac{sM_\beta(x)^d}{f(x)}\biggr\}^{2\beta/d}\,dx.$$

Now, similarly, by Lemma S4 we have

$$\max\biggl\{\biggl|\frac{V_d f(x)h_{x,f}^{-1}(s)^d}{s}-1\biggr|,\biggl|\frac{V_d f(y)h_{y,f}^{-1}(s)^d}{s}-1\biggr|\biggr\} \lesssim \biggl\{\frac{sM_\beta(x)^d}{f(x)}\biggr\}^{\beta/d}.$$

It follows that there exist $C, C' > 0$, depending only on ϑ, such that

$$\begin{aligned} V_d f(x)h_{y,f}^{-1}(s)^d &\geq \frac{f(x)}{f(y)}\biggl[s - Cs\biggl\{\frac{sM_\beta(x)^d}{f(x)}\biggr\}^{\beta/d}\biggr] \\ &\geq \frac{f(x)}{f(y)}\biggl[V_d f(x)h_{x,f}^{-1}(s)^d - 2Cs\biggl\{\frac{sM_\beta(x)^d}{f(x)}\biggr\}^{\beta/d}\biggr] \\ &\geq V_d f(x)h_{x,f}^{-1}(s)^d - 2C's\biggl\{\frac{sM_\beta(x)^d}{f(x)}\biggr\}^{\beta/d}, \end{aligned}$$

where the final bound is from Lemma S5. Hence,

$$\biggl|\int_{\mathcal{X}_{m,n}}\int_{\mathcal{X}_{m,n}}\{f(y)-f(x)\}f(y)^2\phi_{10}\bigl(f(y),g(y)\bigr)\bigl(\mathbb{1}_{\{\|x-y\|\leq h_{y,f}^{-1}(s)\}} - \mathbb{1}_{\{\|x-y\|\leq h_{x,f}^{-1}(s)\}}\bigr)\,dy\,dx\biggr|$$

$$\lesssim \int_{\mathcal{X}_{m,n}} \int_{\mathcal{X}_{m,n}} f(x)^{1+\kappa_1} g(x)^{\kappa_2} |f(y) - f(x)| \mathbb{1}_{\{h_{y,f}^{-1}(s) \leq \|x-y\| \leq h_{x,f}^{-1}(s)\}} \, dy \, dx$$

$$\lesssim \int_{\mathcal{X}_{m,n}} f(x)^{2+\kappa_1} g(x)^{\kappa_2} M_\beta(x)^\beta h_{x,f}^{-1}(s)^\beta \frac{s}{f(x)} \left\{ \frac{s M_\beta(x)^d}{f(x)} \right\}^{\beta/d} dx$$

(S22)

$$\lesssim s \int_{\mathcal{X}_{m,n}} f(x)^{1+\kappa_1} g(x)^{\kappa_2} \left\{ \frac{s M_\beta(x)^d}{f(x)} \right\}^{2\beta/d} dx.$$

It now follows from (S19), (S20), (S21) and (S22) that

$$\bigg| \int_{\mathcal{I}_{m,X}} \int_{\mathcal{I}_{n,Y}} \mathrm{B}_{k_X, m-k_X}(s) \mathrm{B}_{k_Y, n+1-k_Y}(t) \int_{\mathcal{X}_{m,n}} f(x)^2 \phi_{10}\big(f(x), g(x)\big) \left\{ \frac{V_d f(x) h_{x,f}^{-1}(s)^d}{s} - 1 \right\} dx \, dt \, ds \bigg|$$

$$\lesssim \int_{\mathcal{I}_{m,X}} \int_{\mathcal{I}_{n,Y}} \mathrm{B}_{k_X, m-k_X}(s) \mathrm{B}_{k_Y, n+1-k_Y}(t) \int_{\mathcal{X}_{m,n}} f(x)^{1+\kappa_1} g(x)^{\kappa_2} \left\{ \frac{s M_\beta(x)^d}{f(x)} \right\}^{2\beta/d} dx \, dt \, ds + \left(\frac{k_X}{m} \right)^{\lambda_1(1-\zeta)-\epsilon}$$

$$= O\bigg(\max\bigg\{ \left(\frac{k_X}{m} \right)^{2\beta/d}, \left(\frac{k_X}{m} \right)^{\lambda_1(1-\zeta)-\epsilon} \bigg\} \bigg).$$

The other terms in the expansion can be dealt with in the same way, and we conclude that

$$\mathbb{E}\widetilde{T}_{m,n} = \int_{\mathcal{I}_{m,X}} \int_{\mathcal{I}_{n,Y}} \mathrm{B}_{k_X, m-k_X}(s) \mathrm{B}_{k_Y, n+1-k_Y}(t)$$
$$\times \int_{\mathcal{X}} \sum_{\ell_1=0}^{\beta^*-1} \sum_{\ell_2=0}^{\beta^*-1-\ell_1} \frac{f(x)^{1+\ell_1} g(x)^{\ell_2} \phi_{\ell_1 \ell_2}\big(f(x), g(x)\big)}{\ell_1! \ell_2!} \left(\frac{k_X}{ms} - 1 \right)^{\ell_1} \left(\frac{k_Y}{nt} - 1 \right)^{\ell_2} dx \, dt \, ds$$
$$+ O\bigg(\max\bigg\{ \left(\frac{k_X}{m} \right)^{\lambda_1(1-\zeta)-\epsilon}, \left(\frac{k_Y}{n} \right)^{\lambda_2(1-\zeta)-\epsilon}, \left(\frac{k_X}{m} \right)^{2\beta/d}, \left(\frac{k_Y}{n} \right)^{2\beta/d}, k_X^{-\beta_1^*/2}, k_Y^{-\beta_2^*/2} \bigg\} \bigg).$$

The result therefore follows from (S16). □

S1.4. *Proof of Proposition 11 on asymptotic variance.*

PROOF OF PROPOSITION 11. We initially consider the unweighted estimator $\widetilde{T}_{m,n}$, deferring the extension to the weighted estimator $\widehat{T}_{m,n}^{w_X,w_Y}$ to the end of the proof. We start by writing

$$\mathrm{Var}(\widetilde{T}_{m,n}) = \frac{1}{m} \mathrm{Var}\, \phi\big(\widehat{f}_{(k_X),1}, \widehat{g}_{(k_Y),1}\big)$$
$$+ \left(1 - \frac{1}{m}\right) \mathrm{Cov}\Big(\phi\big(\widehat{f}_{(k_X),1}, \widehat{g}_{(k_Y),1}\big), \phi\big(\widehat{f}_{(k_X),2}, \widehat{g}_{(k_Y),2}\big) \Big). \quad \text{(S23)}$$

Taking $\mathcal{X}_{m,n}, \mathcal{I}_{m,X}$ and $\mathcal{I}_{n,Y}$ as defined in the proof of Proposition S1, and letting S_1, S_2 and S_3 be error terms, we now write

$$\mathbb{E}\big\{ \phi\big(\widehat{f}_{(k_X),1}, \widehat{g}_{(k_Y),1}\big)^2 \big\}$$

$$= \int_{\mathcal{X}_{m,n}} f(x) \int_{\mathcal{I}_{m,X}} \int_{\mathcal{I}_{n,Y}} \phi(u_{x,s}, v_{x,t})^2 \mathrm{B}_{k_X,m-k_X}(s)\mathrm{B}_{k_Y,n+1-k_Y}(t)\,ds\,dt\,dx + S_1$$

$$= \int_{\mathcal{X}_{m,n}} f(x) \int_{\mathcal{I}_{m,X}} \int_{\mathcal{I}_{n,Y}} \phi\Big(\frac{k_X f(x)}{ms}, \frac{k_Y g(x)}{nt}\Big)^2 \times \mathrm{B}_{k_X,m-k_X}(s)\mathrm{B}_{k_Y,n+1-k_Y}(t)\,ds\,dt\,dx + S_1 + S_2$$

$$= \mathbb{E}\big\{(\phi_{X_1})^2\big\} + \sum_{j=1}^{3} S_j.$$

We show in Section S1.9 that

$$\sum_{j=1}^{3} S_j = O\bigg(\max\bigg\{\Big(\frac{k_X}{m}\Big)^{\lambda_1(1-2\zeta)-\epsilon}, \Big(\frac{k_Y}{n}\Big)^{\lambda_2(1-2\zeta)-\epsilon}, \Big(\frac{k_X}{m}\Big)^{\frac{2\wedge\beta}{d}}, \Big(\frac{k_Y}{n}\Big)^{\frac{2\wedge\beta}{d}}, k_X^{-\frac{1}{2}}, k_Y^{-\frac{1}{2}}\bigg\}\bigg)$$

for every $\epsilon > 0$. Using Proposition S1 we can now see that

$$\bigg|\frac{1}{m}\mathrm{Var}\,\phi\big(\widehat{f}_{(k_X),1}, \widehat{g}_{(k_Y),1}\big) - \frac{\mathrm{Var}(\phi_{X_1})}{m}\bigg| = o\Big(\frac{1}{m}\Big). \tag{S24}$$

We now turn to the second term in (S23). Let $F_{m,n,x,y} : [0,1]^4 \to [0,1]$ denote the conditional distribution function of

$$\big(h_{x,f}(\rho_{(k_X),1,X}), h_{y,f}(\rho_{(k_X),2,X}), h_{x,g}(\rho_{(k_Y),1,Y}), h_{y,g}(\rho_{(k_Y),2,Y})\big)|X_1 = x, X_2 = y.$$

Moreover, for $s_1, s_2, t_1, t_2 \in [0,1]$ such that $s_1 + s_2 \le 1$ and $t_1 + t_2 \le 1$ define

$$G_m^{(1)}(s_1,s_2) := \int_0^{s_1}\int_0^{s_2} \mathrm{B}_{k_X,k_X,m-2k_X-1}(u_1,u_2)\,du_1\,du_2$$

$$G_n^{(2)}(t_1,t_2) := \int_0^{t_1}\int_0^{t_2} \mathrm{B}_{k_Y,k_Y,n-2k_Y+1}(v_1,v_2)\,dv_1\,dv_2$$

$$G_{m,n}(s_1,s_2,t_1,t_2) := G_m^{(1)}(s_1,s_2)G_n^{(2)}(t_1,t_2),$$

so that we have $F_{m,n,x,y}(s_1,s_2,t_1,t_2) = G_{m,n}(s_1,s_2,t_1,t_2)$ for s_1,s_2,t_1,t_2,x and y such that $\|x-y\| > \max\big(h_{x,f}^{-1}(s_1) + h_{y,f}^{-1}(s_2), h_{x,g}^{-1}(t_1) + h_{y,g}^{-1}(t_2)\big)$. We will also use the shorthand $h(s_1,s_2,t_1,t_2) := \phi(u_{x,s_1}, v_{x,t_1})\phi(u_{y,s_2}, v_{y,t_2})$ and

$$H_m^{(1)}(s_1,s_2) := G_m^{(1)}(s_1,s_2) - \int_0^{s_1}\int_0^{s_2} \mathrm{B}_{k_X,m-k_X}(u_1)\mathrm{B}_{k_X,m-k_X}(u_2)\,du_1\,du_2$$

$$H_n^{(2)}(t_1,t_2) := G_n^{(2)}(t_1,t_2) - \int_0^{t_1}\int_0^{t_2} \mathrm{B}_{k_Y,n+1-k_Y}(v_1)\mathrm{B}_{k_Y,n+1-k_Y}(v_2)\,dv_1\,dv_2$$

$$H_{m,n}(s_1,s_2,t_1,t_2) := H_m^{(1)}(s_1,s_2)G_n^{(2)}(t_1,t_2) + G_m^{(1)}(s_1,s_2)H_n^{(2)}(t_1,t_2) - H_m^{(1)}(s_1,s_2)H_n^{(2)}(t_1,t_2).$$

With this newly-defined notation, we now have

$$\mathrm{Cov}\Big(\phi\big(\widehat{f}_{(k_X),1}, \widehat{g}_{(k_Y),1}\big), \phi\big(\widehat{f}_{(k_X),2}, \widehat{g}_{(k_Y),2}\big)\Big)$$

$$= \int_{\mathcal{X}\times\mathcal{X}} f(x)f(y) \int_{[0,1]^4} h(s_1,s_2,t_1,t_2)\big\{dF_{m,n,x,y}(s_1,s_2,t_1,t_2) - d(H_m^{(1)} - G_m^{(1)})(s_1,s_2)d(H_n^{(2)} - G_n^{(2)})(t_1,t_2)\big\}\,dx\,dy$$

$$= \int_{\mathcal{X}\times\mathcal{X}} f(x)f(y) \int_{\mathcal{I}^2_{m,X}} \int_{\mathcal{I}^2_{n,Y}} h(s_1,s_2,t_1,t_2)\bigl\{d(F_{m,n,x,y} - G_{m,n})(s_1,s_2,t_1,t_2) + dH_{m,n}(s_1,s_2,t_1,t_2)\bigr\}\,dx\,dy + o(m^{-2}+n^{-2}), \tag{S25}$$

where the bound on the final term follows from the fact that $\zeta < 1/2$, Lemma S9 and Cauchy–Schwarz. We first study the second term in this expansion. The intuition behind the following expansion is that, when X_1 and X_2 do not share nearest neighbours, the dependence between $(\widehat{f}_{(k_X),1}, \widehat{g}_{(k_Y),1})$ and $(\widehat{f}_{(k_X),2}, \widehat{g}_{(k_Y),2})$ is relatively weak, and we may expand the functions $\phi, h_{x,f}^{-1}, h_{x,g}^{-1}$ as in the proof of Proposition S1 and approximate integrals. We therefore make use of the shorthand

$$h^{(1)}(s_1,s_2,t_1,t_2) := \Bigl\{\phi\bigl(f(x),v_{x,t_1}\bigr) + \Bigl(\frac{k_X}{ms_1}-1\Bigr)f(x)\phi_{10}\bigl(f(x),v_{x,t_1}\bigr)\Bigr\} \times \Bigl\{\phi\bigl(f(y),v_{y,t_2}\bigr) + \Bigl(\frac{k_X}{ms_2}-1\Bigr)f(y)\phi_{10}\bigl(f(y),v_{y,t_2}\bigr)\Bigr\}$$

$$h^{(2)}(s_1,s_2,t_1,t_2) := \Bigl\{\phi\bigl(u_{x,s_1},g(x)\bigr) + \Bigl(\frac{k_Y}{nt_1}-1\Bigr)g(x)\phi_{01}\bigl(u_{x,s_1},g(x)\bigr)\Bigr\} \times \Bigl\{\phi\bigl(u_{y,s_2},g(y)\bigr) + \Bigl(\frac{k_Y}{nt_2}-1\Bigr)g(y)\phi_{01}\bigl(u_{y,t_2},g(y)\bigr)\Bigr\}$$

for linearised versions of h. We also write, for example,

$$(h\,dH_m^{(1)}\,dG_n^{(2)})(s_1,s_2,t_1,t_2) := h(s_1,s_2,t_1,t_2)\,dH_m^{(1)}(s_1,s_2)\,dG_n^{(1)}(t_1,t_2).$$

Writing T_1, T_2 and T_3 for error terms, we therefore have

$$\begin{aligned}
&\int_{\mathcal{X}^2} f(x)f(y)\int_{\mathcal{I}^2_{m,X}}\int_{\mathcal{I}^2_{n,Y}} (h\,dH_{m,n})(s_1,s_2,t_1,t_2)\,dx\,dy \\
&= \int_{\mathcal{X}^2_{m,f}} f(x)f(y)\int_{\mathcal{I}^2_{m,X}}\int_{\mathcal{I}^2_{n,Y}} (h\,dH_m^{(1)}\,dG_n^{(2)})(s_1,s_2,t_1,t_2)\,dx\,dy + T_1 \\
&\quad + \int_{\mathcal{X}^2_{n,g}} f(x)f(y)\int_{\mathcal{I}^2_{m,X}}\int_{\mathcal{I}^2_{n,Y}} (h\,d(G_m^{(1)}-H_m^{(1)})\,dH_n^{(2)})(s_1,s_2,t_1,t_2)\,dx\,dy \\
&= \int_{\mathcal{X}^2_{m,f}} f(x)f(y)\int_{\mathcal{I}^2_{m,X}}\int_{\mathcal{I}^2_{n,Y}} (h^{(1)}\,dH_m^{(1)}\,dG_n^{(2)})(s_1,s_2,t_1,t_2)\,dx\,dy + T_1 + T_2 \\
&\quad + \int_{\mathcal{X}^2_{n,g}} f(x)f(y)\int_{\mathcal{I}^2_{m,X}}\int_{\mathcal{I}^2_{n,Y}} (h^{(2)}d(G_m^{(1)}-H_m^{(1)})dH_n^{(2)})(s_1,s_2,t_1,t_2)\,dx\,dy \\
&= -\frac{1}{m}\int_{\mathcal{X}^2_{m,f}} f(x)f(y)\int_{\mathcal{I}^2_{n,Y}} \Bigl\{2f(x)\phi_{10}\bigl(f(x),v_{x,t_1}\bigr)\phi\bigl(f(y),v_{y,t_2}\bigr) \\
&\qquad + f(x)\phi_{10}\bigl(f(x),v_{x,t_1}\bigr)f(y)\phi_{10}\bigl(f(y),v_{y,t_2}\bigr)\Bigr\}dG_n^{(2)}(t_1,t_2)\,dx\,dy \\
&\quad - \frac{1}{n}\int_{\mathcal{X}^2_{n,g}} f(x)f(y)\int_{\mathcal{I}^2_{m,X}} g(x)\phi_{01}\bigl(u_{x,s_1},g(x)\bigr)g(y)\phi_{01}\bigl(u_{y,s_2},g(y)\bigr) \\
&\qquad \times d(G_m^{(1)}-H_m^{(1)})(s_1,s_2)\,dx\,dy + T_1 + T_2 + T_3 \\
&= -\frac{2}{m}\mathbb{E}\bigl\{(f\phi_{10})_{X_1}\bigr\}\mathbb{E}(\phi_{X_1}) - \frac{1}{m}\bigl\{\mathbb{E}(f\phi_{10})_{X_1}\bigr\}^2 - \frac{1}{n}\bigl\{\mathbb{E}(g\phi_{01})_{X_1}\bigr\}^2
\end{aligned}$$

(S26)
$$+T_1+T_2+T_3+o(1/m+1/n),$$

where the bound on the final term follows from (S8), Lemma S3(i), Lemma S6 and tail bounds similar to (S5). We show in Section S1.9 that

$$\sum_{j=1}^{3} T_j = O\biggl(\max\biggl\{\Bigl(\frac{k_X}{m}\Bigr)^{1+\lambda_1(1-\zeta)-\epsilon}, \Bigl(\frac{k_Y}{n}\Bigr)^{1+\lambda_2(1-\zeta)-\epsilon}, \frac{k_X^{\frac{1}{2}+\frac{2\wedge\beta}{d}}}{m^{1+\frac{2\wedge\beta}{d}}}, \frac{k_Y^{\frac{1}{2}+\frac{2\wedge\beta}{d}}}{n^{1+\frac{2\wedge\beta}{d}}},$$
$$\Bigl(\frac{k_X}{m}\Bigr)^{1+\frac{2(2\wedge\beta)}{d}}, \Bigl(\frac{k_Y}{n}\Bigr)^{1+\frac{2(2\wedge\beta)}{d}}, \frac{\log m}{mk_X^{\frac{1}{2}}}, \frac{\log n}{nk_Y^{\frac{1}{2}}}\biggr\}\biggr) + o(1/m+1/n). \tag{S27}$$

We now consider the contribution of the first term in (S25). In Section S1.9, we show that

$$U_0 := \int_{\mathcal{X}\times\mathcal{X}_{m,n}^c} f(x)f(y)\int_{\mathcal{I}_{m,X}^2}\int_{\mathcal{I}_{n,Y}^2} h(s_1,s_2,t_1,t_2)$$
$$\times\, d(F_{m,n,x,y}-G_{m,n})(s_1,s_2,t_1,t_2)\,dx\,dy$$
$$= O\biggl(\max\biggl\{\Bigl(\frac{k_X}{m}\Bigr)^{2\lambda_1(1-\zeta)-\epsilon}, \Bigl(\frac{k_Y}{n}\Bigr)^{2\lambda_2(1-\zeta)-\epsilon}\biggr\}\biggr), \tag{S28}$$

so that we may restrict attention to $x\in\mathcal{X}_{m,n}$, in which case $F_{m,n,x,y}-G_{m,n}$ is only non-zero when x and y are close and we may approximate $f(y)\approx f(x)$ and $g(y)\approx g(x)$. Let $p_\cap^{(1)} := \int_{B_x(h_{x,f}^{-1}(s_1))\cap B_y(h_{y,f}^{-1}(s_2))} f(w)\,dw$ and $p_\cap^{(2)} := \int_{B_x(h_{x,g}^{-1}(t_1))\cap B_y(h_{y,g}^{-1}(t_2))} g(w)\,dw$, and let

(S29)
$$(N_1^{(1)},N_2^{(1)},N_3^{(1)},N_4^{(1)}) \sim \mathrm{Multi}\bigl(m-2; s_1-p_\cap^{(1)}; s_2-p_\cap^{(1)}, p_\cap^{(1)}, 1-s_1-s_2+p_\cap^{(1)}\bigr)$$
$$(N_1^{(2)},N_2^{(2)},N_3^{(2)},N_4^{(2)}) \sim \mathrm{Multi}\bigl(n; t_1-p_\cap^{(2)}; t_2-p_\cap^{(2)}, p_\cap^{(2)}, 1-t_1-t_2+p_\cap^{(2)}\bigr).$$

Now set

$$F_{m,x,y}^{(1)}(s_1,s_2) := \mathbb{P}(N_1^{(1)}+N_3^{(1)} \geq k_X - \mathbb{1}_{\{\|x-y\|\leq h_{x,f}^{-1}(s_1)\}},$$
$$N_2^{(1)}+N_3^{(1)} \geq k_X - \mathbb{1}_{\{\|x-y\|\leq h_{y,f}^{-1}(s_2)\}})$$
$$F_{n,x,y}^{(2)}(t_1,t_2) := \mathbb{P}(N_1^{(2)}+N_3^{(2)} \geq k_Y, N_2^{(2)}+N_3^{(2)} \geq k_Y), \tag{S30}$$

so that $F_{m,n,x,y}(s_1,s_2,t_1,t_2) = F_{m,x,y}^{(1)}(s_1,s_2)F_{n,x,y}^{(2)}(t_1,t_2)$. We use the decomposition

$$F_{m,n,x,y}-G_{m,n} = F_{m,x,y}^{(1)}F_{n,x,y}^{(2)} - G_m^{(1)}G_n^{(2)}$$
$$= (F_{m,x,y}^{(1)}-G_m^{(1)})(F_{n,x,y}^{(2)}-G_n^{(2)}) + (F_{m,x,y}^{(1)}-G_m^{(1)})G_n^{(2)} + G_m^{(1)}(F_{n,x,y}^{(2)}-G_n^{(2)}), \tag{S31}$$

so that each term is of product form and involves at least one of the marginal errors. We will see that the first term is asymptotically negligible, while the second and third terms can be studied through the normal approximation given in Lemma S11. For a general distribution function F, for $a_-\leq a_+$ and for a smooth $h:[a_-,a_+]^2\to\mathbb{R}$ with first partial derivatives h_{10}, h_{10} and mixed second partial derivative h_{11}, we will use the formula

$$\int_{[a_-,a_+]^2}(h\,dF)(u,v) - \int_{a_-}^{a_+}\int_{a_-}^{a_+}(h_{11}F(u,v))\,du\,dv$$
$$= \int_{a_-}^{a_+}\bigl[(h_{10}F)(u,a_-)-(h_{10}F)(u,a_+)\bigr]du + (hF)(a_-,a_-) - (hF)(a_+,a_-)$$

(S32) $$+\int_{a_-}^{a_+}\bigl[(h_{01}F)(a_-,v)-(h_{01}F)(a_+,v)\bigr]\,dv+(hF)(a_+,a_+)-(hF)(a_-,a_+).$$

We now deal with each of the three terms on the right-hand side of (S31) in turn, starting with $F = F^{(1)}F^{(2)} = (F^{(1)}_{m,x,y} - G^{(1)}_m)(F^{(2)}_{n,x,y} - G^{(2)}_n)$. For remainder terms U_1, U_2 and U_3 to be bounded later, we write

$$\int_{\mathcal{X}\times\mathcal{X}_{m,n}} f(x)f(y)\int_{\mathcal{I}^2_{m,X}}\int_{\mathcal{I}^2_{n,Y}}(h\,dF)(s_1,s_2,t_1,t_2)\,dx\,dy$$

$$=\int_{\mathcal{X}\times\mathcal{X}_{m,n}} f(x)f(y)\int_{\mathcal{I}^2_{m,X}}\int_{\mathcal{I}^2_{n,Y}}\bigl(h_{0011}\,(dF^{(1)})\,F^{(2)}\bigr)(s_1,s_2,t_1,t_2)\,dt_1\,dt_2\,dx\,dy+U_1$$

$$=\int_{\mathcal{X}\times\mathcal{X}_{m,n}} f(x)f(y)\int_{\mathcal{I}^2_{n,Y}}F^{(2)}(t_1,t_2)\biggl\{\int_{\mathcal{I}^2_{m,X}}(h_{1111}F^{(1)})(s_1,s_2,t_1,t_2)\,ds_1\,ds_2$$

$$-\int_{\mathcal{I}_{m,X}}(h_{1011}F^{(1)})(s_1,a^+_{m,X},t_1,t_2)\,ds_1$$

$$-\int_{\mathcal{I}_{m,X}}(h_{0111}F^{(1)})(a^+_{m,X},s_2,t_1,t_2)\,ds_2\biggr\}\,dt_1\,dt_2\,dx\,dy+U_1+U_2$$

(S33)
$$=\sum_{j=1}^{3}U_j.$$

We show in Section S1.9 that

(S34) $$\sum_{j=1}^{3}U_j = O\biggl(\max\biggl\{\frac{1}{m^2},\frac{1}{n^2},\frac{\log^2 m}{mk_X},\frac{\log^2 n}{nk_Y}\biggr\}\biggr).$$

We next consider $F = F^{(1)}F^{(2)} = (F^{(1)}_{m,x,y} - G^{(1)}_m)G^{(2)}_n$, and recall from Lemma S11 that $\alpha_z = \mu_d\bigl(B_0(1)\cap B_z(1)\bigr)/V_d$, that

$$\Sigma = \begin{pmatrix} 1 & \alpha_z \\ \alpha_z & 1\end{pmatrix},$$

and the definitions of the normal distribution functions Φ_{I_2} and Φ_Σ. Then, for remainder terms U_4, U_5, U_6 to be bounded below, we use the change of variables $y = x + (\frac{k_X}{mV_d f(x)})^{1/d}z$ and the approximation $\frac{\partial}{\partial s}\phi(u_{x,s},v_{x,t}) \approx f(x)\phi_{10}(f(x),g(x))/s$ to write

$$\int_{\mathcal{X}\times\mathcal{X}_{m,n}} f(x)f(y)\int_{\mathcal{I}^2_{m,X}}\int_{\mathcal{I}^2_{n,Y}}(h\,dF)(s_1,s_2,t_1,t_2)\,dx\,dy$$

$$=\int_{\mathcal{X}\times\mathcal{X}_{m,n}} f(x)f(y)\int_{\mathcal{I}^2_{n,Y}}\biggl\{\int_{\mathcal{I}^2_{m,X}}(h_{1100}\,F^{(1)})(s_1,s_2,t_1,t_2)\,ds_1\,ds_2$$

$$-\int_{\mathcal{I}_{m,X}}(h_{1000}F^{(1)})(s_1,a^+_{m,X},t_1,t_2)\,ds_1$$

$$-\int_{\mathcal{I}_{m,X}}(h_{0100}F^{(1)})(a^+_{m,X},s_2,t_1,t_2)\,ds_2\biggr\}dG^{(2)}_n(t_1,t_2)\,dx\,dy+U_4$$

$$=\frac{1}{mV_d}\int_{\mathcal{X}_{m,n}} f(x)\int_{\mathbb{R}^d}\biggl\{(f\phi_{10})^2_x\int_{\mathbb{R}^2}(\Phi_\Sigma-\Phi_{I_2})(u_1,u_2)\,du_1\,du_2$$

$$+ 2(f\phi_{10})_x \phi_x \mathbb{1}_{\{\|z\| \leq 1\}} \biggr\} \, dz \, dx + U_4 + U_5$$

(S35)
$$= \frac{1}{m} \mathbb{E}\bigl\{(f\phi_{10})^2_{X_1}\bigr\} + \frac{2}{m} \mathbb{E}\bigl\{(f\phi_{10})_{X_1} \phi_{X_1}\bigr\} + \sum_{j=4}^{6} U_j.$$

We show in Section S1.9 that

$$\sum_{j=4}^{6} U_j = O\biggl(\frac{1}{m} \max\biggl\{\frac{\log^{\frac{5}{2}} m}{k_X^{1/2}}, \frac{\log^{\frac{1}{2}} n}{k_Y^{1/2}}, \log^2 m \Bigl(\frac{k_X}{m}\Bigr)^{\frac{1 \wedge \beta}{d}}, \Bigl(\frac{k_X}{m}\Bigr)^{\lambda_1(1-2\zeta)-\epsilon},$$

(S36)
$$\Bigl(\frac{k_Y}{n}\Bigr)^{\lambda_2(1-2\zeta)-\epsilon}, \Bigl(\frac{k_Y}{n}\Bigr)^{\frac{2 \wedge \beta}{d}}\biggr\}\biggr)$$

for every $\epsilon > 0$. The final term in (S31) can be approximated by writing $F = F^{(1)} F^{(2)} = G_m^{(1)}(F_{n,x,y}^{(2)} - G_n^{(2)})$, using the changes of variables $y = x + (\frac{k_Y}{nV_d g(x)})^{1/d} z, t_i = (k_Y + k_Y v_i)/n$ for $i = 1, 2$ and using the approximation $\frac{\partial}{\partial t} \phi(u_{x,s}, v_{x,t}) \approx g(x)\phi_{01}(f(x), g(x))/t$ to write

$$\int_{\mathcal{X} \times \mathcal{X}_{m,n}} f(x) f(y) \int_{\mathcal{I}^2_{m,X}} \int_{\mathcal{I}^2_{n,Y}} (h \, dF)(s_1, s_2, t_1, t_2) \, dx \, dy$$

$$= \int_{\mathcal{X} \times \mathcal{X}_{m,n}} f(x) f(y) \int_{\mathcal{I}^2_{m,X}} \int_{\mathcal{I}^2_{n,Y}} (h_{0011} \, dG_m^{(1)} F^{(2)})(s_1, s_2, t_1, t_2) \, dt_1 \, dt_2 \, dx \, dy + U_7$$

$$= \frac{1}{nV_d} \int_{\mathcal{X}_{m,n}} g(x) (f\phi_{01})^2_x \int_{\mathbb{R}^d} \int_{\mathbb{R}^2} (\Phi_\Sigma - \Phi_{I_2})(v_1, v_2) \, dv_1 \, dv_2 \, dz \, dx + U_7 + U_8$$

(S37)
$$= \frac{1}{n} \int_{\mathcal{X}_{m,n}} g(x) (f\phi_{01})^2_x \, dx + U_7 + U_8.$$

Let $\epsilon_0 = \epsilon_0(\lambda_1, \lambda_2, \kappa_1, \kappa_2, C) \in (0, \lambda_1 \wedge \lambda_2)$ be sufficiently small that

$$\frac{2 + 2\kappa_1 - \epsilon/(\lambda_1 \wedge \lambda_2)}{1 - \epsilon_0/(\lambda_1 \wedge \lambda_2)} > 2 + 2\kappa_1 - 1/C \quad \text{and} \quad \frac{2\kappa_2 - 1}{1 - \epsilon_0/(\lambda_1 \wedge \lambda_2)} > 2\kappa_2 - 1 - 1/C.$$

Then, by Hölder's inequality, we have that

$$\sup_{(f,g) \in \mathcal{F}_{d,\vartheta}} \int_{\mathcal{X}} f(x)^{2+2\kappa_1} g(x)^{2\kappa_2 - 1} \biggl[\Bigl\{\frac{M_\beta(x)^d}{f(x)}\Bigr\}^{\epsilon_0} + \Bigl\{\frac{M_\beta(x)^d}{g(x)}\Bigr\}^{\epsilon_0}\biggr] \, dx$$

(S38)
$$\leq 2 \sup_{(f,g) \in \mathcal{F}_{d,\vartheta}} \max_{i=1,2} C^{\epsilon_0/\lambda_i} \biggl[\int_{\mathcal{X}} f(x)^{\frac{2+2\kappa_1 - \epsilon_0/\lambda_i}{1-\epsilon_0/\lambda_i}} g(x)^{\frac{2\kappa_2 - 1}{1-\epsilon_0/\lambda_i}} \, dx\biggr]^{1-\epsilon_0/\lambda_i} < \infty.$$

It follows that

$$\int_{\mathcal{X}^c_{m,n}} g(x) (f\phi_{01})^2_x \, dx \lesssim \int_{\mathcal{X}^c_{m,n}} f(x)^{2+2\kappa_1} g(x)^{-1+2\kappa_2} \, dx$$

$$\leq \int_{\mathcal{X}} f(x)^{2+2\kappa_1} g(x)^{2\kappa_2 - 1} \biggl[\Bigl\{\frac{k_X \log m M_\beta(x)^d}{mf(x)}\Bigr\}^{\epsilon_0} + \Bigl\{\frac{k_Y \log n M_\beta(x)^d}{ng(x)}\Bigr\}^{\epsilon_0}\biggr] \, dx$$

(S39)
$$= O\biggl(\max\biggl\{\Bigl(\frac{k_X \log m}{m}\Bigr)^{\epsilon_0}, \Bigl(\frac{k_Y \log n}{n}\Bigr)^{\epsilon_0}\biggr\}\biggr)$$

We show in Section S1.9 that

$$U_7 + U_8 = O\biggl(\frac{1}{n}\max\biggl\{\frac{\log^{5/2} n}{k_Y^{1/2}}, \log^2 n\Bigl(\frac{k_Y}{n}\Bigr)^{(1\wedge\beta)/d}, \Bigl(\frac{k_Y}{n}\Bigr)^{\epsilon_0/2},$$

$$\frac{\log^{1/2} m}{k_X^{1/2}}, \Bigl(\frac{k_X}{m}\Bigr)^{(2\wedge\beta)/d}, \Bigl(\frac{k_X}{m}\Bigr)^{\epsilon_0/2}\biggr\}\biggr). \tag{S40}$$

It now follows from (S23), (S24), (S25), (S26), (S27), (S28), (S33), (S34), (S35), (S36), (S37), (S39) and (S40) that

$$\begin{aligned}\mathrm{Var}(\widetilde{T}_{m,n}) &= \frac{1}{m}\Bigl[\mathrm{Var}(\phi_{X_1}) - 2\mathbb{E}\bigl\{(f\phi_{10})_{X_1}\bigr\}\mathbb{E}(\phi_{X_1}) - \bigl\{\mathbb{E}(f\phi_{10})_{X_1}\bigr\}^2 + \mathbb{E}\bigl\{(f\phi_{10})^2_{X_1}\bigr\} \\ &\quad + 2\mathbb{E}\bigl\{(f\phi_{10}\phi)_{X_1}\bigr\}\Bigr] + \frac{1}{n}\Bigl[\mathbb{E}\bigl\{(f\phi_{01})^2_{Y_1}\bigr\} - \bigl\{\mathbb{E}(g\phi_{01})_{X_1}\bigr\}^2\Bigr] + o(1/m + 1/n) \\ &= \frac{v_1}{m} + \frac{v_2}{n} + o(1/m + 1/n).\end{aligned}$$

For the general, weighted case, we rely on the decomposition

$$\begin{aligned}\mathrm{Var}(\widehat{T}_{m,n}) = \sum_{j_X,\ell_X=1}^{k_X}\sum_{j_Y,\ell_Y=1}^{k_Y} & w_{X,j_X}w_{X,\ell_X}w_{Y,j_Y}w_{Y,\ell_Y} \\ & \times\biggl\{\frac{1}{m}\mathrm{Cov}\bigl(\phi(\widehat{f}_{(j_X),1},\widehat{g}_{(j_Y),1}),\phi(\widehat{f}_{(\ell_X),1},\widehat{g}_{(\ell_Y),1})\bigr) \\ & + \Bigl(1-\frac{1}{m}\Bigr)\mathrm{Cov}\bigl(\phi(\widehat{f}_{(j_X),1},\widehat{g}_{(j_Y),1}),\phi(\widehat{f}_{(\ell_X),2},\widehat{g}_{(\ell_Y),2})\bigr)\biggr\}.\end{aligned} \tag{S41}$$

Now, for example, when $\ell_X > j_X$, we have

$$\bigl(h_{x,f}(\rho_{(j_X),1,X}), h_{x,f}(\rho_{(\ell_X),1,X}), 1-h_{x,f}(\rho_{(\ell_X),1,X})\bigr)|X_1 = x \sim \mathrm{Dir}(j_X, \ell_X - j_X, m - \ell_X),$$

and it may therefore be deduced similarly to the arguments leading to (S24) that

$$\max_{\substack{j_X,\ell_X: w_{X,j_X},w_{X,\ell_X}\neq 0 \\ j_Y,\ell_Y: w_{Y,j_Y},w_{Y,\ell_Y}\neq 0}}\Bigl|\mathrm{Cov}\bigl(\phi(\widehat{f}_{(j_X),1},\widehat{g}_{(j_Y),1}),\phi(\widehat{f}_{(\ell_X),1},\widehat{g}_{(\ell_Y),1})\bigr) - \mathrm{Var}(\phi_{X_1})\Bigr| = o(1). \tag{S42}$$

The second term on the right-hand side of (S41) is handled using relatively small modifications of the arguments used to study the covariance term in (S23). These modifications are required to account for the fact that the k_X that appears twice in the covariance term in (S23) is now replaced with j_X and ℓ_X (with similar changes to k_Y). Thus, for instance, the joint conditional distribution function of

$$\bigl(h_{x,f}(\rho_{(j_X),1,X}), h_{y,f}(\rho_{(\ell_X),2,X}), h_{x,g}(\rho_{(j_Y),1,Y}), h_{y,g}(\rho_{(\ell_Y),2,Y})\bigr)|X_1 = x, X_2 = y,$$

is now given by

$$\begin{aligned}&F_{m,n,x,y}(s_1,s_2,t_1,t_2) \\ &= \mathbb{P}\bigl(N_1^{(1)} + N_3^{(1)} \geq j_X - \mathbb{1}_{\{\|x-y\|\leq h_{x,f}^{-1}(s_1)\}}, N_2^{(1)} + N_3^{(1)} \geq \ell_X - \mathbb{1}_{\{\|x-y\|\leq h_{y,f}^{-1}(s_2)\}}\bigr) \\ &\qquad\qquad \times \mathbb{P}(N_1^{(2)} + N_3^{(2)} \geq j_Y, N_2^{(2)} + N_3^{(2)} \geq \ell_Y).\end{aligned}$$

Following the arguments through reveals that

$$\max_{\substack{j_X,\ell_X:w_{X,j_X},w_{X,\ell_X}\neq 0\\ j_Y,\ell_Y:w_{Y,j_Y},w_{Y,\ell_Y}\neq 0}} \biggl|\mathrm{Cov}\bigl(\phi(\widehat{f}_{(j_X),1},\widehat{g}_{(j_Y),1}),\phi(\widehat{f}_{(\ell_X),2},\widehat{g}_{(\ell_Y),2})\bigr) - \frac{v_1 - \mathrm{Var}(\phi_{X_1})}{m} - \frac{v_2}{n}\biggr|$$

$$= o(1/m+1/n). \tag{S43}$$

Finally, then, we can deduce from (S41), (S42) and (S43), and using our hypotheses on $\|w_X\|_1$ and $\|w_Y\|_1$, that

$$\mathrm{Var}(\widehat{T}_{m,n}) - \frac{v_1}{m} - \frac{v_2}{n} = o\biggl(\Bigl(\frac{1}{m}+\frac{1}{n}\Bigr)\|w_X\|_1^2\|w_Y\|_1^2\biggr) = o\Bigl(\frac{1}{m}+\frac{1}{n}\Bigr),$$

as required. □

S1.5. *Proofs of Theorems 3 and 5 on asymptotic normality and confidence intervals.* Since the proof of Theorem 3 depends on Proposition 4, we prove Proposition 4 first.

PROOF OF PROPOSITION 4. Where it does not cause confusion, we will supress suffices to write k instead of k_X or k_Y. For any $\ell \geq \max(k+1,i)$ we use the shorthand

$$\widehat{f}_{(k),i,\ell} := \frac{k}{\ell V_d \rho^d_{(k),i,\ell}},$$

and we write $\phi^g_x(\cdot) := \phi\bigl(\cdot, g(x)\bigr)$. We will first study the difference $T_m^{(1)} - T_m^{(1),\mathrm{P}}$ by bounding its first and second conditional moments given M. On the event that $|m/M-1| \leq 1/L$, when $m \geq (1+1/L)(1+k)\log(em)$, we have that

$$\begin{aligned}
&\mathbb{E}\{T_m^{(1)} - T_m^{(1),\mathrm{P}}|M\} \\
&= \mathbb{E}T_m^{(1)} - \frac{M}{m}\mathbb{E}(T_M^{(1)}|M) + \frac{M}{m}\mathbb{E}\biggl\{\phi^g_{X_1}\bigl(\widehat{f}_{(k),1,M}\bigr) - \phi^g_{X_1}\biggl(\frac{M}{m}\widehat{f}_{(k),1,M}\biggr)\biggm| M\biggr\} \\
&\quad + \Bigl(\frac{M}{m}-1\Bigr)\int_{\mathcal{X}} f(x)\{\phi_x + (f\phi_{10})_x\}\,dx \\
&= (\mathbb{E}T_m^{(1)} - T) - \frac{M}{m}\{\mathbb{E}(T_M^{(1)}|M) - T\} + \Bigl(\frac{M}{m}-1\Bigr)\int_{\mathcal{X}} f(x)(f\phi_{10})_x\,dx \\
&\quad + \frac{M}{m}\int_{\mathcal{X}} f(x)\int_0^1 \Bigl\{\phi^g_x\Bigl(\frac{m}{M}u_{x,s}\Bigr) - \phi^g_x(u_{x,s})\Bigr\}\mathrm{B}_{k,M-k}(s)\,ds\,dx \\
&= \mathbb{E}T_m^{(1)} - \mathbb{E}(T_M^{(1)}|M) + \Bigl(\frac{M}{m}-1\Bigr)\int_{\mathcal{X}} f(x)(f\phi_{10})_x\,dx + o\biggl(m^{-1/2} + \Bigl|\frac{M}{m}-1\Bigr|\biggr) \\
&\quad + \frac{M}{m}\int_{\mathcal{X}_{m,f}} f(x)\int_{\mathcal{I}_{m,X}} \Bigl\{\phi^g_x\Bigl(\frac{m}{M}u_{x,s}\Bigr) - \phi^g_x(u_{x,s})\Bigr\}\mathrm{B}_{k,M-k}(s)\,ds\,dx \\
&= \mathbb{E}T_m^{(1)} - \mathbb{E}(T_M^{(1)}|M) + \Bigl(\frac{M}{m}-1\Bigr)\int_{\mathcal{X}} f(x)(f\phi_{10})_x\,dx + o\biggl(m^{-1/2} + \Bigl|\frac{M}{m}-1\Bigr|\biggr) \\
&\quad + \Bigl(1-\frac{M}{m}\Bigr)\int_{\mathcal{X}_{m,f}} f(x)\int_{\mathcal{I}_{m,X}} u_{x,s}\phi_{10}\bigl(u_{x,s},g(x)\bigr)\mathrm{B}_{k,M-k}(s)\,ds\,dx \\
&= \mathbb{E}T_m^{(1)} - \mathbb{E}(T_M^{(1)}|M) + o\biggl(m^{-1/2} + \Bigl|\frac{M}{m}-1\Bigr|\biggr).
\end{aligned}$$

It now follows from the one-sample ($n = \infty$) version of Proposition S1 and the fact that, for $a > 0$, we have $(k/m)^a - (k/M)^a = o(|M/m - 1|)$, that on the event that $|m/M - 1| \leq 1/L$ we have

$$\mathbb{E}\{T_m^{(1)} - T_m^{(1),\mathrm{p}}|M\} = o\biggl(m^{-1/2} + \biggl|\frac{M}{m} - 1\biggr|\biggr).$$

We now bound the conditional variance of $T_m^{(1)} - T_m^{(1),\mathrm{p}}$ on the event $A_m := \{|M/m - 1| \leq 1/\log(em)\}$. We first see that, when $m \geq (k+1)\log(em)/\{1 - 1/\log(em)\}$, we have

$$\begin{aligned}
&\mathrm{Var}\biggl\{\frac{1}{m}\sum_{i=1}^{M}\phi^g_{X_i}\biggl(\frac{M}{m}\widehat{f}_{(k),i,M}\biggr) - \frac{1}{m}\sum_{i=1}^{m}\phi^g_{X_i}\biggl(\frac{M}{m}\widehat{f}_{(k),i,M}\biggr)\biggm| M\biggr\} \\
&\quad = \frac{|M-m|}{m^2}\mathrm{Var}\biggl\{\phi^g_{X_1}\biggl(\frac{M}{m}\widehat{f}_{(k),1,M}\biggr)\biggm| M\biggr\} \\
&\qquad + \frac{|M-m|(|M-m|-1)}{m^2}\mathrm{Cov}\biggl\{\phi^g_{X_1}\biggl(\frac{M}{m}\widehat{f}_{(k),1,M}\biggr), \phi^g_{X_2}\biggl(\frac{M}{m}\widehat{f}_{(k),2,M}\biggr)\biggm| M\biggr\} \\
&\quad = \frac{(M-m)^2}{m^2}\mathrm{Cov}(\phi_{X_1}, \phi_{X_2}) + O\biggl(\frac{|M-m|}{m^2}\biggr) + o\biggl(\frac{(M-m)^2}{m^2}\biggr) \\
&\quad = O\biggl(\frac{|M-m|}{m^2}\biggr) + o\biggl(\frac{(M-m)^2}{m^2}\biggr).
\end{aligned}$$

To bound the conditional variance of $T_m^{(1)} - T_m^{(1),\mathrm{p}}$, it now suffices to bound $\mathrm{Var}(D_m|M)$, where

$$D_m := \frac{1}{m}\sum_{i=1}^{m}\biggl\{\phi^g_{X_i}\biggl(\frac{k}{mV_d\rho^d_{(k),i,m}}\biggr) - \phi^g_{X_i}\biggl(\frac{k}{mV_d\rho^d_{(k),i,M}}\biggr)\biggr\}.$$

To proceed, we will now use the Efron–Stein inequality; see, for example, Boucheron, Lugosi and Massart (2013, Theorem 3.1). Given M, the random variable $D_m = D_m(X_1, \ldots, X_M)$ is a function of the independent random variables $X_1, \ldots, X_M$; letting $X_1', \ldots, X_M'$ denote an independent copy of these random variables, for $j = 1, \ldots, M$, write $D_m^{(j)} := D_m(X_1, \ldots, X_{j-1}, X_j', X_{j+1}, \ldots, X_M)$ for the random variable calculated by replacing X_j in D_m by X_j'. Similarly define $\rho^{(j)}_{(k),i,\ell}$. The Efron–Stein inequality gives that

$$\mathrm{Var}(D_m|M) \leq \frac{1}{2}\sum_{j=1}^{M}\mathbb{E}\bigl\{(D_m - D_m^{(j)})^2|M\bigr\}.$$

For now, we will work on the event $\{M > m\}$. Observe that for $i = 1, \ldots, m$ and $j = m+1, \ldots, M$ we have $\rho^{(j)}_{(k),i,M} = \rho_{(k),i,M}$ unless either X_j is one of the k nearest neighbours of X_i in the sample $X_1, \ldots, X_M$ or X_j' is one of the k nearest neighbours of X_i in the sample $X_1, \ldots, X_{j-1}, X_j', X_{j+1}, \ldots, X_M$. For $j = m+1, \ldots, M$ we have, by arguments similar to those in the proof of Proposition 11, using the fact that $\mathbb{P}(\|X_j - X_1\| \leq \rho_{(k),1,M}|M) = k/(M-1)$ and splitting up into the cases $X_1 \in \mathcal{X}_{m,f}$ and $X_1 \notin \mathcal{X}_{m,f}$, that

$$\begin{aligned}
&\mathbb{E}\{(D_m - D_m^{(j)})^2|M\} \\
&\leq 4\mathbb{E}\biggl[\biggl\{\frac{1}{m}\sum_{i=1}^{m}\mathbb{1}_{\{\|X_j - X_i\| \leq \rho_{(k),i,M}\}}\biggl(\phi^g_{X_i}\biggl(\frac{k}{mV_d\rho^d_{(k),i,M}}\biggr) - \phi_{X_i}\biggr)\biggr\}^2\biggm| M\biggr]
\end{aligned}$$

$$
\begin{aligned}
&= 4(1-1/m)\mathbb{E}\biggl[\mathbb{1}_{\{\|X_j-X_1\|\le\rho_{(k),1,M}\}}\mathbb{1}_{\{\|X_j-X_2\|\le\rho_{(k),2,M}\}}\biggl\{\phi_{X_1}^g\biggl(\frac{k}{mV_d\rho_{(k),1,M}^d}\biggr)-\phi_{X_1}\biggr\} \\
&\qquad\qquad\qquad\times\biggl\{\phi_{X_2}^g\biggl(\frac{k}{mV_d\rho_{(k),2,M}^d}\biggr)-\phi_{X_2}\biggr\}\biggm| M\biggr] \\
&\qquad + \frac{4}{m}\mathbb{E}\biggl[\mathbb{1}_{\{\|X_j-X_1\|\le\rho_{(k),1,M}\}}\biggl\{\phi_{X_1}^g\biggl(\frac{k}{mV_d\rho_{(k),1,M}^d}\biggr)-\phi_{X_1}\biggr\}^2\biggm| M\biggr] \\
&\lesssim \frac{k}{m}\biggl(\frac{k}{m}+\frac{1}{m}\biggr)\int_{\mathcal{X}_{m,f}} f(x)^{1+2\kappa_1}g(x)^{2\kappa_2}\max\biggl[\frac{(M-m)^2}{m^2},\frac{\log m}{k},\biggl\{\frac{kM_\beta(x)^d}{mf(x)}\biggr\}^{\frac{2(2\wedge\beta)}{d}}\biggr]\,dx \\
&\qquad + \frac{k}{m}\biggl\{\biggl(\frac{k}{m}\biggr)^{2\lambda_1(1-\zeta)-\epsilon}+\frac{1}{m}\biggl(\frac{k}{m}\biggr)^{\lambda_1(1-\zeta)-\epsilon}\biggr\}+o(m^{-2}) \\
&= O\biggl(\frac{k}{m}\max\biggl\{\frac{k(M-m)^2}{m^3},\frac{\log m}{m},\biggl(\frac{k}{m}\biggr)^{1+\frac{2(2\wedge\beta)}{d}},\biggl(\frac{k}{m}\biggr)^{2\lambda_1(1-\zeta)-\epsilon}\biggr\}\biggr) \\
&= o\biggl(\max\biggl\{\frac{(M-m)^2}{m^{5/2}},\frac{1}{m^{3/2}}\biggr\}\biggr).
\end{aligned}
$$

Now for $j=1,\ldots,m$ we have

$$
\begin{aligned}
&\mathbb{E}\{(D_m-D_m^{(j)})^2|M\} = \mathbb{E}\{(D_m-D_m^{(1)})^2|M\} \\
&\quad\le 2\mathbb{E}\biggl[\biggl\{\frac{1}{m}\sum_{i=2}^m\biggl(\phi_{X_i}^g\biggl(\frac{k}{mV_d\rho_{(k),i,m}^d}\biggr)-\phi_{X_i}^g\biggl(\frac{k}{mV_d(\rho_{(k),i,m}^{(1)})^d}\biggr) \\
&\qquad\qquad -\phi_{X_i}^g\biggl(\frac{k}{mV_d\rho_{(k),i,M}^d}\biggr)+\phi_{X_i}^g\biggl(\frac{k}{mV_d(\rho_{(k),i,M}^{(1)})^d}\biggr)\biggr)\biggr\}^2\biggm| M\biggr]+o(m^{-2}).
\end{aligned}
$$

Write $\rho_i^{(-1)}$ for the kth nearest neighbour distance of X_i in the sample $X_2,X_3,\ldots,X_M$. The ith term in the above sum is equal to zero unless $\{X_{m+1},\ldots,X_M\}\cap B_{X_i}(\rho_i^{(-1)})\neq\emptyset$ and either $X_1\in B_{X_i}(\rho_i^{(-1)})$ or $X_1'\in B_{X_i}(\rho_i^{(-1)})$. Thus, by similar arguments to those used in the proof of Proposition 11, splitting up into the cases $X_2\in\mathcal{X}_{m,f}$ and $X_2\notin\mathcal{X}_{m,f}$, we have

$$
\begin{aligned}
&\mathbb{E}\{(D_m-D_m^{(j)})^2|M\} \\
&\lesssim \biggl|\mathbb{E}\biggl[\mathbb{1}_{\{\|X_2-X_1\|\le\rho_2^{(-1)},\|X_3-X_1\|\le\rho_3^{(-1)}\}}\biggl\{\phi_{X_2}^g\bigl(\widehat{f}_{(k),2,m}\bigr)-\phi_{X_2}^g\biggl(\frac{M}{m}\widehat{f}_{(k),2,M}\biggr)\biggr\} \\
&\qquad\qquad\times\biggl\{\phi_{X_3}^g\bigl(\widehat{f}_{(k),3,m}\bigr)-\phi_{X_3}^g\biggl(\frac{M}{m}\widehat{f}_{(k),3,M}\biggr)\biggr\}\biggm| M\biggr]\biggr| \\
&\qquad + \frac{1}{m}\mathbb{E}\biggl[\mathbb{1}_{\{\|X_1-X_2\|\le\rho_2^{(-1)}\}}\biggl\{\phi_{X_2}^g\bigl(\widehat{f}_{(k),2,m}\bigr)-\phi_{X_2}^g\biggl(\frac{M}{m}\widehat{f}_{(k),2,M}\biggr)\biggr\}^2\biggm| M\biggr]+o(m^{-2}) \\
&\lesssim |M-m|\biggl\{\biggl(\frac{k}{m}\biggr)^3+\frac{1}{m}\biggl(\frac{k}{m}\biggr)^2\biggr\}\int_{\mathcal{X}_{m,f}} f(x)^{1+2\kappa_1}g(x)^{2\kappa_2} \\
&\qquad\qquad\times\max\biggl\{\frac{(M-m)^2}{m^2},\frac{\log m}{k},\biggl(\frac{kM_\beta(x)^d}{mf(x)}\biggr)^{\frac{2(2\wedge\beta)}{d}}\biggr\}\,dx
\end{aligned}
$$

$$
\begin{aligned}
&+|M-m|\Big(\frac{k}{m}\Big)^2\bigg\{\Big(\frac{k}{m}\Big)^{2\lambda_1(1-\zeta)-\epsilon}+\frac{1}{m}\Big(\frac{k}{m}\Big)^{\lambda_1(1-\zeta)-\epsilon}\bigg\}+o(m^{-2})\\
&\lesssim |M-m|\Big(\frac{k}{m}\Big)^2\max\bigg\{\frac{k(M-m)^2}{m^3},\frac{\log m}{m},\Big(\frac{k}{m}\Big)^{1+\frac{2(2\wedge\beta)}{d}},\Big(\frac{k}{m}\Big)^{2\lambda_1(1-\zeta)-\epsilon}\bigg\}+o(m^{-2})\\
&=o\bigg(\max\bigg\{\frac{1}{m^2},\frac{|M-m|^3}{m^{7/2}}\bigg\}\bigg).
\end{aligned}
$$

It now follows by the Efron–Stein inequality that, on the event A_m, we have

$$
\mathrm{Var}\{T_m^{(1)}-T_m^{(1),\mathrm{p}}|M\}=o\bigg(\max\bigg\{\frac{|M-m|^3}{m^{5/2}},\frac{1}{m}\bigg\}\bigg).
$$

We now bound the contribution from the event A_m^c. We will use the fact that for $x\geq 0$ we have

$$
\mathbb{P}\Big(\Big|\frac{M}{m}-1\Big|\geq x\Big)\leq 2\exp\Big(-\frac{mx^2}{2(1+x)}\Big).
$$

It follows from this that

$$
\mathbb{P}(A_m^c)\leq 2\exp\Big(-\frac{m}{4\log^2(em)}\Big).
$$

Moreover, we have for any $a\geq 1$ that

$$
\begin{aligned}
\mathbb{E}\Big[\Big(\frac{M}{m}\Big)^a\mathbb{1}_{A_m^c}\Big]&\leq\int_0^\infty\mathbb{P}\bigg(\Big|\frac{M}{m}-1\Big|\geq\max\Big\{\frac{1}{\log(em)},x^{1/a}-1\Big\}\bigg)\,dx\\
&\leq 2^a\mathbb{P}(A_m^c)+2a\int_2^\infty y^{a-1}e^{-my/8}\,dy\\
&\leq 2^a\mathbb{P}(A_m^c)+2a(a-1)\log(16/a)\int_2^\infty e^{-(m-1/2)y/8}\,dy\\
&\leq 2^{a+1}\exp\Big(-\frac{m}{4\log^2(em)}\Big)+\frac{2a(a-1)\log(16/a)}{m-1/2}e^{-\frac{2m-1}{8}}.
\end{aligned}
$$

It now follows using Lemma S8 that, when $\log(em)>2d\kappa_1^-/\alpha$, we have

$$
\begin{aligned}
&\mathbb{E}\{(T_m^{(1)}-T_m^{(1),\mathrm{p}})^2\mathbb{1}_{A_m^c}\}\\
&\leq 3\mathbb{P}(A_m^c)\mathbb{E}\{(T_m^{(1)})^2\}+3\mathbb{E}\bigg(\Big|\frac{M}{m}-1\Big|\mathbb{1}_{A_m^c}\bigg)\bigg|\int_{\mathcal{X}}f(x)\{\phi_x+(f\phi_{10})_x\}\,dx\bigg|\\
&\qquad+3\mathbb{E}\bigg\{\frac{M}{m}\phi_{X_1}^g\bigg(\frac{k}{mV_d\rho_{(k),1,M}^d}\bigg)^2\mathbb{1}_{A_m^c\cap\{M\geq(k+1)\log(em)\}}\bigg\}\\
&\lesssim\mathbb{E}\bigg[\frac{M}{m}\mathbb{1}_{A_m^c\cap\{M\geq(k+1)\log(em)\}}\int_{\mathcal{X}}f(x)g(x)^{2\kappa_2}\int_0^1\bigg\{\frac{k}{mV_dh_{x,f}^{-1}(s)^d}\bigg\}^{2\kappa_1}\mathrm{B}_{k,M-k}(s)\,ds\,dx\bigg]\\
&\qquad\mathbb{P}(A_m^c)+\mathbb{E}\bigg(\frac{M}{m}\mathbb{1}_{A_m^c}\bigg)\\
&\lesssim\mathbb{E}\bigg[\frac{M}{m}\mathbb{1}_{A_m^c}\int_{\mathcal{X}}f(x)g(x)^{2\kappa_2}\max\bigg\{1+\|x\|^{2d\kappa_1^-},\Big(\frac{M}{m}\Big)^{2\kappa_1^+}\bigg\}\,dx\bigg]+\mathbb{P}(A_m^c)+\mathbb{E}\bigg(\frac{M}{m}\mathbb{1}_{A_m^c}\bigg)
\end{aligned}
$$

(S44)

$$
\lesssim\mathbb{P}(A_m^c)+\mathbb{E}\bigg\{\Big(\frac{M}{m}\Big)^{1+2\kappa_1^+}\mathbb{1}_{A_m^c}\bigg\}=o(1/m).
$$

Hence,

$$\begin{aligned}\mathbb{E}\{(T_m^{(1)}-T_m^{(1),\mathrm{p}})^2\} &= \mathbb{E}\big[\mathrm{Var}\big(T_m^{(1)}-T_m^{(1),\mathrm{p}}\mathbb{1}_{A_m}\bigm| M\big)\mathbb{1}_{A_m} \\ &\qquad + \big\{\mathbb{E}\big(T_m^{(1)}-T_m^{(1),\mathrm{p}}\bigm| M\big)\big\}^2\mathbb{1}_{A_m}\big] + \mathbb{E}\big\{(T_m^{(1)}-T_m^{(1),\mathrm{p}})^2\mathbb{1}_{A_m^c}\big\} \\ &= o\biggl(\mathbb{E}\max\biggl\{\frac{|M-m|^3}{m^{5/2}},\frac{1}{m}\biggr\}\biggr) = o(1/m),\end{aligned}$$

as required.

We now turn our attention to $T_n^{(2)}-T_n^{(2),\mathrm{p}}$, for which similar arguments apply. We write $\phi_x^f(\cdot):=\phi(f(x),\cdot)$. We have, on the event $|n/N-1|\leq 1/L$ and when $n\geq(1+1/L)k\log(en)$, that

$$\begin{aligned}\mathbb{E}\big\{T_n^{(2)}-T_n^{(2),\mathrm{p}}|N\big\} &= \mathbb{E}\phi_{X_1}^f\biggl(\frac{k}{nV_d\rho_{(k),1,n}^d}\biggr) - \mathbb{E}\biggl\{\phi_{X_1}^f\biggl(\frac{k}{nV_d\rho_{(k),1,N}^d}\biggr)\biggm| N\biggr\} \\ &\qquad + \Bigl(\frac{N}{n}-1\Bigr)\int_{\mathcal{X}} f(x)(g\phi_{01})_x\,dx \\ &= \mathbb{E}\biggl\{\phi_{X_1}^f\biggl(\frac{k}{NV_d\rho_{(k),1,N}^d}\biggr) - \phi_{X_1}^f\biggl(\frac{k}{nV_d\rho_{(k),1,N}^d}\biggr)\biggm| N\biggr\} \\ &\qquad + \Bigl(\frac{N}{n}-1\Bigr)\int_{\mathcal{X}} f(x)(g\phi_{01})_x\,dx + o\biggl(n^{-1/2}+\Bigl|\frac{N}{n}-1\Bigr|\biggr) \\ &= o\biggl(n^{-1/2}+\Bigl|\frac{N}{n}-1\Bigr|\biggr).\end{aligned}$$

To bound the conditional variance of $T_n^{(2)}-T_n^{(2),\mathrm{p}}$ on the event that $|N/n-1|\leq 1/\log(en)$, we again appeal to the Efron–Stein inequality. Similar to before, for $\ell\geq k$ and $x\in\mathcal{X}$, we define

$$\widehat{g}_{(k),\ell}(x) := \frac{k}{\ell V_d\rho_{(k),\ell}^d(x)}.$$

We redefine

$$D_n := \int_{\mathcal{X}} f(x)\biggl\{\phi_x^f\bigl(\widehat{g}_{(k),n}(x)\bigr) - \phi_x^f\biggl(\frac{N}{n}\widehat{g}_{(k),N}(x)\biggr)\biggr\}\,dx.$$

Similarly to above, letting $Y_1',Y_2',\ldots$ be independent copies of $Y_1,Y_2,\ldots$, for $j\in[N]$ write $D_n^{(j)}$ for the value of D_n when it is computed on $Y_1,\ldots,Y_{j-1},Y_j',Y_{j+1},\ldots,Y_N$ instead of $Y_1,\ldots,Y_N$. On the event $\{N>n\}$, for $j=n+1,\ldots,N$, splitting up into the cases $X_1\in\mathcal{X}_{n,g}$ and $X_1\notin\mathcal{X}_{n,g}$, we have

$$\begin{aligned}\mathbb{E}\big\{(D_n-D_n^{(j)})^2|N\big\} &\leq 4\mathbb{E}\biggl[\biggl\{\int_{\{x:\|Y_j-x\|\leq\rho_{(k),N}(x)\}} f(x)\phi_x^f\Bigl(\frac{N}{n}\widehat{g}_{(k),N}(x)\Bigr)\,dx\biggr\}^2\biggm| N\biggr] \\ &= 4\mathbb{E}\biggl[\mathbb{1}_{\{\|Y_j-X_1\|\leq\rho_{(k),N}(X_1)\}}\phi_{X_1}^f\biggl(\frac{N}{n}\widehat{g}_{(k),N}(X_1)\biggr) \\ &\qquad\times\mathbb{1}_{\{\|Y_j-X_2\|\leq\rho_{(k),N}(X_2)\}}\phi_{X_2}^f\biggl(\frac{N}{n}\widehat{g}_{(k),N}(X_2)\biggr)\biggm| N\biggr] \\ &\lesssim \Bigl(\frac{k}{n}\Bigr)^2\int_{\mathcal{X}_{n,g}} f(x)^{2+2\kappa_1}g(x)^{2\kappa_2-1}\,dx + \Bigl(\frac{k}{n}\Bigr)^{1+2\lambda_2(1-\zeta)-\epsilon} = o\bigl(n^{-3/2}\bigr).\end{aligned}$$

On the other hand, for $j \in [n]$ and on the same event $\{N > n\}$, we have

$$\begin{aligned}
&\mathbb{E}\{(D_n - D_n^{(j)})^2 | N\} \\
&\leq 4\mathbb{E}\biggl[\biggl\{\int_{\{x:\|Y_j - x\| \leq \rho_{(k),N}(x)\}} f(x)\biggl(\phi_x^f\bigl(\widehat{g}_{(k),n}(x)\bigr) - \phi_x^f\biggl(\frac{N}{n}\widehat{g}_{(k),N}(x)\biggr)\biggr)\,dx\biggr\}^2 \biggm| N\biggr] \\
&\lesssim (N-n)\Bigl(\frac{k}{n}\Bigr)^3 \int_{\mathcal{X}_{n,g}} f(x)^{2+2\kappa_1} g(x)^{2\kappa_2 - 1}\,dx + (N-n)\Bigl(\frac{k}{n}\Bigr)^{2+2\lambda_2(1-\zeta)-\epsilon} \\
&= o\biggl(\frac{|N-n|}{n^{5/2}}\biggr).
\end{aligned}$$

On the event $\{N < n\}$, the same final bound holds, and it follows by the Efron–Stein inequality that, on the event that $|N/n - 1| \leq 1/\log(en)$, we have

$$\mathrm{Var}\bigl(T_n^{(2)} - T_n^{(2),\mathrm{p}} \mid N\bigr) = o\biggl(\frac{|N-n|}{n^{3/2}}\biggr).$$

Now, similarly to (S44), redefining $A_n := \{|N/n - 1| \leq 1/\log(en)\}$ we have

$$\mathbb{E}\bigl\{(T_n^{(2)} - T_n^{(2),\mathrm{p}})^2 \mathbb{1}_{A_n^c}\bigr\} \lesssim \mathbb{P}(A_n^c) + \mathbb{E}\biggl\{\Bigl(\frac{N}{n}\Bigr)^{1+2\kappa_2^+} \mathbb{1}_{A_n^c}\biggr\} = o(1/n),$$

and the result follows. □

Our second preparatory result provides a convenient partition of (minor modifications of) $\mathcal{X}_{m,f}$ and $\mathcal{X}_{n,g}$ so that, under the Poisson sampling scheme, the k-nearest neighbour distances of points in distant pieces are roughly independent.

PROPOSITION S2. *Let $f \in \mathcal{F}_d$ be $\underline{\beta} := (\lceil \beta \rceil - 1)$-times differentiable. Then there exists $n_0 = n_0(d, \beta)$ such that, for all $n \geq n_0$ and $k \in [3, n/\log n)$, we can find a partition $\{C_j : j \in 1, \ldots, V_n\}$ of $\mathcal{X}_n := \{x : f(x)/M_{f,\beta}(x)^d \geq (k/n)\log^2 n\}$ and points $\{x_j : j = 1, \ldots, V_n\}$ in $\widetilde{\mathcal{X}}_n := \{x : f(x)/M_{f,\beta}(x)^d \geq (k/n)\log^{7/4} n\}$ satisfying the following properties for each $j = 1, \ldots, V_n$:*

(i) *we have $C_j \subseteq B_{x_j}\Bigl(3\bigl(\frac{k\log n}{nV_d f(x_j)}\bigr)^{1/d}\Bigr)$;*

(ii) *we have $\Bigl|\Bigl\{j' = 1, \ldots, V_n : \mathrm{dist}(C_j, C_{j'}) \leq 4\bigl(\frac{k}{nV_d f(x_j)}\bigr)^{1/d}\Bigr\}\Bigr| \leq 2^{2+4d}\log n.$*

PROOF OF PROPOSITION S2. Let $\{x_j : j = 1, \ldots, V_n\}$ be a Poisson process on $\widetilde{\mathcal{X}}_n$ with intensity function $nf(\cdot)/k$, and let P denote the corresponding Poisson random measure. Writing $\mathrm{sargmin}(S)$ for the smallest element of an ordered set $\mathrm{argmin}(S)$, we may partition $\mathcal{X}_n$ into the associated (random) Voronoi cells $\{C_j : j = 1, \ldots, V_n\}$, where $C_j := \{x \in \mathcal{X}_n : \mathrm{sargmin}_{j'=1,\ldots,V_n} \|x - x_{j'}\| = j\}$. We proceed by showing that, for n and k sufficiently large, there is an event of positive probability on which $\{C_j : j = 1, \ldots, V_n\}$ and $\{x_j : j = 1, \ldots, V_n\}$ satisfy (i) and (ii), and we therefore deduce the existence of such a partition. First, let $z_1, \ldots, z_N \in \mathcal{X}_n$ be such that

$$\|z_i - z_j\| \geq h_{z_i,f}^{-1}(k/n) + h_{z_j,f}^{-1}(k/n) =: r(z_i, z_j)$$

for all $i \neq j$, and such that $\sup_{x \in \mathcal{X}_n} \min_{j=1,\ldots,N} \|x - z_j\|/r(x, z_j) < 1$. (We can construct this set inductively: first, choose $z_1 \in \mathcal{X}_n$ arbitrarily. If the second condition is not satisfied once $z_1, \ldots, z_N$ have been defined, then there exists $x \in \mathcal{X}_n$ such that $\|x - z_j\| \geq r(x, z_j)$ for all $j=1,\ldots,N$ and we can set $z_{N+1} := x$.) For all $i \neq j$, the intersection $B_{z_i}\big(h^{-1}_{z_i,f}(k/n)\big) \cap B_{z_j}\big(h^{-1}_{z_j,f}(k/n)\big)$ has Lebesgue measure zero and thus

$$1 \geq \sum_{j=1}^{N} h_{z_j,f}\Big(h^{-1}_{z_j,f}(k/n)\Big) = \frac{Nk}{n}.$$

In particular, $N \leq n/k$.

We now show that if $x \in \mathcal{X}_n$ is such that $\|x-z\| < r(x,z)$ for some $z \in \{z_1,\ldots,z_N\} \subseteq \mathcal{X}_n$ then $f(x) \approx f(z)$. Suppose initially that $r_2 := \{M_{f,\beta}(z)^d \log n\}^{-1/d} \leq \|x-z\| < r(x,z)$. Then, writing $r_1 := \|x-z\| - r_2/2$, writing $\bar{x}$ for the point on the line segment between x and z such that $\|\bar{x}-z\| = r_2$ and writing $I(s) := \int_0^s \mathrm{B}_{(d+1)/2,1/2}(t)\,dt$, we have by Lemma S5 that, for $n \geq n_0(d,\beta)$ sufficiently large,

$$\begin{aligned}
\int_{B_x(r_1)} f(w)\,dw &\geq \int_{B_x(r_1)\cap B_z(r_2)} f(w)\,dw \geq \frac{f(z)}{2}\mu_d\big(B_x(r_1)\cap B_z(r_2)\big) \\
&\geq \frac{f(z)}{2}\mu_d\big(B_{\bar{x}}(r_2/2)\cap B_z(r_2)\big) = \frac{V_d f(z)}{2}\Big\{\Big(\frac{r_2}{2}\Big)^d I(15/16) + r_2^d I(15/64)\Big\} \\
&\geq \frac{V_d}{2^{d+1}} I(15/16)\frac{k\log n}{n} \geq \frac{k}{n}.
\end{aligned}$$

It follows, by Lemma S4 and the fact that $z \in \mathcal{X}_n$, that there exists $n_1 = n_1(d,\beta) \geq n_0$, such that for $n \geq n_1$,

$$\|x-z\| \leq r_1 + h^{-1}_{z,f}(k/n) \leq r_1 + 2\Big(\frac{k}{nV_d f(z)}\Big)^{1/d} \leq r_1 + \frac{r_2}{4} = \|x-z\| - \frac{r_2}{4},$$

which is a contradiction. Thus, for $n \geq n_1$ we have that $\|x-z\| \leq r_2$. In particular, by Lemma S5, for $x,z \in \mathcal{X}_n$ with $\|x-z\| < r(x,z)$, and for $n \geq n_1$, we have that

$$\Big|\frac{f(x)}{f(z)} - 1\Big| \leq \frac{2}{\log^{(1\wedge\beta)/d} n}. \tag{S45}$$

To establish (i), first we define the event

$$\Omega_0 := \bigcap_{j=1}^{N}\Big\{P\Big\{B_{z_j}\Big(h^{-1}_{z_j,f}(k\log n/n)\Big)\Big\} \geq 1\Big\}.$$

By Lemmas S4 and S5 and very similar arguments to those leading up to (S78), there exists $n_2 = n_2(d,\beta) \geq n_1$ such that $B_{z_j}\big(h^{-1}_{z_j,f}(k\log n/n)\big) \subseteq \widetilde{\mathcal{X}}_n$ for all $n \geq n_2$ and $j = 1,\ldots,V_n$. Then, for $n \geq n_2$ we have that

$$\mathbb{P}(\Omega_0^c) \leq N\exp\Big(-\frac{n}{k}\frac{k\log n}{n}\Big) \leq \frac{1}{k}.$$

Let $j \in \{1,\ldots,V_n\}$ be given, and suppose that $x \in C_j$. Let z be in our covering set such that $\|x-z\| < r(x,z)$ and, on the event Ω_0, let $j' \in \{1,\ldots,V_n\}$ be such that $\|x_{j'}-z\| \leq h^{-1}_{z,f}(k\log n/n)$. By (S45), Lemma S4 and Lemma S5, there exists $n_3 = n_3(d,\beta) \geq n_2$ such that, for $n \geq n_3$, we have that $h^{-1}_{z,f}(k\log n/n) \leq \frac{3}{2}\big(\frac{k\log n}{nV_d f(z)}\big)^{1/d}$ and hence that

$$\begin{aligned}
\|x_{j'}-x\| \leq \|x_{j'}-z\| + \|z-x\| &< h^{-1}_{z,f}(k\log n/n) + h^{-1}_{z,f}(k/n) + h^{-1}_{x,f}(k/n) \\
&\leq 2\Big(\frac{k\log n}{nV_d f(x_{j'})}\Big)^{1/d}.
\end{aligned}$$

If $j'=j$ then we are done, so suppose instead that $\|x-x_j\| \leq \|x-x_{j'}\|$. Then

$$\|x_j - x_{j'}\| \leq 2\|x-x_{j'}\| \leq 4\Big(\frac{k\log n}{nV_d f(x_{j'})}\Big)^{1/d}$$

so we can use Lemma S5 to argue that $f(x_j) \approx f(x_{j'})$. In particular, there exists $n_4 = n_4(d,\beta) \geq n_3$ such that, for $n \geq n_4$ we have that

$$\|x-x_j\| \leq \|x-x_{j'}\| \leq 2\Big(\frac{k\log n}{nV_d f(x_{j'})}\Big)^{1/d} \leq 3\Big(\frac{k\log n}{nV_d f(x_j)}\Big)^{1/d}.$$

So, for $n \geq n_4$, we have that (i) holds on Ω_0.

Now, by Lemma S5, there exists $n_5 = n_5(d,\beta) \geq n_4$ such that, for $n \geq n_5$ we have that $\frac{n}{k} h_{z_j,f}\big(16(\frac{k\log n}{nV_d f(z_j)})^{1/d}\big) \leq 2^{1+4d}\log n$ for all $j \in \{1,\ldots,N\}$, and hence, by Bennett's inequality, that the event

$$\Omega_1 := \bigcap_{j=1}^{N} \Big\{ P\Big\{B_{z_j}\Big(16\Big(\frac{k\log n}{nV_d f(z_j)}\Big)^{1/d}\Big)\Big\} \leq 2^{2+4d}\log n \Big\}$$

satisfies

$$\mathbb{P}(\Omega_1^c) \leq N\exp\bigg(-\frac{(2^{2+4d}\log n - 2^{1+4d}\log n)^2}{2^{3+4d}\log n}\bigg) \leq \frac{n}{k}\exp(-2^{4d-1}\log n) \leq \frac{1}{k}.$$

Now, on Ω_0, if $\mathrm{dist}(C_j, C_{j'}) \leq 4\big(\frac{k}{nV_d f(x_j)}\big)^{1/d}$ then we must have

$$\|x_j - x_{j'}\| \leq 4\Big(\frac{k}{nV_d f(x_j)}\Big)^{1/d} + 3\Big(\frac{k\log n}{nV_d f(x_j)}\Big)^{1/d} + 3\Big(\frac{k\log n}{nV_d f(x_{j'})}\Big)^{1/d}. \tag{S46}$$

Using Lemma S4, there exists $n_6 = n_6(d,\beta) \geq n_5$ such that $\|x_j - x_{j'}\| \leq 6h^{-1}_{x_j,f}(k\log n/n) + 6h^{-1}_{x_{j'},f}(k\log n/n)$ for $n \geq n_6$ and hence, by a very similar argument to that leading up to (S45), we have that $|f(x_{j'})/f(x_j) - 1| \leq 2\log^{-(1\wedge\beta)/(2d)} n$ for $n \geq n_6$. Thus, writing z_j^* for an element of our covering set with $\|x_j - z_j^*\| < r(x_j, z_j^*)$, there exists $n_7 = n_7(d,\beta) \geq n_6$ such that, on $\Omega_0 \cap \Omega_1$, for all $n \geq n_7$ we have that

$$\begin{aligned}
&\Big|\Big\{j' \in V_n : \mathrm{dist}(C_j, C_{j'}) \leq 4\Big(\frac{k}{nV_d f(x_j)}\Big)^{1/d}\Big\}\Big| \\
&\qquad \leq \Big|\Big\{j' \in V_n : \|x_j - x_{j'}\| \leq 8\Big(\frac{k\log n}{nV_d f(x_j)}\Big)^{1/d}\Big\}\Big| \\
&\qquad \leq \Big|\Big\{j' \in V_n : \|z_j^* - x_{j'}\| \leq 16\Big(\frac{k\log n}{nV_d f(x_j)}\Big)^{1/d}\Big\}\Big| \\
&\qquad \leq 2^{2+4d}\log n
\end{aligned}$$

for all $j \in V_n$. This establishes that, for $n \geq n_7$, with probability at least $1 - 2/k$ we have that both (i) and (ii) hold. Thus, since $k \geq 3$, there is a positive probability of both (i) and (ii) holding simultaneously and we can deduce the existence of the required partition. □

PROOF OF THEOREM 3. We start by linearising our unweighted estimator. Consider

$$\begin{aligned}
E_{m,n} &:= \frac{1}{m}\sum_{i=1}^{m}\Big\{\phi\big(\widehat{f}_{(k_X),i}, \widehat{g}_{(k_Y),i}\big) - \phi\big(\widehat{f}_{(k_X),i}, g(X_i)\big) - \phi\big(f(X_i), \widehat{g}_{(k_Y),i}\big) + \phi_{X_i}\Big\} \\
&= \frac{1}{m}\sum_{i=1}^{m}\phi^*\big(\widehat{f}_{(k_X),i}, \widehat{g}_{(k_Y),i}\big)
\end{aligned}$$

with $\phi^*(u,v) := \phi(u,v) - \phi\big(u, g(x)\big) - \phi\big(f(x), v\big) + \phi\big(f(x), g(x)\big)$. This is of the same form as the estimators we have already considered, and we have $\phi^*(f(x), g(x)) \equiv \phi^*_{10}(f(x), g(x)) \equiv \phi^*_{01}(f(x), g(x)) \equiv 0$. Therefore, by very similar arguments to those used in the proof of Proposition 11, we have that $\mathrm{Var}(E_{m,n}) = o(1/m + 1/n)$. Further, we have that

$$\mathbb{E}\biggl[\mathrm{Var}\biggl(\frac{1}{m}\sum_{i=1}^m \bigl\{\phi\bigl(f(X_i), \widehat{g}_{(k_Y),i}\bigr) - \phi_{X_i}\bigr\} \,\bigg|\, Y_1, \ldots, Y_n\biggr)\biggr]$$
$$= \frac{1}{m}\mathbb{E}\Bigl[\mathrm{Var}\Bigl(\phi\bigl(f(X_1), \widehat{g}_{(k_Y),1}\bigr) - \phi_{X_1} \,\Big|\, Y_1, \ldots, Y_n\Bigr)\Bigr]$$
$$\leq \frac{1}{m}\mathbb{E}\Bigl[\bigl\{\phi\bigl(f(X_1), \widehat{g}_{(k_Y),1}\bigr) - \phi_{X_1}\bigr\}^2\Bigr] = o(1/m).$$

Recalling the definitions of $T_m^{(1)}$ and $T_n^{(2)}$ in (12), we therefore have that

$$\mathrm{Var}(\widehat{T}_{m,n} - T_m^{(1)} - T_n^{(2)}) \leq 2\mathrm{Var}\biggl(T_n^{(2)} - \frac{1}{m}\sum_{i=1}^m \bigl\{\phi\bigl(f(X_i), \widehat{g}_{(k_Y),i}\bigr) - \phi_{X_i}\bigr\}\biggr) + 2\mathrm{Var}(E_{m,n})$$
$$= 2\mathrm{Var}\biggl(\mathbb{E}\biggl\{T_n^{(2)} - \frac{1}{m}\sum_{i=1}^m \bigl\{\phi\bigl(f(X_i), \widehat{g}_{(k_Y),i}\bigr) - \phi_{X_i}\bigr\} \,\bigg|\, Y_1, \ldots, Y_n\biggr\}\biggr) + o(1/m + 1/n)$$
$$= o(1/m + 1/n).$$

It now follows immediately from Proposition 4 that $\mathrm{Var}(\widehat{T}_{m,n} - T_m^{(1),\mathrm{p}} - T_n^{(2),\mathrm{p}}) = o(1/m + 1/n)$. Noting that $T_m^{(1),\mathrm{p}}$ depends only on $M, X_1, \ldots, X_m$ and $T_n^{(2),\mathrm{p}}$ depends only on $N, Y_1, \ldots, Y_n$ (so they are independent), we now proceed to establish the asymptotic normality of these two random variables separately, and then the result will follow.

We start with $T_m^{(1),\mathrm{p}}$, and adopt the notation of Proposition 4. Define the events $A_{i,m} := \{h_{X_i,f}(\rho_{(k),i,M}) \in \mathcal{I}_{m,X}\}$ for $i = 1, \ldots, M$, similarly to in (S72), and define

$$\mathcal{X}_{m,f} := \biggl\{x : f(x) M_\beta(x)^{-d} \geq \frac{k_X \log^2 m}{m}\biggr\}.$$

By separately considering the event that $|M/m - 1| \leq 1/k_X$ and its complement we may use similar arguments to those in Proposition 4 and Lemma S9 to see that $\mathbb{P}(A_{1,m}^c) = o(m^{-4})$, and moreover that

$$\mathbb{E}\biggl[\mathbb{1}_{A_{1,m}^c} \phi_{X_1}^g\biggl(\frac{M}{m}\widehat{f}_{(k_X),1,M}\biggr)^2\biggr] = o(m^{-4}).$$

Further,

$$\mathbb{E}\biggl[\biggl\{\frac{1}{m}\sum_{i=1}^M \mathbb{1}_{A_{i,m}} \mathbb{1}_{\{X_i \notin \mathcal{X}_{m,f}\}} \phi_{X_i}^g\biggl(\frac{M}{m}\widehat{f}_{(k_X),i,M}\biggr)\biggr\}^2\biggr]$$
$$= \frac{1}{m^2}\mathbb{E}\biggl[M(M-1)\mathbb{1}_{A_{1,m} \cap A_{2,m}} \mathbb{1}_{\{X_1, X_2 \notin \mathcal{X}_{m,f}\}} \phi_{X_1}^g\biggl(\frac{M}{m}\widehat{f}_{(k_X),1,M}\biggr) \phi_{X_2}^g\biggl(\frac{M}{m}\widehat{f}_{(k_X),2,M}\biggr)\biggr]$$
$$+ \frac{1}{m^2}\biggl[M\mathbb{1}_{A_{1,m}} \mathbb{1}_{\{X_1 \notin \mathcal{X}_{m,f}\}} \phi_{X_1}^g\biggl(\frac{M}{m}\widehat{f}_{(k_X),1,M}\biggr)\biggr] = o(1/m).$$

Writing

$$\widetilde{T}_m^{(1),\mathrm{p}} := \frac{1}{m}\sum_{i=1}^M \mathbb{1}_{A_{i,m}} \mathbb{1}_{\{X_i \in \mathcal{X}_{m,f}\}}\biggl\{\phi_{X_i}^g\biggl(\frac{M}{m}\widehat{f}_{(k_X),i,M}\biggr) - \int_{\mathcal{X}} f(x)\{\phi_x + (f\phi_{10})_x\}\,dx\biggr\},$$

we may now see that $\mathrm{Var}(T_m^{(1),\mathrm{p}} - \widetilde{T}_m^{(1),\mathrm{p}}) = o(1/m)$. Letting $\{C_j : j \in 1,\ldots,V_m\}$ denote a partition of $\mathcal{X}_{m,f}$ as in the statement of Proposition S2, and writing $\mathcal{X}_{m,f}^{(j)} := C_j \cap \mathcal{X}_{m,f}$, for $j = 1,\ldots,V_m$ define

$$W_j := \frac{1}{m}\sum_{i=1}^{M} \mathbb{1}_{A_{i,m}} \mathbb{1}_{\{X_i \in \mathcal{X}_{m,f}^{(j)}\}} \biggl\{ \phi_{X_i}^g\biggl(\frac{M}{m}\widehat{f}_{(k_X),i,M}\biggr) - \int_{\mathcal{X}} f(x)\{\phi_x + (f\phi_{10})_x\}\,dx \biggr\}$$

so that $\widetilde{T}_m^{(1),\mathrm{p}} = \sum_{j=1}^{V_m} W_j$. For $j,j' = 1,\ldots,V_m$, write $j \sim j'$ if W_j and $W_{j'}$ are dependent. Because we are working on the events $A_{i,m}$, the random variable W_j is only a function of those X_i that lie within distance $\sup_{x \in \mathcal{X}_{m,f}^{(j)}} h_{x,f}^{-1}(a_{m,X}^+)$ of the set C_j. Hence, by the independence properties of Poisson processes, we can only have $j \sim j'$ if

$$\mathrm{dist}(C_j, C_{j'}) \le \sup_{x \in \mathcal{X}_{m,f}^{(j)}} h_{x,f}^{-1}(a_{m,X}^+) + \sup_{x' \in \mathcal{X}_{m,f}^{(j')}} h_{x',f}^{-1}(a_{m,X}^+).$$

Hence, by Lemma S5 and property (i) of the partition and arguing as after (S46), there exists $m_0 = m_0(d,\vartheta)$ such that for $m \ge m_0$, we can only have $j \sim j'$ if

$$\mathrm{dist}(C_j, C_{j'}) \le \sup_{x \in \mathcal{X}_{m,f}^{(j)}} \biggl(\frac{3k_X}{2mV_d f(x)}\biggr)^{1/d} + \sup_{x' \in \mathcal{X}_{m,f}^{(j')}} \biggl(\frac{3k_X}{2mV_d f(x')}\biggr)^{1/d} \le 4\biggl(\frac{k_X}{mV_d f(x_j)}\biggr)^{1/d},$$

where $\{x_j : j = 1,\ldots,V_m\}$ are the points associated to our partition given in Proposition S2. By property (ii) of our partition, then, for each $j = 1,\ldots,V_m$, we have $|\{j' : j' \sim j\}| \le 2^{2+4d}\log m$. For $j = 1,\ldots,V_m$ and $p \in \mathbb{N}$, we write $L_j^{(p)}$ for the number of connected subsets of $\{1,\ldots,V_m\}$ (with edge relations defined by $\sim$) of cardinality at most p containing j. Then

$$L_j^{(p)} \le 2^{(p-1)(2+4d)} \log^{p-1} m$$

for $p = 3,4$. Now, by Lemma S5 and property (i) of our partition, for any $j = 1,\ldots,V_m$ we have

$$\sup_{x \in C_j} \max\biggl\{\biggl|\frac{f(x)}{f(x_j)} - 1\biggr|, \biggl|\frac{g(x)}{g(x_j)} - 1\biggr|\biggr\} \le \frac{2 \times 3^{1\wedge\beta}}{(V_d \log^{3/4} m)^{(1\wedge\beta)/d}}. \tag{S47}$$

Moreover, by very similar methods to those used in the proof of Proposition 11, we may see that

$$\mathrm{Var}(\widetilde{T}_m^{(1),\mathrm{p}}) = \mathrm{Var}(T_m^{(1)}) + o(1/m) = \frac{v_1}{m} + o(1/m). \tag{S48}$$

Hence, using (S47), (S48) and the facts that $v_1 \ge 1/C$ and $p_{m,f,(j)} := \mathbb{P}(X_1 \in \mathcal{X}_{m,f}^{(j)}) \lesssim 9(k_X/m)\log m$, we have that for $p = 3,4$,

$$\begin{aligned}
&\frac{1}{\mathrm{Var}^{p/2}(\widetilde{T}_m^{(1)})} \sum_{j=1}^{V_m} L_j^{(p)} \mathbb{E}\{|W_j - \mathbb{E}W_j|^p\} \\
&\quad\lesssim m^{-p/2}\log^{p-1} m \sum_{j=1}^{V_m} \mathbb{E}\biggl\{\biggl[\sum_{i=1}^{m} \mathbb{1}_{A_i^X} \mathbb{1}_{\{X_i \in \mathcal{X}_{m,f}^{(j)}\}} \biggl\{\biggl|\phi_{X_i}^g\biggl(\frac{M}{m}\widehat{f}_{(k_X),i,M}\biggr)\biggr| + 1\biggr\}\biggr]^p\biggr\} \\
&\quad\lesssim m^{-p/2}\log^{p-1} m \sum_{j=1}^{V_m} f(x_j)^{p\kappa_1} g(x_j)^{p\kappa_2} \bigl\{m^p p_{m,f,(j)}^p + m p_{m,f,(j)}\bigr\} \\
&\quad\lesssim \frac{k_X^{p-1}\log^{2p-2} m}{m^{p/2-1}} \int_{\mathcal{X}} f(x)^{1+p\kappa_1} g(x)^{p\kappa_2}\,dx \to 0.
\end{aligned}$$

It now follows from Theorem 1 of Baldi and Rinott (1989) that

$$d_{\mathrm{K}}\biggl(\mathcal{L}\Bigl(\frac{m^{1/2}\{\widetilde{T}_m^{(1)}-\mathbb{E}\widetilde{T}_m^{(1)}\}}{v_1^{1/2}}\Bigr),N(0,1)\biggr)\to 0.$$

We now take a similar approach to establish the asymptotic normality of $\widetilde{T}_n^{(2)}$. Letting $\{C_j : j=1,\ldots,V_n\}$ denote a partition of $\mathcal{X}_{n,g}$ as in the statement of Proposition S2, we may write $\rho_{(k_Y),Y}(x) := \|Y_{(k_Y)}(x)-x\|$, $A_n := \bigl\{x : h_{x,g}(\rho_{(k_Y),N}(x)) \in \mathcal{I}_{n,Y}\bigr\}$, $\mathcal{X}_{n,g}^{(j)} := C_j \cap \mathcal{X}_{n,g}$, and

$$W_j := \int_{\mathcal{X}_{n,g}^{(j)}\cap A^Y} f(x)\phi_x^f\Bigl(\frac{k_Y}{nV_d\rho_{(k_Y),N}(x)^d}\Bigr)\,dx - \frac{|\{i : Y_i \in \mathcal{X}_{n,g}^{(j)}\}|}{n}\int_{\mathcal{X}} f(x)(g\phi_{01})_x\,dx.$$

Writing $\widetilde{T}_n^{(2),\mathrm{p}} := \sum_{j=1}^{V_n} W_j$ and arguing as above, we can see that $\mathrm{Var}(T_n^{(2),\mathrm{p}} - \widetilde{T}_n^{(2),\mathrm{p}}) = o(1/n)$. By properties (i) and (ii) of our partition we again have that $L_j^{(p)} \lesssim \log^{p-1} n$, as above. Recall the definition of the conditional distribution function $F_{n,x,y}^{(2)}$ from the proof of Proposition 11. By similar but simpler arguments to those used in Proposition 11, we have that

$$\begin{aligned}\mathrm{Var}(\widetilde{T}_n^{(2),\mathrm{p}}) &= \mathrm{Var}(T_n^{(2)}) + o(1/n) = \mathrm{Var}\biggl(\int_{\mathcal{X}} f(x)\phi_x^f\Bigl(\frac{k_Y}{nV_d\rho_{(k_Y),n}(x)^d}\Bigr)\,dx\biggr) + o(1/n)\\ &= \int_{\mathcal{X}} f(x)f(y)\int_{\mathcal{I}_{n,Y}}\int_{\mathcal{I}_{n,Y}} \phi\bigl(f(x),v_{x,t_1}\bigr)\phi\bigl(f(y),v_{y,t_2}\bigr)\\ &\qquad\times\bigl\{dF_{n,x,y}^{(2)}(t_1,t_2) - \mathrm{B}_{k_Y,n+1-k_Y}(t_1)\mathrm{B}_{k_Y,n+1-k_Y}(t_2)\,dt_1\,dt_2\bigr\}\,dx\,dy + o(1/n)\\ &= \frac{v_2}{n} + o(1/n).\end{aligned} \tag{S49}$$

Now, using an analogous statement to that in (S47), using (S49) and the facts that $\mathbb{P}(Y_1 \in \mathcal{X}_{n,g}^{(j)}) \lesssim (k_Y/n)\log n$ and that $v_2 \geq 1/C$, we have for $p=3,4$ that

$$\begin{aligned}&\frac{1}{\mathrm{Var}^{p/2}\bigl(\widetilde{T}_n^{(2)}\bigr)}\sum_{j=1}^{V_n} L_j^{(p)}\mathbb{E}\{|W_j - \mathbb{E}W_j|^p\}\\ &\qquad\lesssim \frac{k_Y^{p-1}\log^{2p-2} n}{n^{p/2-1}}\biggl\{\int_{\mathcal{X}_{n,g}} f(x)^{p+p\kappa_1}g(x)^{-(p-1)+p\kappa_2}\,dx + 1\biggr\}\to 0.\end{aligned}$$

By Theorem 1 of Baldi and Rinott (1989) we now have that

$$d_{\mathrm{K}}\biggl(\mathcal{L}\Bigl(\frac{n^{1/2}\{\widetilde{T}_n^{(2)}-\mathbb{E}\widetilde{T}_n^{(2)}\}}{v_2^{1/2}}\Bigr),N(0,1)\biggr)\to 0.$$

For our weighted estimator $\widehat{T}_{m,n}$, we can define weighted analogues $\widehat{T}_m^{(1)}$ and $\widehat{T}_n^{(2)}$ of $\widetilde{T}_m^{(1)}$ and $\widetilde{T}_n^{(2)}$ and deduce that

$$\widehat{T}_{m,n} - \mathbb{E}(\widehat{T}_{m,n}) = \widehat{T}_{m,n}^{(1)} - \mathbb{E}(\widehat{T}_{m,n}^{(1)}) + \widehat{T}_{m,n}^{(2)} - \mathbb{E}(\widehat{T}_{m,n}^{(2)}) + o_p(m^{-1/2}+n^{-1/2}), \tag{S50}$$

where

$$d_{\mathrm{K}}\biggl(\mathcal{L}\Bigl(\frac{m^{1/2}\{\widehat{T}_m^{(1)}-\mathbb{E}\widehat{T}_m^{(1)}\}}{v_1^{1/2}}\Bigr),N(0,1)\biggr)$$

$$+ d_{\mathrm{K}}\biggl(\mathcal{L}\Bigl(\frac{n^{1/2}\{\widehat{T}_n^{(2)} - \mathbb{E}\widehat{T}_n^{(2)}\}}{v_2^{1/2}}\Bigr), N(0,1)\biggr) = o(1). \tag{S51}$$

If W, X, Y, Z are independent random variables it can be seen by simple conditioning arguments that

$$d_{\mathrm{K}}\bigl(\mathcal{L}(W+X), \mathcal{L}(Y+Z)\bigr) \leq d_{\mathrm{K}}\bigl(\mathcal{L}(W), \mathcal{L}(Y)\bigr) + d_{\mathrm{K}}\bigl(\mathcal{L}(X), \mathcal{L}(Z)\bigr). \tag{S52}$$

Thus, by (S50), (S51), (S52) and Corollary 7, we may write

$$\widehat{Z}_{m,n} := \frac{\widehat{T}_{m,n} - T}{\{v_1/m + v_2/n\}^{1/2}} = Z^*_{m,n} + W_{m,n},$$

where $d_{\mathrm{K}}\bigl(\mathcal{L}(Z^*_{m,n}), N(0,1)\bigr) \to 0$ and $W_{m,n} = o_p(1)$. Thus, for any $\epsilon > 0$,

$$\begin{aligned}
d_{\mathrm{K}}\bigl(\widehat{Z}_{m,n}, N(0,1)\bigr) &\leq \sup_{x\in\mathbb{R}}\bigl|\mathbb{P}\bigl(\widehat{Z}_{m,n} \leq x, |W_{m,n}| \leq \epsilon\bigr) - \Phi(x)\bigr| + \mathbb{P}(|W_{m,n}| > \epsilon) \\
&\leq \sup_{x\in\mathbb{R}} \max\bigl\{\mathbb{P}(Z^*_{m,n} \leq x+\epsilon) - \Phi(x), \Phi(x) - \mathbb{P}(Z^*_{m,n} \leq x-\epsilon)\bigr\} + 2\mathbb{P}(|W_{m,n}| > \epsilon) \\
&\leq d_{\mathrm{K}}\bigl(Z^*_{m,n}, N(0,1)\bigr) + \frac{\epsilon}{(2\pi)^{1/2}} + 2\mathbb{P}(|W_{m,n}| > \epsilon),
\end{aligned}$$

so the result follows. □

PROOF OF THEOREM 5. The main task is to establish the consistency of $\widehat{V}^{(1)}_{m,n}$ and $\widehat{V}^{(2)}_{m,n}$. For the first of these, we start by noting that

$$\begin{aligned}
\mathbb{E}\Bigl[\bigl\{\phi_{X_1} + (f\phi_{10})_{X_1}\bigr\}^4\Bigr] &\leq 16L^4 C^{8L+4(|\kappa_1|+|\kappa_2|)} \int_{\mathcal{X}} f(x)^{1+4\kappa_1} g(x)^{4\kappa_2}\,dx \\
&\leq 16L^4 C^{1+8L+4(|\kappa_1|+|\kappa_2|)}.
\end{aligned} \tag{S53}$$

Using this and Lemmas S3(i), S6 and S7, and writing $\tilde{\phi}(u,v) := \{\phi(u,v) + u\phi_{10}(u,v)\}^2$ and $b_{m,n} := \log m \wedge \log n$, we have that

$$\begin{aligned}
&\Bigl|\mathbb{E}\widehat{V}^{(1),1}_{m,n} - \int_{\mathcal{X}} f(x)\{\phi_x + (f\phi_{10})_x\}^2\,dx\Bigr| \\
&= \biggl|\int_{\mathcal{X}} f(x) \int_0^1\int_0^1 \Bigl[\min\bigl\{\tilde{\phi}(u_{x,s}, v_{x,t}), b_{m,n}\bigr\} - \min\bigl\{\tilde{\phi}_x, b_{m,n}\bigr\}\Bigr] \\
&\qquad\qquad \times \mathrm{B}_{k_X, m-k_X}(s)\mathrm{B}_{k_Y, n+1-k_Y}(t)\,ds\,dt\,dx\biggr| + O\Bigl(\frac{1}{\log m \wedge \log n}\Bigr) \\
&\leq \int_{\mathcal{X}_{m,n}} f(x) \int_{\mathcal{I}_{m,X}}\int_{\mathcal{I}_{n,Y}} \bigl|\tilde{\phi}(u_{x,s}, v_{x,t}) - \tilde{\phi}_x\bigr| \mathrm{B}_{k_X, m-k_X}(s)\mathrm{B}_{k_Y, n+1-k_Y}(t)\,ds\,dt\,dx \\
&\quad + O\biggl(b_{m,n}\max\Bigl\{\Bigl(\frac{k_X\log m}{m}\Bigr)^{\lambda_1}, \Bigl(\frac{k_Y\log n}{n}\Bigr)^{\lambda_2}, \frac{1}{m^4}, \frac{1}{n^4}, \frac{1}{b^2_{m,n}}\Bigr\}\biggr) \\
&\lesssim \int_{\mathcal{X}_{m,n}} f(x)^{1+2\kappa_1} g(x)^{2\kappa_2} \biggl\{\frac{\log^{\frac{1}{2}} m}{k_X^{1/2}} + \frac{\log^{\frac{1}{2}} n}{k_Y^{1/2}} + \Bigl(\frac{k_X M_\beta(x)^d}{m f(x)}\Bigr)^{\frac{2\wedge\beta}{d}} + \Bigl(\frac{k_Y M_\beta(x)^d}{n g(x)}\Bigr)^{\frac{2\wedge\beta}{d}}\biggr\}\,dx \\
&\quad + O\biggl(b_{m,n}\max\Bigl\{\Bigl(\frac{k_X\log m}{m}\Bigr)^{\lambda_1}, \Bigl(\frac{k_Y\log n}{n}\Bigr)^{\lambda_2}, \frac{1}{m^4}, \frac{1}{n^4}, \frac{1}{b^2_{m,n}}\Bigr\}\biggr) = o(1).
\end{aligned}$$

Now, for $i=1,\ldots,m$, write $\xi_i := \min\{\tilde{\phi}(\widehat{f}_{(k_X),i},\widehat{g}_{(k_Y),i}),b_{m,n}\}$, $\xi_i^* := \min\{\tilde{\phi}_{X_i},b_{m,n}\}$ and

$$\widetilde{\mathcal{X}}_{m,n} := \biggl\{x : \frac{f(x)}{M_\beta(x)^d} \geq \frac{k_X^{1/2}}{m^{1/2}}, \frac{g(x)}{M_\beta(x)^d} \geq \frac{k_Y^{1/2}}{n^{1/2}}\biggr\}.$$

We now have that

$$\begin{aligned}
\mathbb{E}\{(\xi_1-\xi_1^*)^2\} &\leq \mathbb{E}\bigl\{\mathbb{1}_{A_1^X\cap A_1^Y}\mathbb{1}_{\{X_1\in\widetilde{\mathcal{X}}_{m,n}\}}(\xi_1-\xi_1^*)^2\bigr\} \\
&\qquad\qquad + O\biggl(b_{m,n}^2\max\biggl\{\frac{1}{m^4},\frac{1}{n^4},\Bigl(\frac{k_X}{m}\Bigr)^{\lambda_1/2},\Bigl(\frac{k_Y}{n}\Bigr)^{\lambda_2/2}\biggr\}\biggr) \\
&\lesssim \int_{\mathcal{X}_{m,n}} f(x)^{1+2\kappa_1}g(x)^{2\kappa_2}\biggl\{\frac{\log m}{k_X}+\frac{\log n}{k_Y}+\Bigl(\frac{k_X M_\beta(x)^d}{mf(x)}\Bigr)^{\frac{2(2\wedge\beta)}{d}}+\Bigl(\frac{k_Y M_\beta(x)^d}{ng(x)}\Bigr)^{\frac{2(2\wedge\beta)}{d}}\biggr\}dx \\
&\qquad + O\biggl(b_{m,n}^2\max\biggl\{\frac{1}{m^4},\frac{1}{n^4},\Bigl(\frac{k_X}{m}\Bigr)^{\lambda_1/2},\Bigl(\frac{k_Y}{n}\Bigr)^{\lambda_2/2}\biggr\}\biggr) \\
&= O\biggl(b_{m,n}^2\max\biggl\{\frac{\log m}{k_X},\frac{\log n}{k_Y},\Bigl(\frac{k_X}{m}\Bigr)^{\frac{2\wedge\beta}{d}},\Bigl(\frac{k_Y}{n}\Bigr)^{\frac{2\wedge\beta}{d}},\Bigl(\frac{k_X}{m}\Bigr)^{\frac{\lambda_1}{2}},\Bigl(\frac{k_Y}{n}\Bigr)^{\frac{\lambda_2}{2}}\biggr\}\biggr).
\end{aligned}$$

It therefore follows by Cauchy–Schwarz that

$$\begin{aligned}
\mathrm{Var}(\widehat{V}_{m,n}^{(1),1}) &= \frac{1}{m}\mathrm{Var}(\xi_1)+2\Bigl(1-\frac{1}{m}\Bigr)\mathrm{Cov}(\xi_1-\xi_1^*,\xi_2^*)+\Bigl(1-\frac{1}{m}\Bigr)\mathrm{Cov}(\xi_1-\xi_1^*,\xi_2-\xi_2^*) \\
&\leq \frac{b_{m,n}^2}{m}+2b_{m,n}[\mathbb{E}\{(\xi_1-\xi_1^*)^2\}]^{1/2}+\mathbb{E}\bigl\{(\xi_1-\xi_1^*)^2\bigr\}=o(1)
\end{aligned}$$

By very similar arguments to those employed in the proof of Proposition S1 we have that $\mathbb{E}(\widehat{V}_{m,n}^{(1),2})-\int_{\mathcal{X}} f(x)\{\phi_x+(f\phi_{10})_x\}\,dx=o(1)$. By Proposition 11 we have that $\mathrm{Var}(\widetilde{T}_{m,n})=o(1)$. Since $\zeta<1/2$, the summands in $\widehat{V}_{m,n}^{(1),2}-\widetilde{T}_{m,n}$ are square integrable and, writing $\xi_i := \widehat{f}_{(k_X),i}\phi_{10}\bigl(\widehat{f}_{(k_X),i},\widehat{g}_{(k_Y),i}\bigr)$ and $\xi_i^* := (f\phi_{10})_{X_i}$, we have by Cauchy–Schwarz again that

$$\mathrm{Var}\biggl(\frac{1}{m}\sum_{i=1}^m \xi_i\biggr) \leq \frac{1}{m}\mathrm{Var}(\xi_1)+2\mathrm{Var}^{1/2}(\xi_2)\mathrm{Var}^{1/2}(\xi_1-\xi_1^*)+\mathrm{Var}(\xi_1-\xi_1^*)=o(1).$$

Combining our bounds on expectations and variances we have now established that, for any $\epsilon>0$,

$$\sup_{\phi\in\Phi(\xi)}\ \sup_{(f,g)\in\widetilde{\mathcal{F}}_{d,\vartheta}}\ \max_{\substack{k_X\in\{k_X^{\mathrm{L}},\ldots,k_X^{\mathrm{U}}\}\\ k_Y\in\{k_Y^{\mathrm{L}},\ldots,k_Y^{\mathrm{U}}\}}} \mathbb{P}(|\widehat{V}_{m,n}^{(1)}-v_1|\geq\epsilon)\to 0. \tag{S54}$$

Now, we have by Cauchy–Schwarz and Lemma S7 that

$$\begin{aligned}
&\sup_{(f,g)\in\widetilde{\mathcal{F}}_{d,\vartheta}}\int_{\mathcal{X}} f(x)\{f(x)^{1+2\kappa_1}g(x)^{-1+2\kappa_2}\}^{3/2}\,dx \\
&\leq \sup_{(f,g)\in\widetilde{\mathcal{F}}_{d,\vartheta}}\biggl\{\int_{\mathcal{X}} g(x)^{1+4(\kappa_2-1)}f(x)^{4(1+\kappa_1)}\,dx\biggr\}^{1/2}\biggl\{\int_{\mathcal{X}} f(x)^{1+2\kappa_1}g(x)^{2\kappa_2}\,dx\biggr\}^{1/2}<\infty.
\end{aligned}$$

Hence, by analogous calculations to those carried out earlier in this proof, we have for any $\epsilon>0$ that

$$\sup_{\phi\in\Phi(\xi)}\ \sup_{(f,g)\in\widetilde{\mathcal{F}}_{d,\vartheta}}\ \max_{\substack{k_X\in\{k_X^{\mathrm{L}},\ldots,k_X^{\mathrm{U}}\}\\ k_Y\in\{k_Y^{\mathrm{L}},\ldots,k_Y^{\mathrm{U}}\}}} \mathbb{P}(|\widehat{V}_{m,n}^{(2)}-v_2|\geq\epsilon)\to 0. \tag{S55}$$

To conclude the proof, given $\epsilon > 0$, we will consider the event $B_\epsilon := \bigl\{\max\bigl(|\widehat{V}^{(1)}_{m,n}/v_1 - 1|, |\widehat{V}^{(2)}_{m,n}/v_2 - 1|\bigr) \le \epsilon\bigr\}$, and define the shorthand

$$\widehat{Z} := \frac{\widehat{T}_{m,n} - T}{\{\widehat{V}^{(1)}_{m,n}/m + \widehat{V}^{(2)}_{m,n}/n\}^{1/2}} \quad \text{and} \quad Z^* := \frac{\widehat{T}_{m,n} - T}{\{v_1/m + v_2/n\}^{1/2}}.$$

For all $\epsilon \in (0, 1/2)$ we have that

$$\begin{aligned}
d_{\mathrm{K}}\bigl(\mathcal{L}(\widehat{Z}), N(0,1)\bigr) &\le \sup_{z\in\mathbb{R}} \bigl|\mathbb{P}(\widehat{Z} \le z) - \mathbb{P}(Z^* \le z)\bigr| + d_{\mathrm{K}}\bigl(\mathcal{L}(Z^*), N(0,1)\bigr) \\
&\le \sup_{z\in\mathbb{R}} \Bigl\{\bigl|\mathbb{P}(Z^* \le (1+\epsilon)z) - \mathbb{P}(Z^* \le z)\bigr| \vee \bigl|\mathbb{P}(Z^* \le z) - \mathbb{P}(Z^* \le (1-\epsilon)z)\bigr|\Bigr\} \\
&\qquad + d_{\mathrm{K}}\bigl(\mathcal{L}(Z^*), N(0,1)\bigr) + 2\mathbb{P}(B_\epsilon^c) \\
&= \sup_{z\in\mathbb{R}} \bigl|\mathbb{P}(Z^* \le (1+\epsilon)z) - \mathbb{P}(Z^* \le (1-\epsilon)z)\bigr| + d_{\mathrm{K}}\bigl(\mathcal{L}(Z^*), N(0,1)\bigr) + 2\mathbb{P}(B_\epsilon^c) \\
&\le 2\epsilon \sup_{z\in\mathbb{R}} \frac{|z|e^{-z^2/8}}{(2\pi)^{1/2}} + 3d_{\mathrm{K}}\bigl(\mathcal{L}(Z^*), N(0,1)\bigr) + 2\mathbb{P}(B_\epsilon^c). \qquad \text{(S56)}
\end{aligned}$$

The first conclusion of Theorem 5 now follows from (S54), (S55) and (S56). The second conclusion is an immediate consequence of the first. □

S1.6. *Proof of Proposition 12.*

PROOF OF PROPOSITION 12. Since f vanishes at infinity, there exists $x_0 > 0$ such that $h(x) \ge 0$ for all $x \le x_0$ and $h(x) \le 0$ for all $x \ge x_0$. Further, as $x \to \infty$, we have by Karamata's theorem (Bingham, Goldie and Teugels, 1989, Proposition 1.5.10) that

$$\begin{aligned}
h(x) &\sim -P'(x) \int_0^x e^{(1-\kappa)P(y)}\, dy = -\frac{P'(x)}{1-\kappa} \int_{e^{(1-\kappa)P(0)}}^{e^{(1-\kappa)P(x)}} \frac{1}{P'\bigl(P^{-1}\bigl(\frac{\log u}{1-\kappa}\bigr)\bigr)}\, du \\
&\sim -\frac{P'(x)}{1-\kappa} \frac{e^{(1-\kappa)P(x)}}{P'(x)} = -\frac{f(x)^{-(1-\kappa)}}{1-\kappa}. \qquad \text{(S57)}
\end{aligned}$$

In particular, since h is continuous, we can now see that $\sup_{x\ge 0} h(x) < \infty$ and $\inf_{x\ge 0} f(x)h(x) > -\infty$. Hence, for $t \ge 0$ sufficiently small, the function $f_t : [0,\infty) \to \mathbb{R}$ defined by

$$f_t(x) := \{1 - th(x)\}f(x),$$

is bounded and takes values in $[0, \infty)$. Moreover, by Fubini's theorem,

$$\begin{aligned}
\int_0^\infty f(x)h(x)\, dx &= \int_0^\infty f'(x) \int_0^x \bigl\{\psi\bigl(f(y)\bigr) - H(f)\bigr\}\, dy\, dx \\
&= -\int_0^\infty f(y)\bigl\{\psi\bigl(f(y)\bigr) - H(f)\bigr\}\, dy = 0,
\end{aligned}$$

so there exists $t_0 > 0$, depending only on κ and f, such that f_t is a density function for $t \in [0, t_0]$.

Observe that the function h defined in (22) solves the differential equation

$$\frac{d}{dx}\biggl(h(x)\frac{f(x)}{f'(x)}\biggr) = \psi\bigl(f(x)\bigr) - H(f) =: g(x). \qquad \text{(S58)}$$

We now derive, for $t \in [0, t_0]$, the density function of the non-negative random variable $f_t(X_1)$ when X_1 has density function f_t on $[0,\infty)$. As $x \to 0$, we have that

$$f_t'(x) = f'(x) - tf''(x)\int_0^x g(y)\,dy - tf'(x)g(x) \sim f'(x) - txg(0)f''(x) - tf'(x)g(0)$$

$$= f'(x)\biggl[1 - tg(0)\frac{x\{P''(x) - P'(x)^2\} + P'(x)}{P'(x)}\biggr]. \tag{S59}$$

We can also see that as $x \to \infty$ we have

$$f''(x)\int_0^x g(y)\,dy + f'(x)g(x) \sim \frac{f''(x)f(x)^{-(1-\kappa)}}{(1-\kappa)P'(x)} + f'(x)f(x)^{-(1-\kappa)}$$

$$= f(x)^\kappa\biggl\{\frac{P'(x)^2 - P''(x)}{(1-\kappa)P'(x)} - P'(x)\biggr\} \sim f(x)^\kappa\frac{\kappa}{1-\kappa}P'(x), \tag{S60}$$

using the fact that $P''(x) \ll P'(x)^2$ as $x \to \infty$ for strictly increasing polynomials P. Finally, we note that $\sup_{x\in[a,b]} f'(x) < 0$ for every $0 < a < b < \infty$. This, together with (S59) and (S60), means that by reducing $t_0 = t_0(\kappa, f) > 0$ if necessary, we may assume that f_t is strictly decreasing on $[0,\infty)$ for $t \in [0,t_0]$. Thus, for $t \in [0,t_0]$, we can define the inverse function f_t^{-1}, and since $f_t(0) = f(0)$, see that when $X_1 \sim f_t$, the density of $f_t(X_1)$ at $z \in (0, f(0))$ is given by

$$\lim_{\delta\to 0}\frac{1}{\delta}\int_0^\infty f_t(x)\bigl(\mathbb{1}_{\{f_t(x)\le z+\delta\}} - \mathbb{1}_{\{f_t(x)\le z\}}\bigr)\,dx = \frac{z}{-f_t'(f_t^{-1}(z))} =: p_t(z).$$

Our goal now is to show that the family $\{p_t : t \in [0,t_0]\}$ is differentiable in quadratic mean at $t = 0$, with score function $g \circ f^{-1}$. For a fixed $z \in (0, f(0))$, let $x = f^{-1}(z)$ and $x_t = f_t^{-1}(z)$. Then we have

$$0 = f_t(x_t) - f(x) = f(x_t) - f(x) - tf(x_t)h(x_t) = (x_t - x)f'(x) - tzh(x) + o(t)$$

as $t \searrow 0$, and hence $\partial x_t/\partial t|_{t=0} = zh(x)/f'(x)$. It now follows from (S58) that

$$\frac{\partial}{\partial t}\Big|_{t=0} p_t(z) = \frac{\partial}{\partial t}\Big|_{t=0}\biggl(\frac{z}{-f_t'(x_t)}\biggr) = z\frac{\frac{\partial}{\partial t}|_{t=0} f_t'(x_t)}{f'(x)^2}$$

$$= -\frac{z}{f'(x)}\biggl\{h(x) + \frac{h'(x)f(x)}{f'(x)} - \frac{h(x)f''(x)f(x)}{f'(x)^2}\biggr\} = p_0(z)g(x)$$

$$= p_0(z)\{\psi(z) - H(f)\}.$$

To prove differentiability in quadratic mean at $t = 0$ with score function $g \circ f^{-1}$, i.e. that

$$\int_0^{f(0)}\biggl[\frac{p_t(z)^{1/2} - p_0(z)^{1/2}}{t} - \frac{1}{2}\{\psi(z) - H(f)\}p_0(z)^{1/2}\biggr]^2\,dz \to 0 \tag{S61}$$

as $t \searrow 0$, it now suffices by the dominated convergence theorem to show that $t^{-2}\{p_t(z)^{1/2} - p_0(z)^{1/2}\}^2$ can be bounded by an integrable function of z for $t \in [0,t_0]$. Define $b_t := (3t/(1-\kappa))^{1/(1-\kappa)}$ and $a_t := f^{-1}(b_t)$. Now, by (S57), (S59) and (S60), it follows that there exists $C' = C'(\kappa, f) > 0$ such that for all $x \le a_t$ and $t \in [0,t_0]$, we have

$$\max\biggl\{\biggl|\frac{f_t(x)}{f(x)} - 1\biggr|, \biggl|\frac{f_t'(x)}{f'(x)} - 1\biggr|\biggr\} \le t\max\biggl\{C', \frac{3f(x)^{-(1-\kappa)}}{2(1-\kappa)}\biggr\} \le \frac{1}{2}.$$

Write $\epsilon_{t,z} := t\max\{C', \frac{3z^{-(1-\kappa)}}{2^\kappa(1-\kappa)}\}$ so that for $z > 2b_t$ we have $\epsilon_{t,z} \le 1/2$ and

$$f_t\biggl(f^{-1}\biggl(\frac{z}{1+\epsilon_{t,z}}\biggr)\biggr) \le \frac{z}{1+\epsilon_{t,z}}\biggl[1 + t\max\biggl\{C', \frac{3(z/(1+\epsilon_{t,z}))^{-(1-\kappa)}}{2(1-\kappa)}\biggr\}\biggr] \le z.$$

We can similarly establish that $f_t\bigl(f^{-1}(z/(1-\epsilon_{t,z}))\bigr) \geq z$. Now, there exists $x_0 \in (0,\infty)$, depending only on f, such that $f''(x) = \{P'(x)^2 - P''(x)\}f(x) \geq 0$ for all $x \geq x_0$. We can therefore see that, by the convexity of P, for $z > 2b_t$ sufficiently small we have

$$\begin{aligned}
|f_t^{-1}(z) - f^{-1}(z)| &\leq f^{-1}\biggl(\frac{z}{1+\epsilon_{t,z}}\biggr) - f^{-1}\biggl(\frac{z}{1-\epsilon_{t,z}}\biggr) \\
&= P^{-1}\biggl(\log\frac{1+\epsilon_{t,z}}{z}\biggr) - P^{-1}\biggl(\log\frac{1-\epsilon_{t,z}}{z}\biggr) \leq \frac{\log\frac{1+\epsilon_{t,z}}{z} - \log\frac{1-\epsilon_{t,z}}{z}}{P'\bigl(P^{-1}(\log\frac{1-\epsilon_{t,z}}{z})\bigr)} \\
&= \frac{\frac{z}{1-\epsilon_{t,z}}\log\frac{1+\epsilon_{t,z}}{1-\epsilon_{t,z}}}{-f'\bigl(f^{-1}(z/(1-\epsilon_{t,z}))\bigr)} \leq \frac{\frac{z}{1-\epsilon_{t,z}}\log\frac{1+\epsilon_{t,z}}{1-\epsilon_{t,z}}}{-f'(f^{-1}(z))} \lesssim_{\kappa,f} \frac{tz^\kappa}{-f'(f^{-1}(z))} \\
&= \frac{tz^{-(1-\kappa)}f^{-1}(z)}{f^{-1}(z)P'(f^{-1}(z))}, \qquad \text{(S62)}
\end{aligned}$$

and $f^{-1}(z)P'\bigl(f^{-1}(z)\bigr) \to \infty$ as $z \searrow 0$. The derivative $P'(x)$ is bounded away from zero for x bounded away from zero, so for z bounded away from $f(0)$ and 0, we can also see that $|f_t^{-1}(z)/f^{-1}(z) - 1| \lesssim_{\kappa,f} t$. As $x \to 0$,

$$\frac{f_t(x)}{f(x)} = 1 + tP'(x)\int_0^x g(y)\,dt = 1 - txP'(x)|g(0)|\{1 + o_{\kappa,f}(1)\},$$

uniformly for $t \in [0,t_0]$. Thus, similarly to in (S62) and by a Taylor expansion, we can see that for z close to $f(0)$ we have that

$$|f_t^{-1}(z) - f^{-1}(z)| \lesssim_{\kappa,f} \frac{tf^{-1}(z)P'(f^{-1}(z))}{-f'(f^{-1}(z))} = \frac{tf^{-1}(z)}{z} \lesssim_{\kappa,f} tf^{-1}(z).$$

Hence, combining this fact with (S62), uniformly over all $z \in \bigl(2b_t, f(0)\bigr)$, we now have that

$$\biggl|\frac{f'(x_t)}{f'(x)} - 1\biggr| = \biggl|\frac{P'(x_t)}{P'(x)}\frac{1}{1-th(x_t)} - 1\biggr| \lesssim_{\kappa,f} \biggl|\frac{x_t}{x} - 1\biggr| + tz^{-(1-\kappa)} \lesssim_{\kappa,f} tz^{-(1-\kappa)}.$$

We deduce that there exists $c = c(\kappa,f) \in (0, \frac{1-\kappa}{3\times 2^{1-\kappa}})$ such that for $t \in [0,t_0]$, when $tz^{-(1-\kappa)} \leq c$ we have $z > 2b_t$ and

$$\biggl|\frac{p_t(z)}{p_0(z)} - 1\biggr| \leq \biggl|\frac{f'(f^{-1}(z))}{f'(f_t^{-1}(z))}\frac{f'(f_t^{-1}(z))}{f_t'(f_t^{-1}(z))} - 1\biggr| \leq \frac{tz^{-(1-\kappa)}}{2c} \leq \frac{1}{2}. \qquad \text{(S63)}$$

Now, after reducing $t_0 = t_0(\kappa,f) > 0$ if necessary, for $t \in [0,t_0]$ and $tf(x)^{-(1-\kappa)} > c$, we have by (S57) that

$$f_t(x) \leq f(x) + \frac{2tf(x)^\kappa}{1-\kappa} \leq tf(x)^\kappa\biggl(\frac{1}{c} + \frac{2}{1-\kappa}\biggr).$$

Thus, when $tz^{-(1-\kappa)} > c$, we have $x_t = f_t^{-1}(z) \leq f^{-1}\bigl((\frac{z}{t(1/c+2/(1-\kappa))})^{1/\kappa}\bigr)$. Moreover, for z bounded away from $f(0)$, we have that $p_0(z) = 1/P'(f^{-1}(z))$ is bounded. Hence, when $t \in [0,t_0]$ and $tz^{-(1-\kappa)} > c$, using (S60) we can see that

$$\begin{aligned}
p_t(z) = \frac{z}{-f_t'(f_t^{-1}(z))} \leq \frac{z}{t\bigl\{f''(x_t)\int_0^{x_t} g(y)\,dy + f'(x_t)g(x_t)\bigr\}} &\leq \frac{2(1-\kappa)z}{\kappa tf(x_t)^\kappa P'(x_t)} \\
&\leq \frac{4 + \frac{2(1-\kappa)}{c}}{\kappa P'(x_t)}, \qquad \text{(S64)}
\end{aligned}$$

so p_t is also bounded uniformly for $t \in [0, t_0]$ and $tz^{-(1-\kappa)} > c$. It now follows from (S63) and (S64) that for $t \in [0, t_0]$,

$$\begin{aligned}\frac{\{p_t(z)^{1/2} - p_0(z)^{1/2}\}^2}{t^2} &\lesssim_{\kappa,f} \mathbb{1}_{\{t \leq cz^{1-\kappa}\}} p_0(z) z^{2\kappa-2} + \mathbb{1}_{\{t > cz^{1-\kappa}\}} t^{-2} \\ &\leq p_0(z) z^{2\kappa-2} + c^{-2} z^{2\kappa-2}.\end{aligned}$$

Since $\kappa > 1/2$, we have

$$\int_0^{f(0)} z^{2\kappa-2} p_0(z)\, dz = \int_0^\infty f(x)^{2\kappa-1}\, dx < \infty,$$

and also the second term is integrable. Finally, then, the differentiability in quadratic mean property (S61) follows from the dominated convergence theorem.

To complete the proof of the first part of Proposition 12, it suffices to study the differentiability properties of the functional H along our path $\{f_t : t \in [0, t_0]\}$. To this end, integrating by parts and using (S57), we may see that

$$\begin{aligned}\frac{d}{dt}\Big|_{t=0} H(f_t) &= \frac{d}{dt}\Big|_{t=0} \int_0^\infty f(x)^\kappa \{1 - th(x)\}^\kappa\, dx = -\kappa \int_0^\infty f(x)^\kappa h(x)\, dx \\ &= -\kappa \int_0^\infty f(x)^{\kappa-1} f'(x) \int_0^x g(y)\, dy\, dx = -\int_0^\infty \frac{d}{dx}\{f(x)^\kappa\} \int_0^x g(y)\, dy\, dx \\ &= \int_0^\infty f(x)^\kappa g(x)\, dx = \int_0^\infty \big\{f(x)^{-(1-\kappa)} - H(f)\big\} g(x) f(x)\, dx \\ &= \int_0^{f(0)} \big\{z^{-(1-\kappa)} - H(f)\big\} g\big(f^{-1}(z)\big) p_0(z)\, dz.\end{aligned}$$

We therefore conclude that the efficient influence function is given by $z \mapsto \big\{z^{-(1-\kappa)} - H(f)\big\}$, and our result now follows from van der Vaart (1998, Theorem 25.21).

We now turn to the second claim of Proposition 12. First, it is clear that $\|f_t\|_\infty = f(0) < \infty$ for all $t \in [0, t_0]$. As shown by (S57), we have that $f_t(x) \lesssim f(x)^\kappa$ uniformly for $x \in [0, \infty)$ and $t \in [0, t_0]$, and it follows that, for any $\alpha > 0$, we have $\sup_{t \in [0,t_0]} \int_0^\infty x^\alpha f_t(x)\, dx < \infty$. For the smoothness condition, in the interests of brevity, we will restrict attention here to $\beta \in (0, 1]$; the arguments extend naturally to any $\beta > 0$. For $\beta \in (0, 1]$ we claim that $\sup_{t \in [0,t_0]} M_{f_t,\beta}(x) \lesssim_{\kappa,f} \max\{1/x, P'(x)\}$, so that we have

$$\sup_{t \in [0,t_0]} \int_0^\infty f_t(x) \left\{\frac{M_{f,\beta}(x)}{f_t(x)}\right\}^\lambda dx \lesssim_{\kappa,f} \int_0^\infty f(x)^{(1-\lambda)\kappa} \max\{x^{-\lambda}, P'(x)^\lambda\}\, dx < \infty$$

for any $\lambda \in (0, 1)$. To establish this claim, we have $\inf_{t \in [0,t_0]} \inf_{x \in [0,1]} f_t(x) > 0$, and so it follows from the smoothness of f and h that for $t \in [0, t_0]$,

$$\begin{aligned}&\sup_{x \in (0,1]} \sup_{y,z \in [0,2x], y \neq z} \frac{|f_t(z) - f_t(y)|}{|z-y|^\beta f_t(x)} \\ &\qquad \lesssim \sup_{x \in (0,1]} \sup_{y,z \in [0,2x], y \neq z} \frac{|f(z) - f(y)| + t_0 |f(z)h(z) - f(y)h(y)|}{|z-y|^\beta} < \infty.\end{aligned}$$

It follows that for $x \in (0, 1]$ we have $M_{f,\beta}(x) \lesssim_{\kappa,f} 1/x$. Writing $\deg(P)$ for the degree of the strictly increasing polynomial P, we have that

$$0 < \inf_{x \in [1,\infty)} \frac{P'(x)}{x^{\deg(P)-1}} \leq \sup_{x \in [1,\infty)} \frac{P'(x)}{x^{\deg(P)-1}} < \infty.$$

Now for $x \geq 1$ and y such that $|y-x| \leq x \wedge \{1/P'(x)\}$ we have that

$$|P(y)-P(x)| \lesssim_f |y-x| \max(x,y)^{\deg(P)-1} \lesssim_f |y-x|P'(x) \leq 1$$

$$|P'(y)-P'(x)| \lesssim_f |y-x| \max(x,y)^{\deg(P)\vee 2-2} \lesssim_f |y-x|P'(x). \tag{S65}$$

It therefore follows that

$$\sup_{x\in[1,\infty)} \sup_{\substack{y,z\in B_x(x\wedge\{1/P'(x)\}) \\ y\neq z}} \frac{|f(z)-f(y)|}{f(x)\{P'(x)|z-y|\}^\beta}$$

$$= \sup_{x\in[1,\infty)} \sup_{\substack{y,z\in B_x(x\wedge\{1/P'(x)\}) \\ y\neq z}} \frac{e^{P(x)-P(y)}\left|e^{P(y)-P(z)}-1\right|}{\{P'(x)|z-y|\}^\beta} < \infty.$$

We conclude from (S57) both that $\sup_{x\in[1,\infty)} f(x)^{1-\kappa}P'(x)|\int_0^x g(y)\,dy| < \infty$ and that $\inf_{x\in[1,\infty)} f_t(x)/\{f(x)+tf(x)^\kappa\} > 0$. Using (S65) we can now see that for $x \in [1,\infty)$ and $y,z \in B_x(x\wedge\{1/P'(x)\})$ we have for $t \in [0,t_0]$ that

$$\frac{|f_t(z)-f_t(y)|}{f_t(x)} \lesssim_{\kappa,f} \frac{|f(z)-f(y)|}{f(x)} + \frac{|f(z)h(z)-f(y)h(y)|}{f(x)^\kappa}$$

$$\lesssim_{\kappa,f} \{P'(x)|z-y|\}^\beta$$

$$+ \frac{|f'(z)\int_y^z g(u)\,du| + |\int_0^y g(u)\,du|\{f(z)|P'(z)-P'(y)| + P'(y)|f(z)-f(y)|\}}{f(x)^\kappa}$$

$$\lesssim_{\kappa,f} \{P'(x)|z-y|\}^\beta + |z-y||f'(x)|/f(x) + |z-y| + |f(z)/f(y)-1|$$

$$\lesssim_{\kappa,f} \{P'(x)|z-y|\}^\beta.$$

This verifies our claim and the result therefore follows. □

S1.7. *Proof of Theorem 14 on the local asymptotic minimax lower bound.*

PROOF OF THEOREM 14. (i) We check the conditions of, and apply, Theorem 3.11.5 of van der Vaart and Wellner (1996), and therefore borrow some of their terminology. Define the Hilbert space $H := \mathbb{R}^2$ with inner product $\langle (t_1,t_2),(t_1',t_2')\rangle_H := t_1t_1'v_1(f,g) + t_2t_2'v_2(f,g)$. We first claim that our sequence of experiments is asymptotically normal. That is to say, for independent normal random variables $Z_1 \sim N(0,v_1)$ and $Z_2 \sim N(0,v_2)$, if we define the iso-Gaussian process $\{\Delta_t = t_1Z_1 + t_2Z_2 : t = (t_1,t_2) \in H\}$ we claim that

$$\log\frac{dP_{n,t}}{dP_{n,0}} = \Delta_{n,t} - \frac{1}{2}\|t\|_H^2$$

with $\Delta_{n,t} \xrightarrow{d} \Delta_t$ for each fixed $t \in H$. Since $\int_{\mathcal{X}} f(x)h_1(x)^2 < \infty$, and since $K(0) = K'(0) = K''(0) = 1$, we have by the dominated convergence theorem that

$$\left|1/c_1(t_1) - 1 - \frac{t_1^2}{2}v_1\right|$$

$$= \left|\int_{\mathcal{X}} f(x)\left\{K(t_1h_1(x)) - 1 - t_1h_1(x) - \frac{t_1^2}{2}h_1(x)^2\right\}dx\right|$$

$$\leq \frac{1}{6}\sup_{w\in[-1,1]}|K'''(w)|\int_{|t_1h_1(x)|\leq 1} f(x)|t_1h_1(x)|^3\,dx$$

$$+ \left\{2\sup_{w\in\mathbb{R}}|K(w)| + 1 + \frac{1}{2}\right\}\int_{|t_1h_1(x)|>1} f(x)\{t_1h_1(x)\}^2\,dx = o(t_1^2)$$

as $t_1 \to 0$, with a similar calculation holding for $1/c_2(t_2)$ since $\int gh_2^2 < \infty$. Therefore, for each fixed $t = (t_1, t_2) \in H$ we have

$$\log \frac{dP_{n,t}}{dP_{n,0}} = \sum_{i=1}^{m} \log \frac{f_{m^{-1/2}t_1}(X_i)}{f(X_i)} + \sum_{j=1}^{n} \log \frac{g_{n^{-1/2}t_2}(Y_j)}{g(Y_j)}$$

$$= \sum_{i=1}^{m} \log K\Big(\frac{t_1 h_1(X_i)}{m^{1/2}}\Big) + m \log c_1(m^{-1/2}t_1) + \sum_{j=1}^{n} \log K\Big(\frac{t_2 h_2(Y_j)}{n^{1/2}}\Big) + n \log c_2(n^{-1/2}t_2)$$

$$= \frac{t_1}{m^{1/2}} \sum_{i=1}^{m} h_1(X_i) + \frac{t_2}{n^{1/2}} \sum_{j=1}^{n} h_2(Y_j) - \frac{1}{2}\|t\|_H^2 + o_p(1) \xrightarrow{d} \Delta_t - \frac{1}{2}\|t\|_H^2,$$

as claimed.

We will now show that the sequence of parameters defined by $\kappa_n(t) := T(f_{m^{-1/2}t_1}, g_{n^{-1/2}t_2})$ is *regular*, in that there exists a continuous linear map $\dot{\kappa} : H \to \mathbb{R}$ and a sequence (r_n) of real numbers such that

$$r_n\{\kappa_n(t) - \kappa_n(0)\} \to \dot{\kappa}(t)$$

for each $t \in H$. Indeed, for any fixed $t = (t_1, t_2) \in H$ we have

$$\kappa_n(t) - \kappa_n(0) = \int_{\mathcal{X}} \Big\{ f_{m^{-1/2}t_1}(x)\phi\big(f_{m^{-1/2}t_1}(x), g_{n^{-1/2}t_2}(x)\big) - f(x)\phi_x \Big\}\, dx$$

$$= \int_{\mathcal{X}} f(x) \Big\{ K\Big(\frac{t_1 h_1(x)}{m^{1/2}}\Big) \phi\Big(K\Big(\frac{t_1 h_1(x)}{m^{1/2}}\Big) f(x), K\Big(\frac{t_2 h_2(x)}{n^{1/2}}\Big) g(x) \Big) - \phi_x \Big\}\, dx$$
$$+ o(m^{-1/2} + n^{-1/2})$$

$$= \int_{\mathcal{X}} f(x) \Big[\frac{t_1 h_1(x)}{m^{1/2}} \{\phi_x + (f\phi_{10})_x\} + \frac{t_2 h_2(x)}{n^{1/2}} (g\phi_{01})_x \Big]\, dx + o(m^{-1/2} + n^{-1/2})$$

$$= \frac{t_1 v_1}{m^{1/2}} + \frac{t_2 v_2}{n^{1/2}} + o(m^{-1/2} + n^{-1/2}).$$

We may therefore take

$$r_n = (v_1/m + v_2/n)^{-1/2} \quad \text{and} \quad \dot{\kappa}(t_1, t_2) = \frac{t_1 v_1 + A^{1/2} t_2 v_2}{(v_1 + A v_2)^{1/2}}$$

to conclude that our sequence of parameters κ_n is regular.

The adjoint $\dot{\kappa}^* : \mathbb{R} \to H$ of $\dot{\kappa}$ is given by

$$\dot{\kappa}^*(b^*) = \Big(\frac{b^*}{(v_1 + A v_2)^{1/2}}, \frac{A^{1/2} b^*}{(v_1 + A v_2)^{1/2}} \Big)$$

as this satisfies $\langle \dot{\kappa}^*(b^*), t \rangle_H = b^* \dot{\kappa}(t)$ for all $b^* \in \mathbb{R}$ and $t \in H$. Since $\|\dot{\kappa}^*(b^*)\|_H^2 = (b^*)^2$ for all $b^* \in \mathbb{R}$, we may therefore take $G \sim N(0,1)$ and apply Theorem 3.11.5 of van der Vaart and Wellner (1996) to deduce that for any estimator sequence $T_{m,n}$,

$$\sup_{I \in \mathcal{I}} \liminf_{n \to \infty} \max_{t \in I} \mathbb{E}_{P_{n,t}} \Big\{ \frac{(T_{m,n} - T)^2}{v_1/m + v_2/n} \Big\} \geq \mathbb{E}(G^2) = 1.$$

This concludes the proof of (i).

(ii) Since $k : \mathbb{R} \to [1/2, 3/2]$ we have that $f(x)/3 \leq f_t(x) \leq 3f(x)$ and $g(x)/3 \leq g_t(x) \leq 3g(x)$ for all $t \in \mathbb{R}$ and $x \in \mathbb{R}^d$ and, to establish the result, it remains to show that

$\max\{M_{f_t,\tilde{\beta}}(x), M_{g_t,\tilde{\beta}}(x)\} \lesssim M_\beta(x)$ for $t \leq 1$, say. For ease of presentation, we first prove this in the case $\beta \in (0,1]$. When $x \in \mathcal{X}$ and $y,z \in B_x\bigl(1/M_\beta(x)\bigr)$, we have that

$$\begin{aligned}\frac{|f_t(z)-f_t(y)|}{f_t(x)} &= \frac{|K(th_1(z))f(z)-K(th_1(y))f(y)|}{K(th_1(x))f(x)} \\ &\leq \frac{3}{f(x)}|f(z)-f(y)| + \frac{2f(y)}{f(x)}|K(th_1(z))-K(th_1(y))| \\ &\leq 3\{M_\beta(x)\|z-y\|\}^\beta + 4|K(th_1(z))-K(th_1(y))|. \end{aligned} \tag{S66}$$

Additionally,

$$\begin{aligned}|h_1(z)-h_1(y)| &\leq |\phi_z - \phi_y| + f(z)|(\phi_{10})_z - (\phi_{10})_y| + |(\phi_{10})_y||f(z)-f(y)| \\ &\leq L\bigl(1 \vee |\phi_y + (f\phi_{10})_y|\bigr)\Bigl(1+\frac{f(z)}{f(y)}\Bigr)\biggl\{3\Bigl|\frac{f(z)}{f(y)}-1\Bigr|^{(\beta^*-1)\wedge 1} + \Bigl|\frac{g(z)}{g(y)}-1\Bigr|^{\beta^*\wedge 1}\biggr\} \\ &\lesssim (1+|h_1(y)|)\{M_\beta(x)\|z-y\|\}^{\tilde{\beta}}.\end{aligned} \tag{S67}$$

In particular, there exists $c = c(d,\vartheta,\xi)$ such that, whenever $\|z-y\|M_\beta(x) \leq c$, we have $|h_1(z)-h_1(y)| \leq \max(1, |h_1(y)| \wedge |h_1(z)|)/2$. Writing $L_{t,y,z}$ for the line segment between $th_1(y)$ and $th_1(z)$, and using the fact that $\sup_{w\in\mathbb{R}}(1+|w|)|K'(w)| < \infty$, we now have for z,y such that $\|z-y\|M_\beta(x) \leq c$ that

$$\begin{aligned}|K(th_1(z))-K(th_1(y))| &\leq t|h_1(z)-h_1(y)| \sup_{w\in L_{t,y,z}} |K'(w)| \\ &\lesssim \frac{(1+t|h_1(y)| \wedge |h_1(z)|)}{1+\inf_{w\in L_{t,y,z}}|w|}\{M_\beta(x)\|z-y\|\}^{\tilde{\beta}} \lesssim \{M_\beta(x)\|z-y\|\}^{\tilde{\beta}}.\end{aligned} \tag{S68}$$

From (S66), (S67) and (S68), we deduce that $M_{f_t,\tilde{\beta}}(x) \lesssim M_\beta(x)$. Moreover, when $y,z \in B_x\bigl(1/M_\beta(x)\bigr)$, we have that

$$\begin{aligned}&|h_2(z)-h_2(y)| \leq f(z)|(\phi_{01})_z - (\phi_{01})_y| + |(\phi_{01})_y||f(z)-f(y)| \\ &\leq \frac{f(z)}{f(y)}\bigl(1 \vee f(y)|(\phi_{01})_y|\bigr)\biggl\{\Bigl|\frac{f(z)}{f(y)}-1\Bigr|^{\beta^*\wedge 1} + \Bigl|\frac{g(z)}{g(y)}-1\Bigr|^{(\beta^*-1)\wedge 1}\biggr\} + |(\phi_{01})_y||f(z)-f(y)| \\ &\lesssim (1+|h_2(y)|)\{M_\beta(x)\|z-y\|\}^{\tilde{\beta}}.\end{aligned}$$

It now follows by very similar arguments to those in (S68) that $M_{g_t,\tilde{\beta}}(x) \lesssim M_\beta(x)$.

We now extend these arguments to cover the $\beta > 1$ case. For a multi-index $\boldsymbol{\alpha} \in \mathbb{N}_0^d$ with $|\boldsymbol{\alpha}| \leq \underline{\tilde{\beta}} := \lceil \tilde{\beta} \rceil - 1$, we have that $\partial^{\boldsymbol{\alpha}}\{K(th_1(x))\}$ can be written as a finite sum of terms of the form

$$t^r(\partial^{\boldsymbol{\alpha}^{(1)}}h_1)\ldots(\partial^{\boldsymbol{\alpha}^{(r)}}h_1)(x)K^{(r)}\bigl(th_1(x)\bigr) \tag{S69}$$

where $r \in \mathbb{N}_0$ satisfies $r \leq |\boldsymbol{\alpha}|$, and the multi-indices $\boldsymbol{\alpha}^{(1)},\ldots,\boldsymbol{\alpha}^{(r)} \in \mathbb{N}_0^d$ satisfy $|\boldsymbol{\alpha}^{(1)}| + \ldots + |\boldsymbol{\alpha}^{(r)}| = |\boldsymbol{\alpha}|$. Moreover, for any $j = 1,\ldots,r$, we have that $\partial^{\boldsymbol{\alpha}^{(j)}}h_1$ is a finite sum of terms of the form

$$(\partial^{\boldsymbol{\beta}^{(1)}}f)\ldots(\partial^{\boldsymbol{\beta}^{(\ell_1)}}f)(\partial^{\boldsymbol{\gamma}^{(1)}}g)\ldots(\partial^{\boldsymbol{\gamma}^{(\ell_2)}}g)(x)\phi_{\ell_1\ell_2}(f(x),g(x)), \tag{S70}$$

where $\ell_1,\ell_2 \in \mathbb{N}_0$ satisfy $\ell_1 + \ell_2 \leq |\boldsymbol{\alpha}^{(j)}| + 1$, and where moreover the multi-indices $\boldsymbol{\beta}^{(1)},\ldots,\boldsymbol{\beta}^{(\ell_1)},\boldsymbol{\gamma}^{(1)},\ldots,\boldsymbol{\gamma}^{(\ell_2)} \in \mathbb{N}_0^d$ satisfy $|\boldsymbol{\beta}^{(1)}| + \ldots + |\boldsymbol{\beta}^{(\ell_1)}| + |\boldsymbol{\gamma}^{(1)}| + \ldots + |\boldsymbol{\gamma}^{(\ell_2)}| =$

$|\boldsymbol{\alpha}^{(j)}|$. Using the fact that $\sup_{w\in\mathbb{R}}(1+|w|^r)|K^{(r)}(w)| < \infty$ for any $r \in \mathbb{N}$ and assumption (i) in the definition of $\tilde{\Phi}$, we therefore have the bounds

$$|\partial^{\boldsymbol{\alpha}^{(j)}} h_1(x)| \lesssim M_\beta(x)^{|\boldsymbol{\alpha}^{(j)}|}(1+|h_1(x)|) \quad \text{and} \quad |\partial^{\boldsymbol{\alpha}}\{K(th_1(x))\}| \lesssim M_\beta(x)^{|\boldsymbol{\alpha}|}.$$

It follows that, for any multi-index $\boldsymbol{\alpha}$ with $|\boldsymbol{\alpha}| \leq \underline{\tilde{\beta}}$ we have that

$$\Bigl|\frac{\partial^{\boldsymbol{\alpha}} f_t(x)}{f_t(x)}\Bigr| \lesssim M_\beta(x)^{|\boldsymbol{\alpha}|}.$$

Since, for any multi-index $\boldsymbol{\alpha}$ with $|\boldsymbol{\alpha}| \leq \underline{\tilde{\beta}}$, we have that $\partial^{\boldsymbol{\alpha}} h_2$ is a finite sum of terms of the form

$$(\partial^{\boldsymbol{\beta}^{(1)}} f)\ldots(\partial^{\boldsymbol{\beta}^{(\ell_1+1)}} f)(\partial^{\boldsymbol{\gamma}^{(1)}} g)\ldots(\partial^{\boldsymbol{\gamma}^{(\ell_2-1)}} g)(x)\phi_{\ell_1\ell_2}(f(x),g(x)),$$

we deduce by similar arguments that $|\partial^{\boldsymbol{\alpha}} g_t(x)| \lesssim g_t(x) M_\beta(x)^{|\boldsymbol{\alpha}|}$ for any multi-index $\boldsymbol{\alpha}$ with $|\boldsymbol{\alpha}| \leq \underline{\tilde{\beta}}$. Now we have for any $\ell_1,\ell_2 \in \mathbb{N}_0$ with $\ell_1+\ell_2 \leq \beta^*-1$ and $y,z \in B_x\bigl(1/M_\beta(x)\bigr)$ that

$$\begin{aligned}
&|\phi_{\ell_1\ell_2}(f(z),g(z)) - \phi_{\ell_1\ell_2}(f(y),g(y))| \\
&\lesssim f(y)^{-\ell_1} g(y)^{-\ell_2}(1 \vee |h_1(y)|)\biggl\{\Bigl|\frac{f(z)}{f(y)}-1\Bigr|^{(\beta^*-\ell_1)\wedge 1} + \Bigl|\frac{g(z)}{g(y)}-1\Bigr|^{(\beta^*-\ell_2)\wedge 1}\biggr\} \\
&\lesssim f(y)^{-\ell_1} g(y)^{-\ell_2}(1 \vee |h_1(y)|)\{M_\beta(x)\|z-y\|\}^{\min\{1,\beta^*-\ell_1,\beta^*-\ell_2\}}.
\end{aligned}$$

It follows from this, together with the representation (S70) and Lemma S5 that, for any multi-index $\boldsymbol{\alpha}$ with $|\boldsymbol{\alpha}| \leq \underline{\tilde{\beta}}$, we have that

$$\begin{aligned}
|\partial^{\boldsymbol{\alpha}} h_1(z) - \partial^{\boldsymbol{\alpha}} h_1(y)| &\lesssim M_\beta(x)^{|\boldsymbol{\alpha}|}(1 \vee |h_1(y)|)\{M_\beta(x)\|z-y\|\}^{\min\{1,\beta-\underline{\tilde{\beta}},\beta^*-1-\underline{\tilde{\beta}}\}} \\
&\lesssim M_\beta(x)^{|\boldsymbol{\alpha}|}(1 \vee |h_1(y)|)\{M_\beta(x)\|z-y\|\}^{\tilde{\beta}-\underline{\tilde{\beta}}}.
\end{aligned}$$

By a similar argument to (S68), and using (S69) and the fact that $\sup_{w\in\mathbb{R}}(1+|w|^r)|K^{(r)}(w)| < \infty$ for any $r \in \mathbb{N}$, we can now see that, for any multi-index $\boldsymbol{\alpha}$ with $|\boldsymbol{\alpha}| \leq \underline{\tilde{\beta}}$,

$$\bigl|\partial^{\boldsymbol{\alpha}}\{K(th_1(z))\} - \partial^{\boldsymbol{\alpha}}\{K(th_1(y))\}\bigr| \lesssim M_\beta(x)^{|\boldsymbol{\alpha}|}\{M_\beta(x)\|z-y\|\}^{\tilde{\beta}-\underline{\tilde{\beta}}}.$$

Using Lemma S5 it then follows that, for any multi-index $\boldsymbol{\alpha}$ with $|\boldsymbol{\alpha}| = \underline{\tilde{\beta}}$, we have

$$|\partial^{\boldsymbol{\alpha}} f_t(z) - \partial^{\boldsymbol{\alpha}} f_t(y)| \lesssim f_t(x) M_\beta(x)^{\underline{\tilde{\beta}}}\{M_\beta(x)\|z-y\|\}^{\tilde{\beta}-\underline{\tilde{\beta}}}$$

and so $M_{f_t,\tilde{\beta}}(x) \lesssim M_\beta(x)$, as required. Similarly, $M_{g_t,\tilde{\beta}}(x) \lesssim M_\beta(x)$, and this completes the proof of the first statement in Theorem 14(ii).

It remains to prove the local asymptotic minimax result for $\widehat{T}_{m,n}$ under the conditions of Theorem 2, together with $\tilde{\beta} = \beta$. Observe that

$$\begin{aligned}
&\sup_{I\in\mathcal{I}} \limsup_{n\to\infty} \max_{t=(t_1,t_2)\in I} n\mathbb{E}_{P_{n,t}}\Bigl[\bigl\{\widehat{T}_{m,n} - T(f_{m^{-1/2}t_1}, g_{n^{-1/2}t_2})\bigr\}^2\Bigr] \\
&\leq \limsup_{n\to\infty} \sup_{(\tilde{f},\tilde{g})\in\mathcal{F}_{d,\tilde{\vartheta}}} \Bigl\{n\mathbb{E}_{\tilde{f},\tilde{g}}\bigl[\{\widehat{T}_{m,n} - T(\tilde{f},\tilde{g})\}^2\bigr] - \frac{n}{m}v_1(\tilde{f},\tilde{g}) - v_2(\tilde{f},\tilde{g})\Bigr\} \\
&\qquad + \sup_{I\in\mathcal{I}} \limsup_{n\to\infty} \max_{t=(t_1,t_2)\in I} \Bigl\{\frac{n}{m}v_1(f_{m^{-1/2}t_1}, g_{n^{-1/2}t_2}) + v_2(f_{m^{-1/2}t_1}, g_{n^{-1/2}t_2})\Bigr\} \\
&\leq \frac{1}{A}v_1(f,g) + v_2(f,g),
\end{aligned}$$

where, in the second inequality, we have applied Theorem 2 to the first term, and used the continuity properties of v_1 and v_2 for the second term. The fact that the inequalities in this display are attained follows from Theorem 14(i), and this completes the proof. □

S1.8. *Auxiliary lemmas.*

LEMMA S3. *Suppose that $\phi \in \Phi(\xi)$ for some $\xi = (\kappa_1, \kappa_2, \beta^*, L) \in \Xi$. Then*

(i) *For all $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2) \in (-1/2, 1/2)^2$ and $\mathbf{z} = (u, v) \in (0, \infty)^2$ we have*

$$\max\bigl\{u|\phi_{10}(\mathbf{z}+\boldsymbol{\epsilon}\circ\mathbf{z}) - \phi_{10}(\mathbf{z})|, v|\phi_{01}(\mathbf{z}+\boldsymbol{\epsilon}\circ\mathbf{z}) - \phi_{01}(\mathbf{z})|\bigr\} \leq 2^{1+|\kappa_1|+|\kappa_2|+2L} L\|\boldsymbol{\epsilon}\| u_\wedge^{\kappa_1} u_\vee^L v_\wedge^{\kappa_2} v_\vee^L.$$

(ii) *For all $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2) \in (-1/2, 1/2)^2$ and $\mathbf{z} = (u, v) \in (0, \infty)^2$ we have*

$$\biggl|\phi(\mathbf{z}+\boldsymbol{\epsilon}\circ\mathbf{z}) - \sum_{\ell_1,\ell_2=0}^{\infty} \mathbb{1}_{\{\ell_1+\ell_2 \leq \beta^*-1\}} \frac{(u\epsilon_1)^{\ell_1}(v\epsilon_2)^{\ell_2}}{\ell_1!\ell_2!} \phi_{\ell_1\ell_2}(\boldsymbol{z})\biggr| \leq 2^{1+|\kappa_1|+|\kappa_2|+2L} L u_\wedge^{\kappa_1} u_\vee^L v_\wedge^{\kappa_2} v_\vee^L (|\epsilon_1| \vee |\epsilon_2|)^{\beta^*}.$$

PROOF OF LEMMA S3. By the definition of the class Φ, for each $\mathbf{z} \in \mathcal{Z}$, the Hessian matrix

$$H(\mathbf{z}) := \begin{pmatrix} u^2\phi_{20}(\mathbf{z}) & uv\phi_{11}(\mathbf{z}) \\ uv\phi_{11}(\mathbf{z}) & v^2\phi_{02}(\mathbf{z}) \end{pmatrix}$$

satisfies $\|H(\mathbf{z})\|_{\mathrm{op}} \leq 2Lu_\wedge^{\kappa_1} u_\vee^L v_\wedge^{\kappa_2} v_\vee^L$. Now, fixing $\mathbf{z} \in \mathcal{Z}$, the function $g : [0,1] \to \mathbb{R}$ given by $g(t) := u\phi_{10}(\mathbf{z}+t\boldsymbol{\epsilon}\circ\mathbf{z})$ is differentiable with $g'(t) = \{H(\mathbf{z}+t\boldsymbol{\epsilon}\circ\mathbf{z})(\epsilon_1,\epsilon_2)^T\}_1$. Thus, by the mean value theorem,

$$\begin{aligned} &u|\phi_{10}(\mathbf{z}+\boldsymbol{\epsilon}\circ\mathbf{z}) - \phi_{10}(\mathbf{z})| = |g(1) - g(0)| \\ &\leq 2L(1+1/2)^{2L} \max\{(1/2)^{-\kappa_1^- - \kappa_2^-}, (1+1/2)^{\kappa_1^+ + \kappa_2^+}\} \|\boldsymbol{\epsilon}\| u_\wedge^{\kappa_1} u_\vee^L v_\wedge^{\kappa_2} v_\vee^L. \end{aligned}$$

A similar calculation with ϕ_{01} completes the proof of part (i).

To prove part (ii) we use the mean value form of the remainder in Taylor's theorem. Fixing $\boldsymbol{z} \in \mathcal{Z}$ and $\boldsymbol{\epsilon} \in (-1/2, 1/2)^2$ define $h : [0,1] \to \mathbb{R}$ by $h(t) = \phi(\mathbf{z}+t\boldsymbol{\epsilon}\circ\mathbf{z})$. Then we have

$$\begin{aligned} &\biggl|\phi(\mathbf{z}+\boldsymbol{\epsilon}\circ\mathbf{z}) - \sum_{\ell_1,\ell_2=0}^{\infty} \mathbb{1}_{\{\ell_1+\ell_2 \leq \beta^*-1\}} \frac{(u\epsilon_1)^{\ell_1}(v\epsilon_2)^{\ell_2}}{\ell_1!\ell_2!} \phi_{\ell_1\ell_2}(\boldsymbol{z})\biggr| = \biggl|h(1) - \sum_{b=0}^{\beta^*-1} \frac{1}{b!} h^{(b)}(0)\biggr| \\ &\leq \sup_{t\in[0,1]} \frac{1}{\beta^*!} |h^{(\beta^*)}(t)| = \sup_{t\in[0,1]} \biggl|\sum_{\ell=0}^{\beta^*} \frac{(u\epsilon_1)^{\ell}(v\epsilon_2)^{\beta^*-\ell}}{\ell!(\beta^*-\ell)!} \phi_{\ell,\beta^*-\ell}(\mathbf{z}+t\boldsymbol{\epsilon}\circ\mathbf{z})\biggr| \\ &\leq L(|\epsilon_1| \vee |\epsilon_2|)^{\beta^*} 2^{|\kappa_1|+|\kappa_2|+2L} u_\wedge^{\kappa_1} u_\vee^L v_\wedge^{\kappa_2} v_\vee^L \sum_{\ell=0}^{\beta^*} \frac{1}{\ell!(\beta^*-\ell)!} \\ &\leq 2^{1+|\kappa_1|+|\kappa_2|+2L} L u_\wedge^{\kappa_1} u_\vee^L v_\wedge^{\kappa_2} v_\vee^L (|\epsilon_1| \vee |\epsilon_2|)^{\beta^*}, \end{aligned}$$

as claimed. □

LEMMA S4. *Fix $f \in \mathcal{F}_d$ and $\beta \in (0, \infty)$, and let $\mathcal{S}_n \subseteq (0,1), \mathcal{X}_n \subseteq \mathbb{R}^d$ be such that*

$$a_n := \sup_{s\in\mathcal{S}_n} \sup_{x\in\mathcal{X}_n} \frac{sM_{f,\beta}(x)^d}{V_d f(x)} \to 0.$$

Then there exist $n_* = n_*(d,\beta,(a_n)) \in \mathbb{N}$*, coefficients* $b_\ell(x)$ *and* $A = A(d,\beta,(a_n)) \in (0,\infty)$ *such that, for all* $n \geq n_*, s \in \mathcal{S}_n$ *and* $x \in \mathcal{X}_n$*, we have*

$$\biggl|V_d f(x) h_{x,f}^{-1}(s)^d - \sum_{\ell=0}^{\lceil\beta/2\rceil-1} b_\ell(x) s^{1+2\ell/d}\biggr| \leq As\Bigl\{\frac{sM_{f,\beta}(x)^d}{f(x)}\Bigr\}^{\beta/d}.$$

Moreover, $b_0(x) = 1$ *and* $|b_\ell(x)| \leq A\{M_{f,\beta}(x)^d/f(x)\}^{2\ell/d}$.

PROOF OF LEMMA S4. By a Taylor expansion, for $r \leq 1/M_{f,\beta}(x)$ we have that

$$\biggl|h_{x,f}(r) - V_d r^d f(x) - \sum_{\ell=1}^{\lceil\beta/2\rceil-1} r^{d+2\ell} c_\ell(x)\biggr| \lesssim_{\beta,d} r^d f(x)\{M_{f,\beta}(x) r\}^\beta \tag{S71}$$

for some coefficients $c_\ell(\cdot)$ satisfying $|c_\ell(x)| \lesssim_{\beta,d} f(x) M_{f,\beta}(x)^{2\ell}$. In particular,

$$\biggl|\frac{h_{x,f}(r)}{V_d r^d f(x)} - 1\biggr| \lesssim_{\beta,d} \{M_{f,\beta}(x) r\}^{2\wedge\beta}.$$

Thus there exists $C = C(d,\beta) > 0$ such that we have $|\frac{h_{x,f}(r)}{V_d r^d f(x)} - 1| \leq 1/2$ whenever $r \leq 1/\{CM_{f,\beta}(x)\}$. Setting $r = \{\frac{2s}{V_d f(x)}\}^{1/d}$ we have

$$rCM_{f,\beta}(x) = 2^{1/d} C\Bigl\{\frac{sM_{f,\beta}(x)^d}{V_d f(x)}\Bigr\}^{1/d} \leq (2a_n)^{1/d} C \to 0.$$

So, for n large enough that $(2a_n)^{1/d} C \leq 1$, we have $h_{x,f}(\{\frac{2s}{V_d f(x)}\}^{1/d}) \geq s$, so $h_{x,f}^{-1}(s) \leq \{\frac{2s}{V_d f(x)}\}^{1/d}$ for all $x \in \mathcal{X}_n$ and $s \in \mathcal{S}_n$. Now, since $M_{f,\beta}(x) h_{x,f}^{-1}(s) \leq \{\frac{2sM_{f,\beta}(x)^d}{V_d f(x)}\}^{1/d} \leq (2a_n)^{1/d} \to 0$, we may substitute $r = h_{x,f}^{-1}(s)$ into (S71) to see that

$$\biggl|\frac{s}{V_d f(x) h_{x,f}^{-1}(s)^d} - 1 - \sum_{\ell=1}^{\lceil\beta/2\rceil-1} \frac{b_\ell(x)}{V_d f(x)} h_{x,f}^{-1}(s)^{2\ell}\biggr| \lesssim_{\beta,d,(a_n)} \Bigl\{\frac{sM_{f,\beta}(x)^d}{f(x)}\Bigr\}^{\beta/d}.$$

This expansion can be inverted to yield the desired result by substituting this bound into itself and expanding functions of the form $r \mapsto r^{2\ell/d}$ about $r = 1$. □

LEMMA S5. *Fix* $f \in \mathcal{F}_d$ *and* $\beta \in (0,\infty)$*, and suppose that* $\max\{\|y-x\|, \|z-x\|\} \leq 1/\{(6d)^{1/(\beta-\underline{\beta})} M_{f,\beta}(x)\}$*. Then, for multi-indices* $t \in \mathbb{N}_0^d$ *with* $|t| \leq \underline{\beta}$*, we have that*

$$\bigl|(\partial^t f)(z) - (\partial^t f)(y)\bigr| \leq 2M_{f,\beta}(x)^{\min(\beta,|t|+1)} f(x) \|z-y\|^{\min(1,\beta-|t|)}.$$

PROOF. First, if $|t| = \underline{\beta}$ then we simply have that

$$\bigl|(\partial^t f)(z) - (\partial^t f)(y)\bigr| \leq \|f^{(\underline{\beta})}(z) - f^{(\underline{\beta})}(y)\| \leq M_{f,\beta}(x)^\beta f(x) \|z-y\|^{\beta-\underline{\beta}},$$

and the claim holds. Henceforth assume that $|t| \leq \underline{\beta} - 1$ and $\underline{\beta} \geq 1$. Writing $\|\!|\cdot|\!\|$ here for the largest absolute entry of an array, writing L_{yz} for the line segment between y and z, and arguing inductively we have that

$$\bigl|\partial^t f(z) - \partial^t f(y)\bigr| \leq \|z-y\| \sup_{w \in L_{yz}} \bigl\|\nabla \partial^t f(w)\bigr\|$$

$$\leq \|z-y\| \|f^{(|t|+1)}(x)\| + d^{1/2} \|z-y\| \Bigl\{\Bigl|\!\Bigl|\!\Bigl| f^{(|t|+1)}(y) - f^{(|t|+1)}(x)\Bigr|\!\Bigr|\!\Bigr|$$

$$+ \sup_{w \in L_{yz}} \left\| \left\| f^{(|t|+1)}(w) - f^{(|t|+1)}(y) \right\| \right\| \Big\}$$

$$\leq \|z-y\| f(x) \Big[M_{f,\beta}(x)^{|t|+1}$$
$$+ 2d^{1/2} M_{f,\beta}(x)^{\min(\beta,|t|+2)} \big\{ \|y-x\|^{\min(1,\beta-|t|-1)} + \|z-y\|^{\min(1,\beta-|t|-1)} \big\} \Big]$$
$$\leq \|z-y\| f(x) \Big\{ M_{f,\beta}(x)^{|t|+1} + M_{f,\beta}(x)^{\min(\beta,|t|+2)-\min(1,\beta-|t|-1)} \Big\}$$
$$= 2M_{f,\beta}(x)^{|t|+1} f(x) \|z-y\|,$$

as required. □

The following lemma presents a tail bound for a $\mathrm{Beta}(a, b-a)$ random variable that is convenient to apply in settings where $a > 0$ is large and a/b is small.

LEMMA S6. *Suppose $b > a > 0$ and $B \sim \mathrm{Beta}(a, b-a)$. Writing $h(t) := t - \log(1+t)$ we have that*

$$\mathbb{P}\Big(\Big| B - \frac{a}{b} \Big| \geq \frac{a^{1/2}u}{b} \Big) \leq 2\exp\Big(-ah\Big(\frac{a^{-1/2}b^{1/2}u}{b^{1/2}+a^{1/2}+u} \Big) \Big) + 2\exp\Big(-bh\Big(\frac{u}{b^{1/2}+a^{1/2}+u} \Big) \Big)$$

for all $u \in [0, \infty)$.

PROOF. Our proof relies on concentration inequalities for gamma random variables, which we establish now. For $a > 0$, letting $\Gamma_a \sim \Gamma(a, 1)$ we have by a Chernoff bound that for $t \geq 0$,

$$\mathbb{P}\Big(\frac{\Gamma_a - a}{a} \geq t \Big) \leq \inf_{\lambda \in (0,a)} e^{-\lambda t - \lambda} \Big(1 - \frac{\lambda}{a} \Big)^{-a} = e^{-ah(t)}.$$

Similarly, for $t \in [0, 1)$ we have that

$$\mathbb{P}\Big(\frac{\Gamma_a - a}{a} \leq -t \Big) \leq \inf_{\lambda > 0} e^{\lambda - \lambda t} \Big(1 + \frac{\lambda}{a} \Big)^{-a} = e^{-ah(-t)} \leq e^{-ah(t)},$$

and thus, for all $t \geq 0$, we have that $\mathbb{P}(|\Gamma_a - a| \geq at) \leq 2e^{-ah(t)}$. Now, for independent random variables $\Gamma_a \sim \Gamma(a, 1)$ and $\Gamma_{b-a} \sim \Gamma(b-a, 1)$ we have that $\Gamma_a/(\Gamma_a + \Gamma_{b-a}) \sim \mathrm{Beta}(a, b)$, and so for $t \geq 0$ and $\epsilon \in (0, 1)$ we have that

$$\mathbb{P}\Big(\Big| B - \frac{a}{b} \Big| \geq t \Big) = \mathbb{P}\Big(\Big| \frac{\Gamma_a - a}{\Gamma_a + \Gamma_{b-a}} + \frac{a}{b}\Big(\frac{b}{\Gamma_a + \Gamma_{b-a}} - 1 \Big) \Big| \geq t \Big)$$
$$\leq \mathbb{P}\Big(\Big| \frac{a}{b}\Big(\frac{b}{\Gamma_a + \Gamma_{b-a}} - 1 \Big) \Big| \geq \epsilon t \Big) + \mathbb{P}\Big(\frac{|\Gamma_a - a|}{b} \geq \frac{(1-\epsilon)t}{1 + \epsilon tb/a} \Big)$$
$$\leq \mathbb{P}\Big(\frac{|\Gamma_a + \Gamma_{b-a} - b|}{b} \geq \frac{\epsilon tb}{a + \epsilon tb} \Big) + \mathbb{P}\Big(\frac{|\Gamma_a - a|}{a} \geq \frac{(1-\epsilon)tb}{a + \epsilon tb} \Big).$$

Choosing $\epsilon = a^{1/2}/(a^{1/2} + b^{1/2})$ and writing $t = a^{1/2}u/b$ we may now see that

$$\mathbb{P}\Big(\Big| B - \frac{a}{b} \Big| \geq \frac{a^{1/2}u}{b} \Big)$$
$$\leq \mathbb{P}\Big(\frac{|\Gamma_a + \Gamma_{b-a} - b|}{b} \geq \frac{u}{a^{1/2}+b^{1/2}+u} \Big) + \mathbb{P}\Big(\frac{|\Gamma_a - a|}{a} \geq \frac{a^{-1/2}b^{1/2}u}{a^{1/2}+b^{1/2}+u} \Big)$$
$$\leq 2\exp\Big(-bh\Big(\frac{u}{b^{1/2}+a^{1/2}+u} \Big) \Big) + 2\exp\Big(-ah\Big(\frac{a^{-1/2}b^{1/2}u}{b^{1/2}+a^{1/2}+u} \Big) \Big),$$

as required. □

LEMMA S7. *Fix $d \in \mathbb{N}$ and $\vartheta = (\alpha, \beta, \lambda_1, \lambda_2, C) \in \Theta$. Suppose that $a, b, c \in [0, \infty)$ are such that $\frac{a}{\lambda_1} + \frac{b}{\lambda_2} + \frac{c}{\alpha} \leq 1$. Then*

$$\sup_{(f,g)\in\mathcal{F}_{d,\vartheta}} \int_{\mathcal{X}} f(x)\Big\{\frac{M_\beta(x)^d}{f(x)}\Big\}^a \Big\{\frac{M_\beta(x)^d}{g(x)}\Big\}^b (1+\|x\|)^c \, dx < \infty.$$

PROOF. By the generalised Hölder inequality (e.g. Folland, 1999, Chapter 6, Exercise 31), if $X \sim f$ we have that

$$\begin{aligned}
&\int_{\mathcal{X}} f(x)\Big\{\frac{M_\beta(x)^d}{f(x)}\Big\}^a \Big\{\frac{M_\beta(x)^d}{g(x)}\Big\}^b (1+\|x\|)^c \, dx \\
&\quad = \mathbb{E}\bigg[\Big\{\frac{M_\beta(X)^d}{f(X)}\Big\}^a \Big\{\frac{M_\beta(X)^d}{g(X)}\Big\}^b (1+\|X\|)^c\bigg] \\
&\quad \leq \mathbb{E}\bigg[\Big\{\frac{M_\beta(X)^d}{f(X)}\Big\}^{\lambda_1}\bigg]^{\frac{a}{\lambda_1}} \mathbb{E}\bigg[\Big\{\frac{M_\beta(X)^d}{g(X)}\Big\}^{\lambda_2}\bigg]^{\frac{b}{\lambda_2}} \mathbb{E}\Big[(1+\|X\|)^{\frac{c}{1-\frac{a}{\lambda_1}-\frac{b}{\lambda_2}}}\Big]^{1-\frac{a}{\lambda_1}-\frac{b}{\lambda_2}} \\
&\quad \leq C^{\frac{a}{\lambda_1}+\frac{b}{\lambda_2}} \big[\mathbb{E}\{(1+\|X\|)^\alpha\}\big]^{1-\frac{a}{\lambda_1}-\frac{b}{\lambda_2}} \leq C^{\frac{a}{\lambda_1}+\frac{b}{\lambda_2}} \{2^\alpha(1+C)\}^{1-\frac{a}{\lambda_1}-\frac{b}{\lambda_2}},
\end{aligned}$$

as required. □

LEMMA S8. *Fix $f \in \mathcal{F}_d$ with $\max\big(\|f\|_\infty, \mu_\alpha(f)\big) \leq C$ and $\beta \in (0, \infty)$. Then for all $x \in \mathcal{X}$ and $s \in (0,1)$,*

$$\begin{aligned}
\Big(\frac{s}{CV_d}\Big)^{1/d} \leq h_{x,f}^{-1}(s) \leq \min\bigg\{&\|x\| + \Big(\frac{C}{1-s}\Big)^{1/\alpha}, \\
&\Big(\frac{2s}{V_d f(x)}\Big)^{1/d}\bigg[1 + (6d)^{1/(\beta-\underline{\beta})} M_{f,\beta}(x)\Big\{\|x\| + \Big(\frac{C}{1-s}\Big)^{1/\alpha}\Big\}\bigg]\bigg\}.
\end{aligned}$$

PROOF. The lower bound is immediate on noting that

$$h_{x,f}(r) \leq CV_d r^d.$$

For the upper bound, by Lemma S5, if $\|y - x\| \leq 1/\{(6d)^{1/(\beta-\underline{\beta})} M_{f,\beta}(x)\}$, then we have that

$$\Big|\frac{f(y)}{f(x)} - 1\Big| \leq 2M_{f,\beta}(x)^{1\wedge\beta} \|y-x\|^{1\wedge\beta} \leq \frac{1}{2}.$$

Thus, whenever $r \leq 1/\{(6d)^{1/(\beta-\underline{\beta})} M_{f,\beta}(x)\}$ we have that

$$\frac{1}{2} V_d r^d f(x) \leq h_{x,f}(r) \leq \frac{3}{2} V_d r^d f(x).$$

Now, by the triangle and Markov's inequalities, for every $s \in (0,1)$,

$$\mathbb{P}\bigg(\|X_1 - x\| > \|x\| + \Big(\frac{C}{1-s}\Big)^{1/\alpha}\bigg) \leq \mathbb{P}\bigg(\|X_1\| > \Big(\frac{C}{1-s}\Big)^{1/\alpha}\bigg) \leq 1-s,$$

so that

$$h_{x,f}^{-1}(s) \leq \|x\| + \Big(\frac{C}{1-s}\Big)^{1/\alpha}.$$

Hence,

$$h_{x,f}^{-1}(s) \leq \Big(\frac{2s}{V_d f(x)}\Big)^{1/d} \mathbb{1}_{\{h_{x,f}^{-1}(s) \leq 1/\{(6d)^{1/(\beta-\underline{\beta})} M_{f,\beta}(x)\}\}\}} + \Big\{\|x\| + \Big(\frac{C}{1-s}\Big)^{1/\alpha}\Big\} \mathbb{1}_{\{h_{x,f}^{-1}(s) > 1/\{(6d)^{1/(\beta-\underline{\beta})} M_{f,\beta}(x)\}\}\}}$$
$$\leq \Big(\frac{2s}{V_d f(x)}\Big)^{1/d}\Big[1 + (6d)^{1/(\beta-\underline{\beta})} M_{f,\beta}(x)\Big\{\|x\| + \Big(\frac{C}{1-s}\Big)^{1/\alpha}\Big\}\Big],$$

as required. □

The following lemma shows that we may restrict our main attention to the events

(S72) $$A_i^X := \{h_{X_i,f}(\rho^d_{(k_X),i,X}) \in \mathcal{I}_{m,X}\}, \quad A_i^Y := \{h_{X_i,g}(\rho^d_{(k_Y),i,Y}) \in \mathcal{I}_{n,Y}\},$$

for $i = 1, \ldots, n$.

LEMMA S9. *Fix $d \in \mathbb{N}$, $\vartheta \in \Theta$, $(\kappa_1, \kappa_2) \in \mathbb{R}^2$ and suppose that*

$$\frac{\kappa_1^-}{\lambda_1} + \frac{\kappa_2^-}{\lambda_2} + \frac{d(\kappa_1^- + \kappa_2^-)}{\alpha} \leq 1.$$

Let $k_X^{\mathrm{L}} \leq k_X^{\mathrm{U}}, k_Y^{\mathrm{L}} \leq k_Y^{\mathrm{U}}$ be deterministic sequences of positive integers such that $k_X^{\mathrm{L}}/\log m \to \infty$, $k_Y^{\mathrm{L}}/\log n \to \infty$, $k_X^{\mathrm{U}}/m \to 0$ and $k_Y^{\mathrm{U}}/n \to 0$. Then

$$\max_{\substack{k_X \in \{k_X^{\mathrm{L}},\ldots,k_X^{\mathrm{U}}\}\\ k_Y \in \{k_Y^{\mathrm{L}},\ldots,k_Y^{\mathrm{U}}\}}} \sup_{(f,g)\in\mathcal{F}_{d,\vartheta}} \mathbb{E}\Big[\max\Big\{\widehat{f}^{\kappa_1}_{(k_X),1}, \widehat{f}^{L}_{(k_X),1}, f(X_1)^{\kappa_1}\Big\} \times \max\Big\{\widehat{g}^{\kappa_2}_{(k_Y),1}, \widehat{g}^{L}_{(k_Y),1}, g(X_1)^{\kappa_2}\Big\}\big(1 - \mathbb{1}_{A_1^X}\mathbb{1}_{A_1^Y}\big)\Big] = o(m^{-4} + n^{-4})$$

as $m, n \to \infty$.

PROOF OF LEMMA S9. Given $a > -\min(k_X, k_Y), b > -\min(m - k_X, n + 1 - k_Y)$ define

$$\Delta^{(1)}_{a,b} := \int_{[0,1]\setminus\mathcal{I}_{m,X}} \mathrm{B}_{k_X+a,m-k_X+b}(s)\,ds, \quad \Delta^{(2)}_{a,b} := \int_{[0,1]\setminus\mathcal{I}_{n,Y}} \mathrm{B}_{k_Y+a,n+1-k_Y+b}(t)\,dt.$$

By Lemma S6 we have that

$$\max_{\substack{k_X \in \{k_X^{\mathrm{L}},\ldots,k_X^{\mathrm{U}}\}\\ k_Y \in \{k_Y^{\mathrm{L}},\ldots,k_Y^{\mathrm{U}}\}}} \sup_{a,b\in[-A,A]} \max(\Delta^{(1)}_{a,b}, \Delta^{(2)}_{a,b}) = o(m^{-9(1-\epsilon)/2} + n^{-9(1-\epsilon)/2})$$

for any fixed $A \geq 0$ and $\epsilon > 0$. Now, by Lemma S8 and writing $\kappa_i^+ := \max(\kappa_i, 0)$ for $i = 1, 2$, we have that

$$\mathbb{E}\Big[\max\Big\{\widehat{f}^{\kappa_1}_{(k_X),1}, \widehat{f}^{L}_{(k_X),1}, f(X_1)^{\kappa_1}\Big\}\max\Big\{\widehat{g}^{\kappa_2}_{(k_Y),1}, \widehat{g}^{L}_{(k_Y),1}, g(X_1)^{\kappa_2}\Big\}\big(1 - \mathbb{1}_{A_1^X}\mathbb{1}_{A_1^Y}\big)\Big]$$
$$= \int_{\mathcal{X}} f(x) \int_0^1\int_0^1 \max\big(u^{\kappa_1}_{x,s}, u^{L}_{x,s}, f(x)^{\kappa_1}\big)\max\big(v^{\kappa_2}_{x,t}, v^{L}_{x,t}, g(x)^{\kappa_2}\big) \times \max\big(\mathbb{1}_{\{s\notin\mathcal{I}_{m,X}\}}, \mathbb{1}_{\{t\notin\mathcal{I}_{n,Y}\}}\big)\mathrm{B}_{k_X,m-k_X}(s)\mathrm{B}_{k_Y,n+1-k_Y}(t)\,ds\,dt\,dx$$
$$\lesssim \int_{\mathcal{X}} f(x)\int_0^1\int_0^1 \max\Big\{\Big(\frac{k_X}{ms}\Big)^{\kappa_1^+\vee L}, \Big(\frac{msM_\beta(x)^d(1+\|x\|)^d}{k_X f(x)(1-s)^{d/\alpha}}\Big)^{\kappa_1^-}, f(x)^{\kappa_1}\Big\}$$

$$\times \max\Big\{\Big(\frac{k_Y}{nt}\Big)^{\kappa_2^+ \vee L}, \Big(\frac{ntM_\beta(x)^d(1+\|x\|)^d}{k_Y g(x)(1-t)^{d/\alpha}}\Big)^{\kappa_2^-}, g(x)^{\kappa_2}\Big\}$$

$$\times \max\big(\mathbb{1}_{\{s\notin\mathcal{I}_{m,X}\}}, \mathbb{1}_{\{t\notin\mathcal{I}_{n,Y}\}}\big)\mathrm{B}_{k_X,m-k_X}(s)\mathrm{B}_{k_Y,n+1-k_Y}(t)\,ds\,dt\,dx$$

$$\lesssim \max\big(\Delta^{(1)}_{-(\kappa_1^+\vee L),0}, \Delta^{(1)}_{0,0}, \Delta^{(1)}_{\kappa_1^-,-d\kappa_1^-/\alpha}, \Delta^{(2)}_{-(\kappa_2^+\vee L),0}, \Delta^{(2)}_{0,0}, \Delta^{(2)}_{\kappa_2^-,-d\kappa_2^-/\alpha}\big)$$

$$\times \int_{\mathcal{X}} f(x)\Big(\frac{M_\beta(x)^d}{f(x)}\Big)^{\kappa_1^-}\Big(\frac{M_\beta(x)^d}{g(x)}\Big)^{\kappa_2^-}(1+\|x\|)^{d(\kappa_1^-+\kappa_2^-)}\,dx$$

$$\lesssim (m^{-17/4}+n^{-17/4})\int_{\mathcal{X}} f(x)\Big(\frac{M_\beta(x)^d}{f(x)}\Big)^{\kappa_1^-}\Big(\frac{M_\beta(x)^d}{g(x)}\Big)^{\kappa_2^-}(1+\|x\|)^{d(\kappa_1^-+\kappa_2^-)}\,dx.$$

The conclusion follows immediately on appealing to Lemma S7. □

LEMMA S10. *Let $a, b, c \in \mathbb{R}$ be any fixed constants, and let $k^{\mathrm{L}} \le k^{\mathrm{U}}$ be deterministic sequences of positive integers such that $k^{\mathrm{L}} \to \infty$ and $k^{\mathrm{U}}/n \to 0$ as $n \to \infty$. Then*

$$\int_0^1\int_0^1 \Big|\mathrm{B}_{j+a,\ell+b,n+c-j-\ell}(s,t) - \mathrm{B}_{j+a,n-j}(s)\mathrm{B}_{\ell+b,n-\ell}(t)\Big|\,ds\,dt \le \frac{(j\ell)^{\frac{1}{2}}}{n}\{1+o(1)\}$$

as $n \to \infty$, uniformly for $j, \ell \in \{k^{\mathrm{L}}, \dots, k^{\mathrm{U}}\}$.

PROOF. In the following bound we make use the standard asymptotic expansions

$$\log\Gamma(z) = z\log z - z - \frac{1}{2}\log\Big(\frac{z}{2\pi}\Big) + \frac{1}{12z} + O\Big(\frac{1}{z^3}\Big)$$

$$\Psi(z) = \log z - \frac{1}{2z} - \frac{1}{12z^2} + O\Big(\frac{1}{z^4}\Big)$$

as $z \to \infty$. Using these expansions, by Lemma S6 and Pinsker's inequality we have that

$$\int_0^1\int_0^1 \Big|\mathrm{B}_{j+a,\ell+b,n+c-j-\ell}(s,t) - \mathrm{B}_{j+a,n-j}(s)\mathrm{B}_{\ell+b,n-\ell}(t)\Big|\,ds\,dt$$

$$\le \Big\{2\int_0^1\int_0^{1-t} \mathrm{B}_{j+a,\ell+b,n+c-j-\ell}(s,t)\log\Big(\frac{\mathrm{B}_{j+a,\ell+b,n+c-j-\ell}(s,t)}{\mathrm{B}_{j+a,n-j}(s)\mathrm{B}_{\ell+b,n-\ell}(t)}\Big)\,ds\,dt\Big\}^{1/2}$$

$$= 2^{\frac{1}{2}}\Big[\log\Big(\frac{\Gamma(n+a+b+c)\Gamma(n-j)\Gamma(n-\ell)}{\Gamma(n+c-j-\ell)\Gamma(n+a)\Gamma(n+b)}\Big) + (n-c-1)\Psi(n+a+b+c)$$

$$- (n-j-1)\Psi(n+b+c-j) - (n-\ell-1)\Psi(n+a+c-\ell)$$

$$+ (n+c-j-\ell-1)\Psi(n+c-j-\ell)\Big]^{1/2}$$

$$= \frac{(j\ell)^{1/2}}{n}\{1+o(1)\}$$

as $n \to \infty$, uniformly for $j, \ell \in \{k^{\mathrm{L}}, \dots, k^{\mathrm{U}}\}$. □

The following lemma provides bounds on the normal approximation to relevant multinomial distributions.

LEMMA S11. *Fix $f \in \mathcal{F}_d$ and $\beta \in (0,1]$, and let $k^{\mathrm{L}} \le k^{\mathrm{U}}$ be deterministic sequences of positive integers satisfying $k^{\mathrm{L}}/\log n \to \infty$ and $(k^{\mathrm{U}}/n)\log n \to 0$. For $k \in \{k^{\mathrm{L}}, \dots, k^{\mathrm{U}}\}$*

define $\mathcal{X}_n := \{x : f(x)/M_{f,\beta}(x)^d \geq (k/n)\log n\}$. *For* $j, \ell \in \mathbb{N}$ *and* $z \in \mathbb{R}^d$ *define* $y \equiv y_{x,z}^{(j)} := x + (\frac{j}{nV_d f(x)})^{1/d} z$, $\alpha_z(r) := \mu_d(B_0(1) \cap B_z(r))/V_d$, *and*

$$\Sigma := \begin{pmatrix} 1 & (j/\ell)^{1/2}\alpha_z((\ell/j)^{1/d}) \\ (j/\ell)^{1/2}\alpha_z((\ell/j)^{1/d}) & 1 \end{pmatrix}.$$

For $s, t \in (0,1), j, \ell \in \mathbb{N}$ *and* $x, z \in \mathbb{R}^d$ *let* $p_\cap = \int_{B_x(h_{x,f}^{-1}(s)) \cap B_y(h_{y,f}^{-1}(t))} f(w)\,dw$, *define* $(N_1, N_2, N_3, N_4) \sim \mathrm{Multi}(n; s - p_\cap, t - p_\cap, p_\cap, 1 - s - t + p_\cap)$, *let* $(M_1, M_2, M_3) \sim \mathrm{Multi}(n; s, t, 1 - s - t)$, *and*

$$F(s,t) := F_{n,x,z}^{(j),(\ell)}(s,t) := \mathbb{P}(N_1 + N_3 \geq j, N_2 + N_3 \geq \ell)$$
$$G(s,t) := G_n^{(j),(\ell)}(s,t) := \mathbb{P}(M_1 \geq j, M_2 \geq \ell).$$

Then, given $c \in (0,1)$ *and writing* Φ_V *for the distribution function of the bivariate normal distribution with mean zero and covariance matrix* V, *there exists* $A = A(d, \beta, c, (k^{\mathrm{L}}), (k^{\mathrm{U}}))$ *such that*

$$\max\Big\{\Big|F(s,t) - \Phi_\Sigma\Big(\frac{ns-j}{j^{1/2}}, \frac{nt-\ell}{\ell^{1/2}}\Big)\Big|, \Big|G(s,t) - \Phi_{I_2}\Big(\frac{ns-j}{j^{1/2}}, \frac{nt-\ell}{\ell^{1/2}}\Big)\Big|\Big\}$$
$$\leq A \min\bigg\{1, \frac{1}{\|z\|}\bigg(\frac{\log^{1/2} n}{k^{1/2}} + \Big(\frac{kM_{f,\beta}(x)^d}{nf(x)}\Big)^{\beta/d}\bigg)\bigg\}$$

for all $k \in \{k^{\mathrm{L}}, \ldots, k^{\mathrm{U}}\}$, *for all* $j, \ell \in \mathbb{N}$ *such that* $ck \leq j, \ell \leq k$, *for all* $x \in \mathcal{X}_n$, *for all* $s, t \in (0,1)$ *such that* $j^{-1/2}|ns - j| \vee \ell^{-1/2}|nt - \ell| \leq 3\log^{1/2} n$, *and for all* $0 < \|z\| \leq (\frac{nV_d f(x)}{j})^{1/d}\{h_{x,f}^{-1}(s) + h_{y,f}^{-1}(t)\}$.

PROOF. We present here the approximation for $F(s,t)$, the approximation for $G(s,t)$ being similar but much simpler. Let $X_1, \ldots, X_n \overset{\mathrm{iid}}{\sim} f$ and for $i = 1, \ldots, n$ and k, j, ℓ, x, s, t, z in the specified ranges, define $Y_i := (\mathbb{1}_{\{\|X_i - x\| \leq h_{x,f}^{-1}(s)\}}, \mathbb{1}_{\{\|X_i - y\| \leq h_{y,f}^{-1}(t)\}})^T$,

$$V := \mathrm{Cov}(Y_1) = \begin{pmatrix} s(1-s) & p_\cap - st \\ p_\cap - st & t(1-t) \end{pmatrix} \quad \text{and} \quad Z_i := V^{-1/2}(Y_i - (s,t)^T).$$

Then by the Berry–Esseen theorem of Götze (1991) we have

$$\big|\mathbb{P}(N_1 + N_3 \geq j, N_2 + N_3 \geq \ell) - \Phi_{nV}\big(ns - j, nt - \ell\big)\big| \lesssim n^{-1/2}\mathbb{E}(\|Z_1\|^3). \tag{S73}$$

In order to control the right hand side of this bound, we will require bounds on $p_\cap$. Writing α_z for $\alpha_z((\ell/j)^{1/d})$, we have

$$\Big|\frac{np_\cap}{j} - \alpha_z\Big| \leq \frac{n}{j}\Big|p_\cap - f(x)\mu_d\big(B_x(h_{x,f}^{-1}(s)) \cap B_y(h_{y,f}^{-1}(t))\big)\Big|$$
$$+ \Big|\frac{n}{j} f(x)\mu_d\big(B_x(h_{x,f}^{-1}(s)) \cap B_y(h_{y,f}^{-1}(t))\big) - \alpha_z\Big|$$
$$\leq \frac{n}{j} M_{f,\beta}(x)^\beta f(x) \int_{B_x(h_{x,f}^{-1}(s)) \cap B_y(h_{y,f}^{-1}(t))} \|w - x\|^\beta\,dw$$
$$+ \bigg|\frac{1}{V_d}\mu_d\bigg(B_0\bigg(\Big(\frac{nV_d f(x) h_{x,f}^{-1}(s)^d}{j}\Big)^{\frac{1}{d}}\bigg) \cap B_z\bigg(\Big(\frac{nV_d f(x) h_{y,f}^{-1}(t)^d}{j}\Big)^{\frac{1}{d}}\bigg) - \alpha_z\bigg|$$
$$\lesssim \frac{n}{k} M_{f,\beta}(x)^\beta f(x) h_{x,f}^{-1}(s)^{d+\beta} + \Big|\frac{nV_d f(x) h_{x,f}^{-1}(s)^d}{j} - 1\Big| + \Big|\frac{nV_d f(x) h_{y,f}^{-1}(t)^d}{l} - 1\Big|$$

$$\text{(S74)}\quad \lesssim \Big\{\frac{kM_{f,\beta}(x)^d}{nf(x)}\Big\}^{\beta/d} + \frac{\log^{1/2} n}{k^{1/2}},$$

where the final bound follows by Lemma S4 and similar arguments to those in (S76) and (S77) in the bounds on U_0 below. We will also need to bound $s+t-2p_\cap$ below. If $h_{y,f}^{-1}(t) \geq h_{x,f}^{-1}(s)$ then, by the mean value theorem and Lemma S4,

$$\begin{aligned}
&\mu_d\big(B_y(h_{y,f}^{-1}(t)) \cap B_x(h_{x,f}^{-1}(s))^c\big) \geq \mu_d\big(B_y(h_{x,f}^{-1}(s)) \cap B_x(h_{x,f}^{-1}(s))^c\big) \\
&\quad = V_d h_{x,f}^{-1}(s)^d \int_{(1-\frac{\|x-y\|^2}{4h_{x,f}^{-1}(s)^2})_+}^1 \mathrm{B}_{\frac{d+1}{2},\frac{1}{2}}(\xi)\,d\xi \gtrsim h_{x,f}^{-1}(s)^d \frac{\|x-y\| \wedge h_{x,f}^{-1}(s)}{h_{x,f}^{-1}(s)} \gtrsim \frac{k(\|z\|\wedge 1)}{nf(x)}.
\end{aligned}$$

A similar argument applies with (x,s) and (y,t) swapped and so we have

$$\begin{aligned}
s+t-2p_\cap &= \bigg(\int_{B_y(h_{y,f}^{-1}(t))\cap B_x(h_{x,f}^{-1}(s))^c} + \int_{B_x(h_{x,f}^{-1}(s))\cap B_y(h_{y,f}^{-1}(t))^c}\bigg) f(w)\,dw \\
&\gtrsim f(x)\big\{\mu_d\big(B_y(h_{y,f}^{-1}(t)) \cap B_x(h_{x,f}^{-1}(s))^c\big) + \mu_d\big(B_x(h_{x,f}^{-1}(s)) \cap B_y(h_{y,f}^{-1}(t))^c\big)\big\} \\
&\gtrsim \frac{k(\|z\|\wedge 1)}{n}.
\end{aligned}$$

We will also use a lower bound on $|V| := \det(V)$ when $\|z\| \geq 1$. Note that with $e_1 = (1,0,\ldots,0)^T \in \mathbb{R}^d$, when $\|z\| \geq 1$ we have that $\alpha_z = \alpha_{\|z\|e_1} \leq \alpha_{e_1}$. If $\ell/j \geq (3/2)^d$ then

$$\frac{j\alpha_z^2}{\ell} \leq \frac{j\alpha_{e_1}^2}{\ell} \leq \frac{j}{\ell} \leq \Big(\frac{2}{3}\Big)^{d/2} < 1.$$

However if $\ell/j < (3/2)^d$ then

$$\frac{j\alpha_z^2}{\ell} \leq \alpha^2 < V_d^{-2}\mu_d\big(B_0(1)\cap B_{e_1}((3/2)^{1/d})\big)^2 < 1.$$

Thus there exists $c_d \in (0,1)$ such that $j\alpha_z^2/\ell \leq c_d$ whenever $\|z\| \geq 1$. Thus, by (S74), we have that

$$|V| = st(1-s)(1-t) - (p_\cap - st)^2 \geq \frac{(1-c_d)j\ell}{n^2}\{1+o(1)\},$$

uniformly over $\|z\| \geq 1$. Similar to (36), (37) and (38) in the supplement of Berrett, Samworth and Yuan (2019), and splitting up into cases $\|z\| < 1$ and $\|z\| \geq 1$ where necessary, we have that

$$\begin{aligned}
p_\cap \bigg\| V^{-1/2}\begin{pmatrix}1-s\\1-t\end{pmatrix}\bigg\|^3 &\leq p_\cap \min\Big\{\frac{s+t}{|V|}, \frac{1}{p_\cap - st}\Big\}^{3/2} \lesssim \frac{n^{1/2}}{k^{1/2}}, \\
(1-s-t+p_\cap)\bigg\| V^{-1/2}\begin{pmatrix}s\\t\end{pmatrix}\bigg\|^3 &= (1-s-t+p_\cap)\Big\{\frac{st(s+t-2p_\cap)}{|V|}\Big\}^{\frac{3}{2}} \lesssim \frac{k^{\frac{3}{2}}}{n^{\frac{3}{2}}}.
\end{aligned}$$

Likewise,

$$\begin{aligned}
&(s-p_\cap)\bigg\| V^{-1/2}\begin{pmatrix}1-s\\-t\end{pmatrix}\bigg\|^3 \leq (s-p_\cap)t^{3/2}|V|^{-3/2} \\
&= (s-p_\cap)t^{\frac{3}{2}}\Big\{(s+t-2p_\cap)\Big(p_\cap - st + \frac{(s-p_\cap)(t-p_\cap)}{s+t-2p_\cap}\Big)\Big\}^{-\frac{3}{2}} \lesssim \Big(\frac{n}{k\|z\|}\Big)^{\frac{1}{2}},
\end{aligned}$$

with a similar bound holding for $(t-p_\cap)\left\|V^{-1/2}\begin{pmatrix}-s\\1-t\end{pmatrix}\right\|^3$. Thus

$$n^{-1/2}\mathbb{E}\|Z_3\|^3 \lesssim (k\|z\|)^{-1/2},$$

which in combination with (S73) provides a bound on the difference between $F(s,t)$ and $\Phi_{nV}(ns-j,nt-\ell)$. Next, similar to the displayed equation above (39) in the supplement of Berrett, Samworth and Yuan (2019), we have

$$\begin{aligned}
&\big|\Phi_{nV}\big(ns-j,nt-\ell\big)-\Phi_\Sigma\big(j^{-1/2}(ns-j),\ell^{-1/2}(nt-\ell)\big)\big| \\
&\leq \min\left\{1,2\left\|\Sigma^{-\frac12}\begin{pmatrix}\frac{ns(1-s)}{j}-1 & \frac{n(p_\cap-st)}{j^{1/2}\ell^{1/2}}-j^{1/2}\ell^{-1/2}\alpha_z\\ \frac{n(p_\cap-st)}{j^{1/2}\ell^{1/2}}-j^{1/2}\ell^{-1/2}\alpha_z & \frac{nt(1-t)}{\ell}-1\end{pmatrix}\Sigma^{-\frac12}\right\|\right\} \\
&\lesssim \{1/(1-(j/\ell)^{1/2}\alpha_z)+1/(1+(j/\ell)^{1/2}\alpha_z)\}\left\{\frac{\log^{1/2}n}{k^{1/2}}+\frac{j^{1/2}}{\ell^{1/2}}\left|\frac{np_\cap}{j}-\alpha_z\right|\right\} \\
&\lesssim \frac{1}{\|z\|}\left\{\frac{\log^{1/2}n}{k^{1/2}}+\Big(\frac{kM_{f,\beta}(x)^d}{nf(x)}\Big)^{\beta/d}\right\}
\end{aligned}$$

as required. □

S1.9. *Bounds on remainder terms in the proof of Proposition 11.* *To bound S_1:* Since $\zeta<1/2$ we may apply Lemma S9 to see that

$$\begin{aligned}
S_{11}:=\int_{\mathcal{X}} f(x)\int_0^1\int_0^1 \max\big(\mathbb{1}_{\{s\notin\mathcal{I}_{m,X}\}},\mathbb{1}_{\{t\notin\mathcal{I}_{n,Y}\}}\big)\phi(u_{x,s},v_{x,t})^2 \\
\times \mathrm{B}_{k_X,m-k_X}(s)\mathrm{B}_{k_Y,n+1-k_Y}(t)\,ds\,dt\,dx = o(m^{-4}+n^{-4}).
\end{aligned}$$

By Lemma S8 we have that for every $\epsilon>0$,

$$\begin{aligned}
|S_{12}| &:= |S_1-S_{11}| \\
&= \left|\int_{\mathcal{X}_{m,n}^c} f(x)\int_{\mathcal{I}_{m,X}}\int_{\mathcal{I}_{n,Y}} \phi\big(u_{x,s},v_{x,t}\big)^2\mathrm{B}_{k_X,m-k_X}(s)\mathrm{B}_{k_Y,n+1-k_Y}(t)\,ds\,dt\,dx\right| \\
&\lesssim \int_{\mathcal{X}_{m,n}^c} f(x)^{1-2\kappa_1^-}g(x)^{-2\kappa_2^-}M_\beta(x)^{2d(\kappa_1^-+\kappa_2^-)}(1+\|x\|)^{2d(\kappa_1^-+\kappa_2^-)}\,dx \\
&= O\left(\max\left\{\Big(\frac{k_X}{m}\Big)^{\lambda_1(1-2\zeta)-\epsilon},\Big(\frac{k_Y}{n}\Big)^{\lambda_2(1-2\zeta)-\epsilon}\right\}\right),
\end{aligned}$$

where the final bound holds by Lemma S7, as in the bound on R_1.

To bound S_2: Using Lemma S4 we now have that

$$\begin{aligned}
|S_2| = \Bigg|\int_{\mathcal{X}_{m,n}} f(x)\int_{\mathcal{I}_{m,X}}\int_{\mathcal{I}_{n,Y}}\Big\{\phi\big(u_{x,s},v_{x,t}\big)^2 \\
-\phi\Big(\frac{k_Xf(x)}{ms},\frac{k_Yg(x)}{nt}\Big)^2\Big\}\mathrm{B}_{k_X,m-k_X}(s)\mathrm{B}_{k_Y,n+1-k_Y}(t)\,ds\,dt\,dx\Bigg| \\
\lesssim \int_{\mathcal{X}_{m,n}} f(x)^{1+2\kappa_1}g(x)^{2\kappa_2}\int_{\mathcal{I}_{m,X}}\int_{\mathcal{I}_{n,Y}}\mathrm{B}_{k_X,m-k_X}(s)\mathrm{B}_{k_Y,n+1-k_Y}(t) \\
\times\left\{\left|\frac{s}{V_df(x)h_{x,f}^{-1}(s)^d}-1\right|+\left|\frac{t}{V_dg(x)h_{x,g}^{-1}(t)^d}-1\right|\right\}ds\,dt\,dx
\end{aligned}$$

$$\lesssim \int_{\mathcal{X}_{m,n}} f(x)^{1+2\kappa_1} g(x)^{2\kappa_2} \biggl\{ \Bigl(\frac{k_X M_\beta(x)^d}{mf(x)}\Bigr)^{\frac{2\wedge\beta}{d}} + \Bigl(\frac{k_Y M_\beta(x)^d}{ng(x)}\Bigr)^{\frac{2\wedge\beta}{d}} \biggr\}\, dx$$
$$= O\biggl(\max\biggl\{\Bigl(\frac{k_X}{m}\Bigr)^{\frac{2\wedge\beta}{d}}, \Bigl(\frac{k_X}{m}\Bigr)^{\lambda_1(1-2\zeta)-\epsilon}, \Bigl(\frac{k_Y}{n}\Bigr)^{\frac{2\wedge\beta}{d}}, \Bigl(\frac{k_Y}{n}\Bigr)^{\lambda_2(1-2\zeta)-\epsilon}\biggr\}\biggr)$$

for all $\epsilon > 0$, where for the final bound we use Lemma S7 as in (S10) and (S11).

To bound S_3: Using Lemma S9 and Lemma S7 we may write

$$|S_3| = \biggl| \int_{\mathcal{X}_{m,n}} f(x) \int_{\mathcal{I}_{m,X}} \int_{\mathcal{I}_{n,Y}} \phi\Bigl(\frac{k_X f(x)}{ms}, \frac{k_Y g(x)}{nt}\Bigr)^2$$
$$\times \mathrm{B}_{k_X, m-k_X}(s) \mathrm{B}_{k_Y, n+1-k_Y}(t)\, ds\, dt\, dx - \int_{\mathcal{X}} f(x)\phi_x^2\, dx \biggr|$$
$$\leq \int_{\mathcal{X}_{m,n}} f(x) \int_{\mathcal{I}_{m,X}} \int_{\mathcal{I}_{n,Y}} \biggl|\phi\Bigl(\frac{k_X f(x)}{ms}, \frac{k_Y g(x)}{nt}\Bigr)^2 - \phi_x^2\biggr|$$
$$\times \mathrm{B}_{k_X, m-k_X}(s) \mathrm{B}_{k_Y, n+1-k_Y}(t)\, ds\, dt\, dx + \int_{\mathcal{X}_{m,n}^c} f(x)\phi_x^2\, dx + o(m^{-4} + n^{-4})$$
$$\lesssim \bigl(k_X^{-\frac{1}{2}} + k_Y^{-\frac{1}{2}}\bigr) \int_{\mathcal{X}_{m,n}} f(x)^{1+2\kappa_1} g(x)^{2\kappa_2}\, dx + \int_{\mathcal{X}_{m,n}^c} f(x)^{1+2\kappa_1} g(x)^{2\kappa_2}\, dx + o(m^{-4} + n^{-4})$$
$$= O\biggl(\max\biggl\{k_X^{-1/2}, k_Y^{-1/2}, \Bigl(\frac{k_X}{m}\Bigr)^{\lambda_1(1-2\zeta)-\epsilon}, \Bigl(\frac{k_Y}{n}\Bigr)^{\lambda_2(1-2\zeta)-\epsilon}\biggr\}\biggr),$$

for every $\epsilon > 0$.

To bound T_1: We first consider

$$T_{11} := \biggl(\int_{\mathcal{X}^2} - \int_{\mathcal{X}_{m,f}^2}\biggr) \int_{\mathcal{I}_{m,X}^2} \int_{\mathcal{I}_{n,Y}^2} (h\, dH_m^{(1)}\, dG_n^{(2)})(s_1, s_2, t_1, t_2)\, dx\, dy.$$

By symmetry we may write $T_{11} = T_{111} + 2T_{112}$, where

$$T_{111} := \int_{\mathcal{X}_{m,f}^c \times \mathcal{X}_{m,f}^c} f(x)f(y) \int_{\mathcal{I}_{m,X}^2} \int_{\mathcal{I}_{n,Y}^2} (h\, dH_m^{(1)}\, dG_n^{(2)})(s_1, s_2, t_1, t_2)\, dx\, dy$$

and

$$T_{112} := \int_{\mathcal{X}_{m,f} \times \mathcal{X}_{m,f}^c} f(x)f(y) \int_{\mathcal{I}_{m,X}^2} \int_{\mathcal{I}_{n,Y}^2} (h\, dH_m^{(1)}\, dG_n^{(2)})(s_1, s_2, t_1, t_2)\, dx\, dy.$$

Using Lemma S7 and Lemma S8 as in the bounds on S_1, and using Lemma S10 we have that

$$|T_{111}| \lesssim \frac{k_X}{m} \biggl\{ \int_{\mathcal{X}_{m,f}^c} f(x)^{1-\kappa_1^-} g(x)^{-\kappa_2^-} M_\beta(x)^{d(\kappa_1^- + \kappa_2^-)} (1 + \|x\|)^{d(\kappa_1^- + \kappa_2^-)}\, dx \biggr\}^2$$
$$= O\biggl(\Bigl(\frac{k_X}{m}\Bigr)^{1+2\lambda_1(1-\zeta)-\epsilon}\biggr)$$

for all $\epsilon > 0$. We now turn to T_{112}, and similarly write

$$|T_{112}| = \biggl| \int_{\mathcal{X}_{m,f} \times \mathcal{X}_{m,f}^c} f(x)f(y) \int_{\mathcal{I}_{m,X}^2} \int_{\mathcal{I}_{n,Y}^2} (h\, dH_m^{(1)}\, dG_n^{(2)})(s_1, s_2, t_1, t_2)\, dx\, dy \biggr|$$

$$\lesssim \frac{k_X}{m} \int_{\mathcal{X}_{m,f} \times \mathcal{X}_{m,f}^c} \frac{f(x)^{1+\kappa_1} g(x)^{\kappa_2} f(y)}{f(y)^{\kappa_1^-} g(y)^{\kappa_2^-}} M_\beta(y)^{d(\kappa_1^- + \kappa_2^-)} (1 + \|y\|)^{d(\kappa_1^- + \kappa_2^-)} \, dy \, dx$$
$$= O\biggl(\Bigl(\frac{k_X}{m} \Bigr)^{1+\lambda_1(1-\zeta)-\epsilon} \biggr)$$

for all $\epsilon > 0$. Combining our bounds on T_{111} and T_{112} we have that

$$T_{11} = O\biggl(\Bigl(\frac{k_X}{m} \Bigr)^{1+\lambda_1(1-\zeta)-\epsilon} \biggr)$$

for all $\epsilon > 0$. We can develop analogous bounds on

$$T_{12} := \biggl(\int_{\mathcal{X}^2} - \int_{\mathcal{X}_{n,g}^2} \biggr) \int_{\mathcal{I}_{m,X}^2} \int_{\mathcal{I}_{n,Y}^2} (h \, d(G_m^{(1)} - H_m^{(1)}) \, dH_n^{(2)})(s_1, s_2, t_1, t_2) \, dx \, dy$$

to conclude that

$$T_1 = T_{11} + T_{12} = O\biggl(\max\biggl\{ \Bigl(\frac{k_X}{m} \Bigr)^{1+\lambda_1(1-\zeta)-\epsilon}, \Bigl(\frac{k_Y}{n} \Bigr)^{1+\lambda_2(1-\zeta)-\epsilon} \biggr\} \biggr)$$

for all $\epsilon > 0$.

To bound T_2: Here we use the notation

$$L_x^f(s,t) := \phi\bigl(f(x), v_{x,t}\bigr) + \Bigl(\frac{k_X}{ms} - 1 \Bigr) f(x) \phi_{10}\bigl(f(x), v_{x,t}\bigr)$$
$$R_x^f(s,t) := \phi\bigl(u_{x,s}, v_{x,t}\bigr) - L_x^f(s,t)$$

for a linearised version of $\phi(u_{x,s}, v_{x,t})$ and the linearisation error, so that we have $h^{(1)}(s_1, s_2, t_1, t_2) = L_x^f(s_1, t_1) L_y^f(s_2, t_2)$. Again we write $T_2 = T_{21} + T_{22}$, with

$$T_{21} := \int_{\mathcal{X}_{m,f}^2} f(x) \int_{\mathcal{I}_{m,X}^2} \int_{\mathcal{I}_{n,Y}^2} (\{h - h^{(1)}\} \, dH_m^{(1)} \, dG_n^{(2)})(s_1, s_2, t_1, t_2) \, dx \, dy$$
$$= \int_{\mathcal{X}_{m,f}^2} f(x) \int_{\mathcal{I}_{m,X}^2} \int_{\mathcal{I}_{n,Y}^2} \bigl\{ R_x^f(s_1, t_1) R_y^f(s_2, t_2) + 2 L_x^f(s_1, t_1) R_y^f(s_2, t_2) \bigr\}$$
$$\times \, dH_m^{(1)}(s_1, s_2) \, dG_n^{(2)}(t_1, t_2) \, dx \, dy$$
$$=: T_{211} + T_{212}$$

and $T_{22} := T_2 - T_{21}$ having a similar expression. Now

$$|T_{211}| \lesssim \frac{k_X}{m} \biggl[\int_{\mathcal{X}_{m,f}} \frac{f(x)^{1+\kappa_1}}{g(x)^{\kappa_2^-}} M_\beta(x)^{d\kappa_2^-} (1 + \|x\|)^{d\kappa_2^-} \biggl\{ \Bigl(\frac{k_X M_\beta(x)^d}{m f(x)} \Bigr)^{\frac{2 \wedge \beta}{d}} + \frac{\log m}{k_X} \biggr\} \, dx \biggr]^2$$
$$= O\biggl(\frac{k_X}{m} \max\biggl\{ \Bigl(\frac{k_X}{m} \Bigr)^{\frac{2(2 \wedge \beta)}{d}}, \Bigl(\frac{k_X}{m} \Bigr)^{2\lambda_1(1-\zeta)-\epsilon}, \frac{\log^2 m}{k_X^2} \biggr\} \biggr)$$

for every $\epsilon > 0$. When bounding T_{212} we first integrate over s_1 using the facts that

$$\int_0^1 \{ \mathrm{B}_{k_X, k_X, m-2k_X-1}(s_1, s_2) - \mathrm{B}_{k_X, m-k_X}(s_1) \mathrm{B}_{k_X, m-k_X}(s_2) \} \, ds_1$$
$$= \frac{m-1}{m-k_X-1} \mathrm{B}_{k_X, m-k_X-1}(s_2) \Bigl(s_2 - \frac{k_X}{m-1} \Bigr)$$

and

$$\int_0^1 \frac{k_X}{ms_1}\{\mathrm{B}_{k_X,k_X,m-2k_X-1}(s_1,s_2)-\mathrm{B}_{k_X,m-k_X}(s_1)\mathrm{B}_{k_X,m-k_X}(s_2)\}\,ds_1$$
$$=\frac{k_X(m-2)}{m(k_X-1)}\mathrm{B}_{k_X,m-k_X-2}(s_2)\biggl\{1-\frac{(m-1)^2}{(m-k_X-1)(m-k_X-2)}(1-s_2)^2\biggr\}$$
$$=\mathrm{B}_{k_X,m-k_X-2}(s_2)\biggl\{2\Bigl(\frac{k_X}{m-2}-s_2\Bigr)+O\Bigl(\frac{k_X^2}{m^2}+\frac{1}{m}\Bigr)\biggr\},$$

uniformly for $s_2\in\mathcal{I}_{m,X}$. Using (S9) and the fact that $k_X^{3/2}/m\to 0$ we can now see that

$$|T_{212}|\lesssim \frac{k_X^{1/2}}{m}\int_{\mathcal{X}_{m,f}}\frac{f(y)^{1+\kappa_1}}{g(y)^{\kappa_2^-}}M_\beta(y)^{d\kappa_2^-}(1+\|y\|)^{d\kappa_2^-}\biggl\{\Bigl(\frac{k_X M_\beta(y)^d}{mf(y)}\Bigr)^{\frac{2\wedge\beta}{d}}+\frac{\log m}{k_X}\biggr\}\,dy$$
$$=O\biggl(\frac{k_X^{1/2}}{m}\max\biggl\{\Bigl(\frac{k_X}{m}\Bigr)^{\frac{2\wedge\beta}{d}},\Bigl(\frac{k_X}{m}\Bigr)^{\lambda_1(1-\zeta)-\epsilon},\frac{\log m}{k_X}\biggr\}\biggr)$$

for every $\epsilon>0$. Combining our bounds on T_{211} and T_{212} we therefore have that

$$|T_{21}|=O\biggl(\max\biggl\{\Bigl(\frac{k_X}{m}\Bigr)^{1+\lambda_1(1-\zeta)-\epsilon},\frac{\log m}{mk_X^{1/2}},\frac{k_X^{1/2}}{m}\Bigl(\frac{k_X}{m}\Bigr)^{\frac{2\wedge\beta}{d}},\Bigl(\frac{k_X}{m}\Bigr)^{1+\frac{2(2\wedge\beta)}{d}}\biggr\}\biggr)$$

for every $\epsilon>0$. By analogous arguments we can show that

$$|T_{22}|=O\biggl(\max\biggl\{\Bigl(\frac{k_Y}{n}\Bigr)^{1+\lambda_2(1-\zeta)-\epsilon},\frac{\log n}{nk_Y^{1/2}},\frac{k_Y^{1/2}}{n}\Bigl(\frac{k_Y}{n}\Bigr)^{\frac{2\wedge\beta}{d}},\Bigl(\frac{k_Y}{n}\Bigr)^{1+\frac{2(2\wedge\beta)}{d}}\biggr\}\biggr),$$

for every $\epsilon>0$, and this concludes the bound on T_2.

To bound T_3*:* Here we integrate out (s_1,s_2) in the $\mathcal{X}_{m,f}$ term and (t_1,t_2) in the $\mathcal{X}_{n,g}$ term. Now

$$\int_0^1\int_0^{1-s_1} h^{(1)}(s_1,s_2,t_1,t_2)\,dG_m^{(1)}(s_1,s_2)$$
$$-\int_0^1\int_0^1 h^{(1)}(s_1,s_2,t_1,t_2)\mathrm{B}_{k_X,m-k_X}(s_1)\mathrm{B}_{k_X,m-k_X}(s_2)\,ds_1\,ds_2$$
$$=f(x)\phi_{10}(f(x),v_{x,t_1})f(y)\phi_{10}(f(y),v_{y,t_2})$$
$$\times\biggl\{\int_0^1\int_0^{1-s_1}\Bigl(\frac{k_X}{ms_1}-1\Bigr)\Bigl(\frac{k_X}{ms_2}-1\Bigr)dG_m^{(1)}(s_1,s_2)$$
$$-\int_0^1\int_0^1\Bigl(\frac{k_X}{ms_1}-1\Bigr)\Bigl(\frac{k_X}{ms_2}-1\Bigr)\mathrm{B}_{k_X,m-k_X}(s_1)\mathrm{B}_{k_X,m-k_X}(s_2)\,ds_1\,ds_2\biggr\}$$
$$+\{f(x)\phi_{10}(f(x),v_{x,t_1})\phi(f(y),v_{y,t_2})+\phi(f(x),v_{x,t_1})f(y)\phi_{10}(f(y),v_{y,t_2})\}$$
$$\times\int_0^1\frac{k_X}{ms}\{\mathrm{B}_{k_X,m-k_X-1}(s)-\mathrm{B}_{k_X,m-k_X}(s)\}\,ds$$
$$=-\frac{k_X}{(k_X-1)m}\biggl[\biggl\{\frac{k_X(3m-5)}{(k_X-1)m}-2\biggr\}f(x)\phi_{10}(f(x),v_{x,t_1})f(y)\phi_{10}(f(y),v_{y,t_2})$$
$$+\{f(x)\phi_{10}(f(x),v_{x,t_1})\phi(f(y),v_{y,t_2})+\phi(f(x),v_{x,t_1})f(y)\phi_{10}(f(y),v_{y,t_2})\}\biggr].$$

The contribution from the $\mathcal{X}_{n,g}$ term is simpler because the marginals of the $\mathrm{B}_{k_Y,k_Y,n-2k_Y+1}$ density are equal to $\mathrm{B}_{k_Y,n-k_Y+1}$, and we have

$$\begin{aligned}&\int_0^1\int_0^{1-t_1} h^{(2)}(s_1,s_2,t_1,t_2)dG_n^{(2)}(t_1,t_2)\\&\qquad\qquad-\int_0^1\int_0^1 h^{(2)}(s_1,s_2,t_1,t_2)\mathrm{B}_{k_Y,n-k_Y+1}(t_1)\mathrm{B}_{k_Y,n-k_Y+1}(t_2)\,dt_1\,dt_2\\&=g(x)\phi_{01}(u_{x,s_1},g(x))g(y)\phi_{01}(u_{y,s_2},g(y))\biggl\{\int_0^1\int_0^{1-t_1}\frac{k_Y^2}{n^2t_1t_2}dG_n^{(2)}(t_1,t_2)\\&\qquad\qquad\qquad-\int_0^1\int_0^1\frac{k_Y^2}{n^2t_1t_2}\mathrm{B}_{k_Y,n-k_Y+1}(t_1)\mathrm{B}_{k_Y,n-k_Y+1}(t_2)\,dt_1\,dt_2\biggr\}\\&=-\frac{k_Y^2}{(k_Y-1)^2n}g(x)\phi_{01}(u_{x,s_1},g(x))g(y)\phi_{01}(u_{y,s_2},g(y)).\end{aligned}$$

The error T_3 is the error in, for example, $k_Y^2(k_Y-1)^{-2}/n\approx 1/n$, together with the contribution from $(s_1,s_2)\notin\mathcal{I}_{m,X}^2$ and $(t_1,t_2)\notin\mathcal{I}_{n,Y}^2$, and we can use Lemma S6 to see that

$$T_3=o(1/m+1/n).$$

To bound U_0: We write $r_{m,x,y}^{(1)}:=h_{x,f}^{-1}(a_{m,X}^+)+h_{y,f}^{-1}(a_{m,X}^+)$ and $r_{n,x,y}^{(2)}:=h_{x,g}^{-1}(a_{n,Y}^+)+h_{y,g}^{-1}(a_{n,Y}^+)$ as shorthand. For $s_1,s_2\le a_{m,X}^+$ and $t_1,t_2\le a_{n,Y}^+$ we have $F_{m,n,x,y}(s_1,s_2,t_1,t_2)=G_{m,n}(s_1,s_2,t_1,t_2)$ unless we also have $\|y-x\|\le\max\{r_{m,x,y}^{(1)},r_{n,x,y}^{(2)}\}$. Here we will present bounds in the case $\|y-x\|\le r_{m,x,y}^{(1)}$, but the other case follows using very similar arguments. First, by using Lemma S7 and Lemma S8, we have that

$$\begin{aligned}&\int_{\mathcal{X}_{m,n}^c} f(x)\sup_{s\in\mathcal{I}_{m,X},t\in\mathcal{I}_{n,Y}}|\phi(u_{x,s},v_{x,t})|\,dx\\&\qquad\qquad\lesssim\int_{\mathcal{X}_{m,n}^c} f(x)^{1-\kappa_1^-}g(x)^{-\kappa_2^-}M_\beta(x)^{d(\kappa_1^-+\kappa_2^-)}(1+\|x\|)^{d(\kappa_1^-+\kappa_2^-)}\,dx\\&\qquad\qquad=O\biggl(\max\biggl\{\Bigl(\frac{k_X}{m}\Bigr)^{\lambda_1(1-\zeta)-\epsilon},\Bigl(\frac{k_Y}{n}\Bigr)^{\lambda_2(1-\zeta)-\epsilon}\biggr\}\biggr),\end{aligned}\tag{S75}$$

for every $\epsilon>0$, and we proceed by showing that, since x and y are close, the contribution from $\mathcal{X}_{m,n}^c\times\mathcal{X}$ behaves similarly to the contribution from $\mathcal{X}_{m,n}^c\times\mathcal{X}_{m,n}^c$, which can be bounded by the square of the final bound in (S75). It suffices to consider $(x,y)\in\mathcal{X}_{m,n}^c\times\mathcal{X}_{m,n}$, as the contribution from $\mathcal{X}_{m,n}^c\times\mathcal{X}_{m,n}^c$ is more straightforward.

By a very similar argument to that used to establish (S45) in the proof of Proposition S2, we have that $\|x-y\|\le\{M_\beta(y)^d\log^{1/2}m\}^{-1/d}$ for m sufficiently large, and hence, by Lemma S5, that

$$|f(x)/f(y)-1|\le 2\{M_\beta(y)\|y-x\|\}^{1\wedge\beta}\le 1/2,$$

and in particular $f(x)\ge f(y)/2$. Thus, again using Lemma S5, we have that

$$\max_{t=1,\ldots,\underline{\beta}}\biggl(\frac{\|f^{(t)}(x)\|}{f(x)}\biggr)^{1/t}\le 4M_\beta(y).\tag{S76}$$

for m sufficiently large. In addition,

(S77)
$$\sup_{w,z\in B_x(1/\{2M_\beta(y)\})}\frac{\|f^{(\underline{\beta})}(z)-f^{(\underline{\beta})}(w)\|}{\|z-w\|^{\beta-\underline{\beta}}f(w)}\leq \sup_{w,z\in B_y(1/M_\beta(y))}\frac{\|f^{(\underline{\beta})}(z)-f^{(\underline{\beta})}(w)\|}{\|z-w\|^{\beta-\underline{\beta}}f(w)}\leq M_\beta(y)^\beta,$$

and so we have that $M_\beta(x)\leq 4M_\beta(y)$. Using this fact and the previously established fact that $f(x)\geq f(y)/2$, we may apply Lemma S4 to see that in fact

$$\|x-y\|\leq r^{(1)}_{m,x,y}\lesssim\Big(\frac{k_X}{mf(y)}\Big)^{1/d}.$$

Using Lemma S5 we also have that $g(x)\geq g(y)/2$, and therefore that

$$\max\Big\{\frac{f(y)M_\beta(y)^{-d}}{f(x)M_\beta(x)^{-d}},\frac{g(y)M_\beta(y)^{-d}}{g(x)M_\beta(x)^{-d}}\Big\}\leq 2^{2d+1}. \tag{S78}$$

Since $x\in\mathcal{X}^c_{m,n}$, we have now established that

$$\min\Big\{\frac{mf(y)M_\beta(y)^{-d}}{k_X\log m},\frac{ng(y)M_\beta(y)^{-d}}{k_Y\log n}\Big\}\leq 2^{2d+1}.$$

Applying the same bounds as we would for $\mathcal{X}^c_{m,n}$, as in (S75), we can now see that

$$U_0=O\bigg(\max\Big\{\Big(\frac{k_X}{m}\Big)^{2\lambda_1(1-\zeta)-\epsilon},\Big(\frac{k_Y}{n}\Big)^{2\lambda_2(1-\zeta)-\epsilon}\Big\}\bigg),$$

for every $\epsilon>0$, as claimed.

To bound U_1: By Lemma S6 we have that

$$\max_{t_1\in\{a^-_{n,Y},a^+_{n,Y}\}}\sup_{t_2\in[0,1]}|F^{(2)}_{n,x,y}-G^{(2)}_n|(t_1,t_2)$$
$$\bigvee\max_{t_2\in\{a^-_{n,Y},a^+_{n,Y}\}}\sup_{t_1\in[0,1]}|F^{(2)}_{n,x,y}-G^{(2)}_n|(t_1,t_2)=o(n^{-4}). \tag{S79}$$

In order to use this to bound U_1, corresponding to the right-hand side of (S32), we must first develop bounds on the derivatives of h. Writing $S_x(r):=\{y\in\mathbb{R}^d:\|x-y\|=r\}$ and $d\mathrm{Vol}_S$ for the associated volume element we have by Lemma S5 that for $r\leq 1/\{(6d)^{1/(\beta-\underline{\beta})}M_\beta(x)\}$,

$$\Big|\frac{h'_{x,f}(r)}{dV_dr^{d-1}f(x)}-1\Big|=\Big|\frac{1}{dV_dr^{d-1}f(x)}\int_{S_x(r)}\{f(y)-f(x)\}\,d\mathrm{Vol}_S(y)\Big|\lesssim\{rM_\beta(x)\}^{2\wedge\beta},$$

with a similar bound holding for $h'_{x,g}(r)$. Using Lemma S4, for $x\in\mathcal{X}_{m,n}$, we have that $\max\{M_\beta(x)^dh^{-1}_{x,f}(a^+_{m,X})^d,M_\beta(x)^dh^{-1}_{x,g}(a^+_{n,Y})^d\}\lesssim 1/\log m\to 0$ and so we have, by Lemma S3(i), that

$$\begin{aligned}
&\Big|\frac{\partial}{\partial s}\phi(u_{x,s},v_{x,t})+\frac{k_Xf(x)}{ms^2}\phi_{10}\Big(\frac{k_Xf(x)}{ms},\frac{k_Yg(x)}{nt}\Big)\Big|\\
&\quad=\Big|-\frac{k_Xd\,\phi_{10}(u_{x,s},v_{x,t})}{mV_dh^{-1}_{x,f}(s)^{d+1}h'_{x,f}(h^{-1}_{x,f}(s))}+\frac{k_Xf(x)}{ms^2}\phi_{10}\Big(\frac{k_Xf(x)}{ms},\frac{k_Yg(x)}{nt}\Big)\Big|\\
&\quad\leq\frac{k_Xd}{mV_dh^{-1}_{x,f}(s)^{d+1}h'_{x,f}(h^{-1}_{x,f}(s))}\Big|\phi_{10}(u_{x,s},v_{x,t})-\phi_{10}\Big(\frac{k_Xf(x)}{ms},\frac{k_Yg(x)}{nt}\Big)\Big|
\end{aligned}$$

$$+\frac{k_X f(x)}{ms^2}\Big|\frac{dV_d h_{x,f}^{-1}(s)^{d-1} f(x)/h'_{x,f}(h_{x,f}^{-1}(s))}{V_d^2 f(x)^2 h_{x,f}^{-1}(s)^{2d}/s^2}-1\Big|\Big|\phi_{10}\Big(\frac{k_X f(x)}{ms},\frac{k_Y g(x)}{nt}\Big)\Big|$$

$$\lesssim \frac{k_X}{ms^2} f(x)^{\kappa_1} g(x)^{\kappa_2}\Big\{\Big(\frac{sM_\beta(x)^d}{f(x)}\Big)^{(2\wedge\beta)/d}+\Big(\frac{tM_\beta(x)^d}{g(x)}\Big)^{(2\wedge\beta)/d}\Big\}$$

$$\lesssim (1/s) f(x)^{\kappa_1} g(x)^{\kappa_2}\Big\{\Big(\frac{k_X M_\beta(x)^d}{mf(x)}\Big)^{(2\wedge\beta)/d}+\Big(\frac{k_Y M_\beta(x)^d}{ng(x)}\Big)^{(2\wedge\beta)/d}\Big\}, \quad \text{(S80)}$$

uniformly for $x\in\mathcal{X}_{m,n}$, $s\in\mathcal{I}_{m,X}$ and $t\in\mathcal{I}_{n,Y}$. In particular, we have that

$$\Big|\frac{\partial}{\partial s}\phi(u_{x,s},v_{x,t})\Big|\lesssim (1/s) f(x)^{\kappa_1} g(x)^{\kappa_2}, \quad \text{(S81)}$$

uniformly for $x\in\mathcal{X}_{m,n}$, $s\in\mathcal{I}_{m,X}$ and $t\in\mathcal{I}_{n,Y}$. Analogous arguments also reveal that $\frac{\partial}{\partial t}\phi(u_{x,s},v_{x,t})\lesssim (1/t) f(x)^{\kappa_1} g(x)^{\kappa_2}$, uniformly for $x\in\mathcal{X}_{m,n}$, $s\in\mathcal{I}_{m,X}$ and $t\in\mathcal{I}_{n,Y}$. Moreover, since $x\in\mathcal{X}_{m,n}$ and $\|y-x\|\le\max\{r_{m,x,y}^{(1)},r_{n,x,y}^{(2)}\}$, we may argue as we did leading up to (S78) to obtain similar bounds on $\frac{\partial}{\partial s}\phi(u_{y,s},v_{y,t})$ and $\frac{\partial}{\partial t}\phi(u_{y,s},v_{y,t})$. Thus, using (S32) and (S79), we find that $U_1=o(n^{-4})$.

To bound U_2: Again using Lemma S6, we have that

$$\max\Big\{\sup_{s_1\in[0,1]}|F_{m,x,y}^{(1)}-G_m^{(1)}|(s_1,a_{m,X}^-),\ \sup_{s_2\in[0,1]}|F_{m,x,y}^{(1)}-G_m^{(1)}|(a_{m,X}^-,s_2),$$

$$|F_{m,x,y}^{(1)}-G_m^{(1)}|(a_{m,X}^+,a_{m,X}^+)\Big\}=o(m^{-4}). \quad \text{(S82)}$$

By similar arguments to those used in the bound on U_1 we have that

$$\Big|\frac{\partial^2}{\partial s\partial t}\phi(u_{x,s},v_{x,t})\Big|=\frac{k_X k_Y d^2|\phi_{11}(u_{x,s},v_{x,t})|}{mnV_d^2 h_{x,f}^{-1}(s)^{d+1} h'_{x,f}\big(h_{x,f}^{-1}(s)\big) h_{x,g}^{-1}(t)^{d+1} h'_{x,g}\big(h_{x,g}^{-1}(t)\big)}$$

$$\lesssim \{1/(st)\} f(x)^{\kappa_1} g(x)^{\kappa_2}, \quad \text{(S83)}$$

uniformly for $x\in\mathcal{X}_{m,n}$, $s\in\mathcal{I}_{m,X}$ and $t\in\mathcal{I}_{n,Y}$; moreover, the same bound also holds for $\frac{\partial^2}{\partial s\partial t}\phi(u_{y,s},v_{y,t})$. We may therefore use (S32), (S80) and (S82) to conclude that $U_2=o(m^{-4})$.

To bound U_3: By Lemma S6, we have that

$$F_{m,x,y}^{(1)}(s_1,a_{m,X}^+)-G_m^{(1)}(s_1,a_{m,X}^+)=\frac{\mathrm{B}_{k_X,m-k_X}(s_1)}{m-1}\mathbb{1}_{\{\|x-y\|\le h_{x,f}^{-1}(s_1)\}}+o(m^{-4}), \quad \text{(S84)}$$

uniformly for $x\in\mathcal{X}_{m,n}$, $\|y-x\|\le r_{m,x,y}^{(1)}$ and $s\in\mathcal{I}_{m,X}$, with an analogous statement holding for $F_{m,x,y}^{(1)}(a_{m,X}^+,s_2)-G_m^{(1)}(a_{m,X}^+,s_2)$. Now, combining this statement with our bounds on the derivatives of h in (S81) and (S83), and applying the bounds $|F^{(1)}(s_1,s_2)|\le\mathbb{1}_{\{\|y-x\|\le r_{m,x,y}^{(1)}\}}$ and $|F^{(2)}(t_1,t_2)|\le\mathbb{1}_{\{\|y-x\|\le r_{n,x,y}^{(2)}\}}$, we may write

$$|U_3|\lesssim\int_{\mathcal{X}\times\mathcal{X}_{m,n}} f(x)^{1+\kappa_1} g(x)^{\kappa_2} f(y)^{1+\kappa_1} g(y)^{\kappa_2}\mathbb{1}_{\{\|y-x\|\le\min\{r_{m,x,y}^{(1)},r_{n,x,y}^{(2)}\}\}}$$

$$\times\Big(\frac{\log m\log n}{k_X k_Y}+\frac{\log n}{k_X k_Y}+\frac{\log^{1/2} m\log n}{m^2 k_X^{1/2} k_Y}\Big)\,dx\,dy$$

$$\lesssim\frac{\log m\log n}{k_X k_Y}\int_{\mathcal{X}\times\mathcal{X}_{m,n}} f(x)^{2+2\kappa_1} g(x)^{2\kappa_2}\mathbb{1}_{\{\|y-x\|\le\min\{r_{m,x,y}^{(1)},r_{n,x,y}^{(2)}\}\}}\,dx\,dy$$

$$\lesssim\frac{\log m\log n}{k_X k_Y}\int_{\mathcal{X}_{m,n}} f(x)^{2+2\kappa_1} g(x)^{2\kappa_2}\min\Big\{\frac{k_X}{mf(x)},\frac{k_Y}{ng(x)}\Big\}\,dx.$$

Since $\min(m,n) \geq 3$, if $m \geq n$, then $(1/m)\log m \leq (1/n)\log n$ and therefore

$$|U_3| \lesssim \frac{\log m \log n}{k_X k_Y} \int_{\mathcal{X}} \frac{k_X}{m} f(x)^{1+2\kappa_1} g(x)^{2\kappa_2}\, dx \lesssim \frac{\log^2 n}{n k_Y}.$$

Similarly, if $n \geq m$ then

$$|U_3| \lesssim \frac{\log m \log n}{k_X k_Y} \int_{\mathcal{X}} \frac{k_Y}{n} f(x)^{2+2\kappa_1} g(x)^{2\kappa_2 - 1}\, dx \lesssim \frac{\log^2 m}{m k_X}.$$

Putting these two statements together,

$$U_3 = O\biggl(\max\biggl\{\frac{\log^2 m}{m k_X}, \frac{\log^2 n}{n k_Y}\biggr\}\biggr),$$

which establishes (S34).

To bound U_4: Using (S32), (S80) and (S82) we have that $U_4 = o(m^{-4})$.

To bound U_5: We first bound the contribution to U_5 from the discontinuous parts of $F_{m,x,y}^{(1)}$, arising due to the indicator functions in (S30). Recalling the definition of the multinomial random vector $(N_1^{(1)}, N_2^{(1)}, N_3^{(1)}, N_4^{(1)})$ in (S29), we have that

$$\begin{aligned}
0 &\leq F_{m,x,y}^{(1)}(s_1,s_2) - \mathbb{P}\bigl(N_1^{(1)} + N_3^{(1)} \geq k_X, N_2^{(1)} + N_3^{(1)} \geq k_X\bigr) \\
&\leq \mathbb{P}\bigl(N_1^{(1)} + N_3^{(1)} = k_X - 1\bigr) + \mathbb{P}\bigl(N_2^{(1)} + N_3^{(1)} = k_X - 1\bigr) \\
&= \binom{m-2}{k_X - 1} s^{k_X - 1}(1-s)^{m-k_X-1} + \binom{m-2}{k_X - 1} t^{k_X - 1}(1-t)^{m-k_X-1} \\
&\leq \frac{2}{(2\pi k_X)^{1/2}}\{1 + o(1)\},
\end{aligned}$$

uniformly for $x \in \mathcal{X}_{m,n}$, $\|y - x\| \leq r_{m,x,y}^{(1)}$ and $(s_1,s_2) \in \mathcal{I}_{m,X}^2$, and we will see is of no larger order than the error in the normal approximation for the continuous part. Now, writing $y = x + \{\frac{k_X}{mV_d f(x)}\}^{1/d} z$, define

$$\begin{aligned}
U_{51} := \int_{\mathcal{X}\times\mathcal{X}_{m,n}} & f(x)f(y) \int_{\mathcal{I}_{n,Y}^2} dG_n^{(2)}(t_1,t_2) \\
\times \biggl[\int_{\mathcal{I}_{m,X}^2} & h_{1100}\biggl\{F^{(1)}(s_1,s_2) - \bigl(\Phi_\Sigma - \Phi_{I_2}\bigr)\biggl(\frac{ms_1 - k_X}{k_X^{1/2}}, \frac{ms_2 - k_X}{k_X^{1/2}}\biggr)\biggr\}\, ds_1\, ds_2 \\
- \int_{\mathcal{I}_{m,X}} & h_{1000}\biggl\{F^{(1)}(s_1, a_{m,X}^+) - \frac{\mathrm{B}_{k_X, m-k_X}(s_1)}{m-1} \mathbb{1}_{\{\|z\| \leq 1\}}\biggr\}\, ds_1 \\
- \int_{\mathcal{I}_{m,X}} & h_{0100}\biggl\{F^{(1)}(a_{m,X}^+, s_2) - \frac{\mathrm{B}_{k_X, m-k_X}(s_2)}{m-1} \mathbb{1}_{\{\|z\| \leq 1\}}\biggr\}\, ds_2\biggr]\, dx\, dy.
\end{aligned}$$

By Lemma S4 we have that

$$\begin{aligned}
\int_{\mathcal{X}} \int_{\mathcal{I}_{m,X}} & \frac{1}{ms} \mathrm{B}_{k_X, m-k_X}(s) \Bigl| \mathbb{1}_{\{\|y-x\| \leq h_{x,f}^{-1}(s)\}} - \mathbb{1}_{\{\|z\| \leq 1\}} \Bigr|\, ds\, dy \\
&\lesssim \frac{1}{k_X} \int_{\mathbb{R}^d} \mathbb{1}_{\{(\frac{k_X}{mV_d f(x)})^{\frac{1}{d}} < \|y-x\| \leq h_{x,f}^{-1}(a_{m,X}^+)\}} \vee \mathbb{1}_{\{h_{x,f}^{-1}(a_{m,X}^-) < \|y-x\| \leq (\frac{k_X}{mV_d f(x)})^{\frac{1}{d}}\}}\, dy \\
&\leq \frac{V_d}{k_X}\bigl\{h_{x,f}^{-1}(a_{m,X}^+)^d - h_{x,f}^{-1}(a_{m,X}^-)^d\bigr\} \lesssim \frac{1}{mf(x)}\biggl\{\frac{\log^{\frac{1}{2}} m}{k_X^{1/2}} + \Bigl(\frac{k_X M_\beta(x)^d}{mf(x)}\Bigr)^{\frac{2\wedge\beta}{d}}\biggr\},
\end{aligned}$$

uniformly for $x \in \mathcal{X}_{m,n}$. Using this bound together with Lemma S11, (S80) and (S84) we may say that

$$|U_{51}| \lesssim \int_{\mathcal{X}_{m,n}} f(x)^{1+2\kappa_1} g(x)^{2\kappa_2} \biggl[\frac{1}{m}\biggl\{\frac{\log^{1/2} m}{k_X^{1/2}} + \biggl(\frac{k_X M_\beta(x)^d}{mf(x)}\biggr)^{(2\wedge\beta)/d}\biggr\}$$
$$+ \frac{k_X}{m}\log^2\Bigl(\frac{a_{m,X}^+}{a_{n,Y}^-}\Bigr) \int_{B_0(2)} \min\biggl\{1, \frac{1}{\|z\|}\biggl(\frac{\log^{\frac{1}{2}} m}{k_X^{1/2}} + \biggl(\frac{k_X M_\beta(x)^d}{mf(x)}\biggr)^{\frac{1\wedge\beta}{d}}\biggr)\biggr\}\,dz\biggr]\,dx$$
$$= O\biggl(\frac{1}{m}\max\biggl\{\frac{\log^{5/2} m}{k_X^{1/2}}, \log^2 m\Bigl(\frac{k_X}{m}\Bigr)^{(1\wedge\beta)/d}, \Bigl(\frac{k_X}{m}\Bigr)^{\lambda_1(1-2\zeta)-\epsilon}\biggr\}\biggr),$$

for every $\epsilon > 0$. In bounding U_5 it therefore remains to approximate the derivatives of h using (S80) and to bound the contribution from the tails of the t_1, t_2, s_1, s_2 integrals. By Lemma S6 and standard normal tail bounds the error from these tail contributions is $o(m^{-4})$, and so, using (S80),

$$|U_{52}| := |U_5 - U_{51}| \lesssim \frac{1}{m}\int_{\mathcal{X}_{m,n}} f(x)^{1+2\kappa_1} g(x)^{2\kappa_2} \biggl\{\frac{\log^{1/2} m}{k_X^{1/2}} + \frac{\log^{1/2} n}{k_Y^{1/2}}$$
$$+ \biggl(\frac{k_X M_\beta(x)^d}{mf(x)}\biggr)^{\frac{2\wedge\beta}{d}} + \biggl(\frac{k_Y M_\beta(x)^d}{ng(x)}\biggr)^{\frac{2\wedge\beta}{d}}\biggr\}\,dx$$
$$= O\biggl(\frac{1}{m}\max\biggl\{\frac{\log^{1/2} m}{k_X^{1/2}}, \frac{\log^{1/2} n}{k_Y^{1/2}}, \Bigl(\frac{k_X}{m}\Bigr)^{\frac{2\wedge\beta}{d}}, \Bigl(\frac{k_Y}{n}\Bigr)^{\frac{2\wedge\beta}{d}},$$
$$\Bigl(\frac{k_X}{m}\Bigr)^{\lambda_1(1-2\zeta)-\epsilon}, \Bigl(\frac{k_Y}{n}\Bigr)^{\lambda_2(1-2\zeta)-\epsilon}\biggr\}\biggr),$$

for every $\epsilon > 0$.

To bound U_6*:* Using Lemma S7 we have that

$$|U_6| \lesssim \frac{1}{m}\int_{\mathcal{X}_{m,n}^c} f(x)^{1+2\kappa_1} g(x)^{2\kappa_2}\,dx = O\biggl(\frac{1}{m}\max\biggl\{\Bigl(\frac{k_X}{m}\Bigr)^{\lambda_1(1-2\zeta)-\epsilon}, \Bigl(\frac{k_Y}{n}\Bigr)^{\lambda_2(1-2\zeta)-\epsilon}\biggr\}\biggr),$$

for every $\epsilon > 0$. This establishes (S36).

To bound U_7*:* Analogously to our bounds on U_1, we may use (S32), (S79) and (S80) to show that $U_7 = o(n^{-4})$.

To bound U_8*:* Using Lemma S11, (S32), (S38), (S79) and (S80), and the change of variables $y = x + (\frac{k_Y}{nV_d g(x)})^{1/d} z$, we have that

$$|U_{81}| := \biggl|\int_{\mathcal{X}\times\mathcal{X}_{m,n}} f(x)f(y)\int_{\mathcal{I}_{m,X}^2}\int_{\mathcal{I}_{n,Y}^2} h_{0011}(s_1,s_2,t_1,t_2)\,dG_m^{(1)}(s_1,s_2)$$
$$\times\Bigl\{F^{(2)}(t_1,t_2) - (\Phi_\Sigma - \Phi_{I_2})\Bigl(\frac{nt_1 - k_Y}{k_Y^{1/2}}, \frac{nt_2 - k_Y}{k_Y^{1/2}}\Bigr)\Bigr\}\,dt_1\,dt_2\,dx\,dy\biggr|$$
$$\lesssim \frac{k_Y}{n}\log^2\Bigl(\frac{a_{n,Y}^+}{a_{n,Y}^-}\Bigr)\int_{\mathcal{X}_{m,n}} f(x)^{2+2\kappa_1} g(x)^{2\kappa_2-1}$$
$$\times\int_{B_0(2)}\min\biggl\{1, \frac{1}{\|z\|}\biggl(\frac{\log^{1/2} n}{k_Y^{1/2}} + \biggl(\frac{k_Y M_\beta(x)^d}{ng(x)}\biggr)^{(1\wedge\beta)/d}\biggr)\biggr\}\,dz\,dx$$

$$\lesssim \frac{\log^2 n}{n} \int_{\mathcal{X}_{m,n}} f(x)^{2+2\kappa_1} g(x)^{2\kappa_2-1} \biggl\{ \frac{\log^{1/2} n}{k_Y^{1/2}} + \biggl(\frac{k_Y M_\beta(x)^d}{n g(x)} \biggr)^{(1\wedge\beta)/d} \biggr\} \, dx$$
$$= O\biggl(\frac{\log^2 n}{n} \max\biggl\{ \frac{\log^{1/2} n}{k_Y^{1/2}}, \Bigl(\frac{k_Y}{n}\Bigr)^{(1\wedge\beta)/d}, \Bigl(\frac{k_Y}{n}\Bigr)^{\epsilon_0} \biggr\} \biggr),$$

As with U_5 we now define $U_{82} := U_8 - U_{81}$ and note that to bound U_{82} we need to control the tails of s_1, s_2, t_1, t_2 integrals and our approximations to the derivatives of h. By (S38), (S80) and Lemma S6 we have that

$$|U_{82}| \lesssim \frac{1}{n} \int_{\mathcal{X}_{m,n}} f(x)^{2+2\kappa_1} g(x)^{2\kappa_2-1} \biggl\{ \frac{\log^{1/2} m}{k_X^{1/2}} + \frac{\log^{1/2} n}{k_Y^{1/2}} + \biggl(\frac{k_X M_\beta(x)^d}{m f(x)} \biggr)^{(2\wedge\beta)/d} + \biggl(\frac{k_Y M_\beta(x)^d}{n g(x)} \biggr)^{(2\wedge\beta)/d} + m^{-2} \biggr\} \, dx$$
$$= O\biggl(\frac{1}{n} \max\biggl\{ \frac{\log^{1/2} n}{k_Y^{1/2}}, \frac{\log^{1/2} m}{k_X^{1/2}}, \Bigl(\frac{k_X}{m}\Bigr)^{(2\wedge\beta)/d}, \Bigl(\frac{k_Y}{n}\Bigr)^{(2\wedge\beta)/d}, \Bigl(\frac{k_X}{m}\Bigr)^{\epsilon_0}, \Bigl(\frac{k_Y}{n}\Bigr)^{\epsilon_0}, m^{-2} \biggr\} \biggr),$$

This establishes (S40), and therefore concludes the proof.

REFERENCES

Baldi, P. and Rinott, Y. (1989) On normal approximations of distributions in terms of dependency graphs. *Ann. Prob.*, **17**, 1646–1650.

Berrett, T. B. and Samworth, R. J. (2023) Efficient functional estimation and the super-oracle phenomenon. *Submitted.*

Berrett, T. B., Samworth, R. J. and Yuan, M. (2019) Efficient multivariate entropy estimation via k-nearest neighbour distances. *Ann. Statist.*, **47**, 288–318.

Biau, G. and Devroye, L. (2015) *Lectures on the Nearest Neighbor Method*. Springer, New York.

Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1989) *Regular Variation*. Cambridge University Press, Cambridge.

Boucheron, S., Lugosi, G. and Massart, P. (2013) *Concentration Inequalities*. Oxford University Press, Oxford.

Folland, G. B. (1999) *Real Analysis: Modern Techniques and their Applications*. Wiley, New York.

Götze, F. (1991) On the rate of convergence in the multivariate CLT. *Ann. Prob.*, **19**, 724–739.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge.

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