Supplementary Material for Extrinsic local regression on manifold-valued data

Appendix

Proof of Theorem 4.1. Recall

$$\widehat{F}(x) = \frac{\frac{1}{n}\sum_{i=1}^{n} J(y_i)K_H(x_i - x)}{\frac{1}{n}\sum_{i=1}^{n} K_H(x_i - x)}.$$

Denote the denominator of $\widehat{F}(x)$ as

$$\widehat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K_H(x_i - x) = \frac{1}{n \mid H \mid}\sum_{i=1}^{n} K(x_i - x).$$

It is standard to show

$$\widehat{f}(x) \xrightarrow{P} f_X(x) \tag{1}$$

where $\xrightarrow{P}$ indicates convergence in probability. For the numerator term of $\widehat{F}(x)$, one has

$$\begin{aligned}
E\left(\frac{1}{n}\sum_{i=1}^{n} J(y_i)K_H(x_i - x)\right) &= \frac{1}{n}\sum_{i=1}^{n} E\left(J(y_i)K_H(x_i - x)\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} \int E\left(J(y_i)K_H(x_i - x) \mid x_i\right) f_X(x_i)dx_i \\
&= \frac{1}{n}\sum_{i=1}^{n} \int \mu(x_i)K_H(x_i - x)f_X(x_i)dx_i \\
&= \int \mu(\widetilde{x})K_H(\widetilde{x} - x)f_X(\widetilde{x})d\widetilde{x}.
\end{aligned}$$

Noting that $\mu(x) = (\mu_1(x), \ldots, \mu_D(x))' \in \mathbb{R}^D$, we slightly abuse the integral notation above meaning that the jth entry of $E\left(n^{-1}\sum_{i=1}^n J(y_i)K_H(x_i - x)\right)$ is given by

$$\int \mu_j(\widetilde{x})K_H(\widetilde{x} - x)f_X(\widetilde{x})d\widetilde{x}.$$

Letting $v = H^{-1}(\widetilde{x} - x)$ by changing of variables, the above equations become

$$E\left(\frac{1}{n}\sum_{i=1}^n J(y_i)K_H(x_i - x)\right) = \int \mu(x + Hv)K(v)f_X(x + Hv)dv.$$

By the multivariate Taylor expansion,

$$f_X(x + Hv) = f_X(x) + (\bigtriangledown f) \cdot (Hv) + R, \tag{2}$$

where $\bigtriangledown f$ is the gradient of f and R is the remainder term of the expansion. The remainder R can be shown to be bounded above by

$$R \leq \frac{C}{2}\|Hv\|^2, \ \|Hv\| = |h_1v_1| + \ldots |h_mv_m|.$$

Note that $\mu(x + Hv)$ is a multivariate map valued in $\mathbb{R}^D$. We can make second order multivariate Taylor expansions for $\mu(x + Hv) = (\mu_1(x + Hv), \ldots, \mu_D(x + Hv))'$ at each of its entries μ_i for $i = 1, \ldots, D$. We have

$$\mu(x + Hv) = \mu(x) + A(Hv) + V + R, \tag{3}$$

where A is a $D \times m$ matrix whose ith row is given by the gradient of μ_i evaluated at x. V is a D-dimensional vector, whose ith term is given by $\frac{1}{2}(Hv)^tT_i(Hv)$, where T_i is the Hessian matrix of $\mu_i(x)$ and R is the remainder vector. Thus,

$$\begin{aligned}
&E\left(\frac{1}{n}\sum_{i=1}^n J(y_i)K_H(x_i - x)\right) \\
&\approx \int \left((f_X(x) + (\bigtriangledown f) \cdot (Hv))K(v)(\mu(x) + A(Hv) + V)\right) dv \\
&= f_X(x)\mu(x) + f_X(x)\int K(v)A(Hv)dv + f_X(x)\int K(v)V dv \qquad (4) \\
&+ \mu(x)\int(\bigtriangledown f) \cdot (Hv)K(v)dv + \int(\bigtriangledown f) \cdot (Hv)K(v)A(Hv)dv + \int(\bigtriangledown f) \cdot (Hv)K(v)V dv. \qquad (5)
\end{aligned}$$

By the property of the kernel function, we have $\int K(u)udu = 0$; therefore the second term of equation (4) is zero by simple algebra. To evaluate the third term of equation (4), we first calculate for $\int K(v)V dv$. From here onward until the end of the proof, we denote $x = (x^1, \ldots, x^m)$ where x^i is the ith coordinate of x. Note that the ith term of V $(i = 1, \ldots, D)$ is given by $\frac{1}{2}(Hv)^t T_i (Hv)$, where T_i is the Hessian matrix of μ_i, which is precisely

$$\frac{1}{2}h_1^2 v_1^2 \left(\frac{\partial^2 \mu_i}{\partial (x^1)^2} + \ldots + \frac{\partial^2 \mu_i}{\partial x^m x^1}\right) + \ldots + \frac{1}{2}h_m^2 v_m^2 \left(\frac{\partial^2 \mu_i}{\partial x^1 x^m} + \ldots + \frac{\partial^2 \mu_i}{\partial (x^m)^2}\right).$$

Therefore, the ith entry of the third term of equation (4) is given by

$$\begin{aligned} U_i = \frac{1}{2} f_X(x) \Big(& h_1^2 \left(\frac{\partial^2 \mu_i}{\partial (x^1)^2} + \ldots + \frac{\partial^2 \mu_i}{\partial x^m x^1}\right) \int v_1^2 K_1(v_1) dv_1 + \ldots \\ & + h_m^2 \left(\frac{\partial^2 \mu_i}{\partial x^1 x^m} + \ldots + \frac{\partial^2 \mu_i}{\partial (x^m)^2}\right) \int v_m^2 K_m(v_m) dv_m \Big). \end{aligned} \tag{6}$$

The first term of equation (5) is given by

$$\mu(x) \int (\bigtriangledown f) \cdot (Hv) K(v) dv = \int \left(h_1 v_1 \frac{\partial f}{\partial x^1} + \ldots + h_m v_m \frac{\partial f}{\partial x^m} \right) K(v) dv = 0.$$

The ith entry of the second term of equation (5) is given by

$$h_1^2 \frac{\partial f}{\partial x^1} \frac{\partial \mu_i}{\partial x^1} \int v_1^2 K_1(v_1) dv_1 + \ldots + h_m^2 \frac{\partial f}{\partial x^m} \frac{\partial \mu_i}{\partial x^m} \int v_m^2 K_m(v_m) dv_m. \tag{7}$$

The third term of equation (5) can be shown to be zero, since odd moments of symmetric kernels are 0. Therefore, we have

$$E\left(\frac{1}{n}\sum_{i=1}^n J(y_i) K_H(x_i - x)\right) \approx f_X(x)\mu(x) + Z, \tag{8}$$

where the ith coordinate of Z is

$$\begin{aligned} Z_i = & h_1^2 \left\{ \frac{\partial f}{\partial x^1} \frac{\partial \mu_i}{\partial x^1} + \frac{1}{2} f_X(x) \left(\frac{\partial^2 \mu_i}{\partial (x^1)^2} + \ldots + \frac{\partial^2 \mu_i}{\partial x^m x^1}\right) \right\} \int v_1^2 K_1(v_1) dv_1 \\ & + \ldots \\ & + h_m^2 \left\{ \frac{\partial f}{\partial x^m} \frac{\partial \mu_i}{\partial x^m} + \frac{1}{2} f_X(x) \left(\frac{\partial^2 \mu_i}{\partial x^1 x^m} + \ldots + \frac{\partial^2 \mu_i}{\partial (x^m)^2}\right) \right\} \int v_m^2 K_m(v_m) dv_m \end{aligned} \tag{9}$$

combining equations (6) and (7). The reminder term of (2) is of order $o(\max\{h_1, \ldots, h_m\})$ and each entry of the remainder vector in (3) is of order $o(\max\{h_1^2, \ldots, h_m^2\})$.

We now look at the covariance matrix of $n^{-1}\sum_{i=1}^n J(y_i) K_H(x_i - x)$, which we denote by $\Sigma(x)$. Denote the jth entry $(j = 1, \ldots, D)$ of $J(y_i)$ as $J_j(y_i)$. Denote $\sigma(y^j, y^k)$ as the conditional covariance between the ith entry and jth entry of y. We have

$$\Sigma_{jk} = E\Big[\left(\frac{1}{n}\sum_{i=1}^{n} J_j(y_i)K_H(x_i-x) - E\left(\frac{1}{n}\sum_{i=1}^{n} J_j(y_i)K_H(x_i-x)\right)\right)$$
$$\left(\frac{1}{n}\sum_{i=1}^{n} J_k(y_i)K_H(x_i-x) - E\left(\frac{1}{n}\sum_{i=1}^{n} J_k(y_i)K_H(x_i-x)\right)\right)\Big]$$
$$= E\Big[\left(\frac{1}{n}\sum_{i=1}^{n}\left(J_j(y_i)K_H(x_i-x) - \int \mu_j(\widetilde{x})K_H(\widetilde{x}-x)f_X(\widetilde{x})d\widetilde{x}\right)\right)$$
$$\left(\frac{1}{n}\sum_{i=1}^{n}\left(J_k(y_i)K_H(x_i-x) - \int \mu_k(\widetilde{x})K_H(\widetilde{x}-x)f_X(\widetilde{x})d\widetilde{x}\right)\right)\Big]$$
$$= \frac{1}{n}\int E\Big[\left(J_j(y_1)K_H(x_1-x) - \int \mu_j(\widetilde{x})K_H(\widetilde{x}-x)f_X(\widetilde{x})d\widetilde{x}\right)$$
$$\left(J_k(y_1)K_H(x_1-x) - \int \mu_k(\widetilde{x})K_H(\widetilde{x}-x)f_X(\widetilde{x})d\widetilde{x}\right) \mid x_1\Big] f_X(x_1)dx_1$$
$$= \frac{1}{n}\int \sigma(J_j(y_1)K_H(x_1-x), J_k(y_1)K_H(x_1-x)f_X(x_1)dx_1$$
$$= \frac{1}{n}\int K_H(x_1-x)^2\sigma(J_j(y_1), J_k(y_1))f_X(x_1)dx_1.$$

By the change of variable $v = H^{-1}(x_1 - x)$, the above equation becomes

$$\Sigma_{jk} = \frac{1}{n|H|}\int K(v)^2\sigma(J_j(y_v), J_k(y_v))f_X(Hv+x)dv$$
$$= \frac{1}{n|H|}\int K(v)^2\sigma(J_j(y_v), J_k(y_v))\left(f_X(x) + \bigtriangledown f\cdot(Hv) + o(\max\{h_1,\dots,h_m\})\right)dv$$
$$= \frac{1}{n|H|}\int K(v)^2\sigma(J_j(y_v), J_k(y_v))f_X(x))dv + o\left(\frac{1}{n|H|}\right). \tag{10}$$

By (1), (8) and (23), and applying central limit theorem and Slustky's theorem, one has

$$\sqrt{n|H|}\left(\widehat{F}(x) - \widetilde{\mu}(x)\right) \xrightarrow{L} N(0, \bar{\Sigma}(x)), \tag{11}$$

where $\widetilde{\mu}(x) = \mu(x) + \frac{Z}{f_X(x)}$ and the ith entry $(i = 1,\dots,D)$ of Z is given by (9) and

$$\bar{\Sigma}_{jk} = \frac{\sigma(J_j(y_v), J_k(y_v))\int K(v)^2dv}{f_X(x)}. \tag{12}$$

One can show

$$\sqrt{n|H|}\left(\widehat{F}_E(x)-\mathcal{P}\left(\widetilde{\mu}(x)\right)\right)=\sqrt{n|H|}d_{\widetilde{\mu}(x)}\mathcal{P}\left(\widehat{F}(x)-\widetilde{\mu}(x)\right)+o_P(1).$$

Therefore, one has

$$\sqrt{n|H|}d_{\widetilde{\mu}(x)}\mathcal{P}\left(\widehat{F}(x)-\widetilde{\mu}(x)\right)\xrightarrow{L}N(0,\widetilde{\Sigma}(x)). \tag{13}$$

Here $\widetilde{\Sigma}(x)=B^T\bar{\Sigma}(x)B$, where B is the $D\times d$ matrix of the differential $d_{\widetilde{\mu}(x)}\mathcal{P}$ with respect to given orthonormal bases of $T_{\widetilde{\mu}(x)}\mathbb{R}^D$ and $T_{\mathcal{P}\widetilde{\mu}(x)}\widetilde{M}$.

□

Proof of Corollary 4.2. In choosing the optimal order of bandwidth, one can consider choosing $(h_1,\dots,h_m)$ such that the mean integrated squared error is minimized. Note that

$$\widehat{F}_E(x)-F(x)=\text{Jacob}(\mathcal{P})_{\mu(x)}\left(\widehat{F}(x)-\mu(x)\right)+o_p(1). \tag{14}$$

Here Jacob($\mathcal{P}$) is the Jacobian matrix of the projection map $\mathcal{P}$. One has

$$\begin{aligned}
MISE(\widehat{F}_E(x)) &= \int E\|\widehat{F}_E(x)-F(x)\|^2dx\\
&= \int E\|\text{Jacob}(\mathcal{P})_{\mu(x)}\left(\widehat{F}(x)-\mu(x)\right)+o_p(1)\|^2dx\\
&= \int E\left(\sum_{i=1}^{D}\left(\sum_{j=1}^{D}\mathcal{P}_{ij}\left(\widehat{F}_j(x)-\mu_j(x)\right)\right)^2+o_p(1)\right)dx\\
&= O(1/n|H|)+\ldots+O(1/n|H|)+O(h_1^4)+\ldots+O(h_m^4).
\end{aligned}$$

The last terms follow from Fatou's lemma, and that the Jacobian map is differentiable at $\mu(x)$ for every x. Therefore, if h_i's $(i=1,\dots,m)$ are taken to be of the same order, that is, of $O(n^{-1/(m+4)})$, then one can obtain MISE($\widehat{F}_E(x)$) with an order of $O(n^{-4/(m+4)})$. □

Proof of Theorem 4.3. Let B be the $D\times d$ matrix of the differential $d_{\widetilde{\mu}(x)}\mathcal{P}$ with respect to given orthonormal basis of tangent space $T_{\widetilde{\mu}(x)}\mathbb{R}^D$ and tangent space $T_{\mathcal{P}\widetilde{\mu}(x)}\widetilde{M}$. Given a canonical choice of basis for tangent space $T_{\widetilde{\mu}(x)}\mathbb{R}^D$, one has the representation for

$$\sup_x\|d_{\widetilde{\mu}(x)}\mathcal{P}\left(\widehat{F}(x)-E(\widehat{F}(x))\right)\|=\sup_x\sqrt{\sum_{i=1}^{d}\left(\sum_{j=1}^{D}B_{ij}^T\left(\widehat{F}_j(x)-E(\widehat{F}_j(x))\right)\right)^2}. \tag{15}$$

Note that the projection map is differentiable around the neighborhood of $\mu(x)$ and $\mathcal{X}$ is compact, so $B_{ij}^T(x)$ are bounded. Let $C_{ij}=\sup_{x\in\mathcal{X}}(B_{ij}^T)^2(x)$ and $C=\max C_{ij}$. For each term note that, by Cauchy-Schwarz inequality,

$$\sup_{x\in\mathcal{X}}\left(\sum_{j=1}^{D}\left(B_{ij}^{T}\left(\widehat{F}_j(x)-E\left(\widehat{F}_j(x)\right)\right)\right)\right)^2 \le \sup_{x}\sum_{j=1}^{D}(B_{ij}^{T})^2\left(\widehat{F}_j(x)-E\left(\widehat{F}_j(x)\right)\right)^2 \tag{16}$$

$$\le C\sum_{j=1}^{D}\sup_{x\in\mathcal{X}}\left(\widehat{F}_j(x)-E\left(\widehat{F}_j(x)\right)\right)^2. \tag{17}$$

By Theorem 2 in Hansen (2008), one can see that

$$\sup_{x\in\mathcal{X}}|\left(\widehat{F}_j(x)-E\left(\widehat{F}_j(x)\right)\right)| = O(r_n), \tag{18}$$

where $r_n = \log^{1/2} n/\sqrt{n|H|}$. Then one has

$$\sup_{x\in\mathcal{X}}\sum_{i=1}^{d}\left(\sum_{j=1}^{D}\left(B_{ij}\left(\widehat{F}_j(x)-E\left(\widehat{F}_j(x)\right)\right)\right)\right)^2 = O(r_n^2). \tag{19}$$

Then one has

$$\sup_{x}\|d_{\widetilde{\mu}(x)}\mathcal{P}\left(\widehat{F}(x)-E(\widehat{F}(x))\right)\| = \sup_{x\in\mathcal{X}}\sqrt{\sum_{i=1}^{d}\left(\sum_{j=1}^{D}\left(B_{ij}\left(\widehat{F}_j(x)-E\left(\widehat{F}_j(x)\right)\right)\right)\right)^2}$$
$$= O(r_n) = O\left(\log^{1/2} n/\sqrt{n|H|}\right).$$

□

Proof of Theorem 4.4. Given the higher order smoothness assumption on $\mu(x)$, one can make higher order approximations and using a local polynomial regression estimate would result in the reduction of bias term in estimating $\mu(x)$. The asymptotic distribution for multivariate local regression estimator for Euclidean responses has been derived (Gu et al., 2014; Ruppert and Wand, 1994; Masry, 1996), and we leverage on some of their results in our proof.

Note that $\widehat{F}(x) = \left(\widehat{F}_1(x),\ldots,\widehat{F}_D(x)\right) \in \mathbb{R}^D$, $E(\widehat{F}(x)) = \left(E(\widehat{F}_1(x)),\ldots,E(\widehat{F}_D(x))\right)^T$ and the expectation taken in each component is with respect to the marginal distribution of $\widetilde{P}(dy|x)$. Then by Theorem 1 of Gu et al. (2014), the following holds:

(1) If p is odd, then for $j = 1,\ldots,D$

$$\begin{aligned}\text{Bias}_j(\widehat{F}(x)) &= E(\widehat{F}_j(x)) - \mu_j(x)\\ &= \left(\mathcal{M}_p^{-1}\mathcal{B}_{p+1}\boldsymbol{H}^{(p+1)}\boldsymbol{m}^{j}_{\boldsymbol{p+1}}(x)\right)_1,\end{aligned} \tag{20}$$

which is of order $O(\|\boldsymbol{h}\|^{p+1})$. Here $(\cdot)_1$ represents the first entry of the vector inside the parenthesis;

(2) If p is even, then for $j = 1, \ldots, D$

$$\begin{aligned}\text{Bias}_j(\widehat{F}(x)) &= E(\widehat{F}_j(x)) - \mu_j(x) \qquad (21)\\ &= \left(\sum_{l=1}^{m} h_l \frac{f_l(x)}{f_X(x)} \left(\mathcal{M}_p^{-1}\mathcal{B}_{p+1}^l - \mathcal{M}_p^{-1}\mathcal{M}_p^l\mathcal{M}_p^{-1}\mathcal{B}_{p+1} \right) \boldsymbol{H}^{(p+1)} \boldsymbol{m}_{\boldsymbol{p+1}}^j(x) + \mathcal{M}_p^{-1}\mathcal{B}_{p+2}\boldsymbol{H}^{(p+2)}\boldsymbol{m}_{\boldsymbol{p+2}}^j(x) \right)_1,\end{aligned}$$

which is of order $O(\|\boldsymbol{h}\|^{p+2})$.

For any $k \in \{0, 1, \ldots, p\}$. Let $N_k = \binom{k+m-1}{m-1}$ and $\mathcal{N}_p = \sum_{k=0}^{p} N_k$. Here $\mathcal{M}_p$ is a $\mathcal{N}_p \times \mathcal{N}_p$ matrix whose (i,j)th block $(0 \le i, j \le p)$ is given by $\int_{\mathbb{R}^m} \boldsymbol{u}^{i+j} K(\boldsymbol{u}) d\boldsymbol{u}$ and $\mathcal{M}_p^l$ $(l = 1, \ldots, m)$ is a $\mathcal{N}_p \times \mathcal{N}_p$ matrix whose (i,j)th block $(0 \le i, j \le p)$ is given by $\int_{\mathbb{R}^m} u_l \boldsymbol{u}^{i+j} K(\boldsymbol{u}) d\boldsymbol{u}$. $\mathcal{B}_{p+1}$ is a $\mathcal{N}_p \times N_{p+1}$ matrix whose $(i, p+1)$th $(i = 1, \ldots, p)$ block is given by $\int_{\mathbb{R}^m} \boldsymbol{u}^{i+p+1} K(\boldsymbol{u}) d\boldsymbol{u}$ and $\mathcal{B}_{p+1}^l$ $(l = 1, \ldots, m)$ is a $\mathcal{N}_p \times N_{p+1}$ matrix whose $(i, p+1)$th $(i = 1, \ldots, p)$ block is given by $\int_{\mathbb{R}^m} u_l \boldsymbol{u}^{i+p+1} K(\boldsymbol{u}) d\boldsymbol{u}$. We have $\boldsymbol{H}^{(p+1)} = \text{Diag}\{h_1^{p+1}, \ldots, h_m^{p+1}\}$, $f_l(x) = \dfrac{\partial f_X(x)}{\partial x^l}$ and $\boldsymbol{m}_{\boldsymbol{p+1}}^j(x)$ $(j = 1, \ldots, D)$ is the vector of all the $p+1$ order partial derivatives of $\mu_j(x)$, that is, $m_{\boldsymbol{p+1}}^j(x) = \left(\dfrac{\partial \mu_j^{p+1}(x)}{\partial (x^1)^{p+1}}, \dfrac{\partial \mu_j^{p+1}(x)}{\partial (x^1)^p \partial (x^2)}, \ldots, \dfrac{\partial \mu_j^{p+1}(x)}{\partial (x^m)^{p+1}} \right)$.

With $\text{Bias}_j(\widehat{F}(x))$ $(j = 1, \ldots, D)$ given above, one has

$$\text{Bias}(x) = E(\widehat{F}(x)) - \mu(x) = \left(\text{Bias}_1(\widehat{F}(x)), \ldots, \text{Bias}_D(\widehat{F}(x)) \right)^T. \qquad (22)$$

Although higher order polynomial regression results in the reduction in the order of bias with the higher order smoothness assumptions on $\mu(x)$, the order and expression of the covariance remains the same. That is,

$$\begin{aligned}\Sigma_{jk} &= \text{Cov}(\widehat{F}_j(x), \widehat{F}_k(x))\\ &= 1/(n|H|) f_X(x)^{-1} \int K(v)^2 \sigma(J_j(y_v), J_k(y_v)) dv + o\left(1/(n|H|)\right), \qquad (23)\end{aligned}$$

where $\sigma(J_j(y_v), J_k(y_v)$ is the covariance between $J_j(y_v)$ and $J_k(y_v)$.

Applying the central limit theorem, one has

$$\sqrt{n|H|} \left(\widehat{F}(x) - \mu(x) - \text{Bias}(x) \right) \xrightarrow{L} N(0, \bar{\Sigma}(x)) \qquad (24)$$

where the jth $(j = 1, \ldots, D)$ entry of $\text{Bias}(x)$ is given in (20) or (21) depending on whether p is odd or even, and

$$\bar{\Sigma}_{jk} = \frac{\sigma(J_j(y_v), J_k(y_v)) \int K(v)^2 dv}{f_X(x)}. \qquad (25)$$

Letting $\widetilde{\mu}(x) = \mu(x) + \mathrm{Bias}(x)$, one has

$$\sqrt{n|H|}\left(\widehat{F}_E(x) - \mathcal{P}\left(\widetilde{\mu}(x)\right)\right) = \sqrt{n|H|}d_{\widetilde{\mu}(x)}\mathcal{P}\left(\widehat{F}(x) - \widetilde{\mu}(x)\right) + o_P(1).$$

Therefore, by applying Slutsky's theorem, one has

$$\sqrt{n|H|}d_{\widetilde{\mu}(x)}\mathcal{P}\left(\widehat{F}(x) - \widetilde{\mu}(x)\right) \xrightarrow{L} N(0, \widetilde{\Sigma}(x)). \tag{26}$$

Here $\widetilde{\Sigma}(x) = B^T\bar{\Sigma}(x)B$ where B is the $D \times d$ matrix of the differential $d_{\widetilde{\mu}(x)}P$ with respect to given orthonormal bases of the tangent space $T_{\widetilde{\mu}(x)}\mathbb{R}^D$ and tangent space $T_{\widetilde{\mu}(x)}\widetilde{M}$.

□

References

Gu, J., Li, Q., and Yang, J.-C. (2014). Multivariate local polynomial kernel estimators: leading bias and asymptotic distribution. *Econometric Reviews*. to appear.

Hansen, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory*, 24:726–748.

Masry, E. (1996). Multivariate local polynomial regression for time series:uniform strong consistency and rates. *Journal of Time Series Analysis*, 17(6):571–599.

Ruppert, D. and Wand, M. P. (1994). Multivariate locally weighted least squares regression. *The Annals of Statistics*, 22(3):1346–1370.