Local Polynomial and Penalized Trigonometric Series Regression

Li-Shan Huang and Kung-Sik Chan

National Tsing Hua University and University of Iowa

Supplementary Material

Appendix

CONDITIONS (A).

(A1). The design density $f(x)$ is bounded away from 0 and ∞, and $f(x)$ has a continuous fourth derivative on a compact support which, with no loss of generality, is taken to be $[0, 1]$.

(A2). The kernel $K(\cdot)$ is a Lipschitz continuous, bounded and symmetric probability density function, having a support on a compact interval, say $[-1, 1]$.

(A3). The error ε is from a symmetric distribution with mean 0, variance σ^2, and finite fourth moment.

(A4). The $(2p + 3)$-th derivative of $m(\cdot)$ exists.

(A5). The bandwidth $h = h_n$ is a non-random sequence of positive numbers satisfying $h \to 0$ and $nh \to \infty$ as $n \to \infty$.

Lemma: Under Conditions (A1)-(A4), H_p^* is a shrinking matrix, i.e. $||H_p^*\mathbf{y}|| \leq ||\mathbf{y}||$.

Proof: From Huang and Chen (2008),

$$\mathbf{y}^T H_p^* \mathbf{y} = n^{-1} \int \sum_{i=1}^{n} \left(\sum_{j=0}^{p} \hat{\beta}_j(x)(X_i - x)^j \right)^2 K_h\left(X_i - x\right) dx.$$

Therefore H_p^* is positive-definite except in degenerating cases. Being both symmetric and positive-definite, the eigenvalues of H_p^* are > 0. From (2.3) and (2.5), $(I - H_p^*)$ is nonnegative-definite. Then for any n-vector v with a Euclidean norm 1, $v^T(I - H_p^*)v \geq 0$ implies $v^T H_p^* v \leq 1$. Therefore H_p^* is a shrinking matrix.

Proof of property c:

Based on (2.6), for $i = 1, \ldots, n$ and $l = 1, \ldots, n$, the (i,l)-th element in H_p^* is

$$\int \sum_{j=1}^{(p+1)} \sum_{k=1}^{(p+1)} K_h(X_i - x) K_h(X_l - x)(X_i - x)^{(j-1)}(X_l - x)^{(k-1)} s_{j,k}(x) dx, \qquad \text{(S0.1)}$$

where $s_{j,k}(x)$ denotes the asymptotic expression for (j,k)-th element of $(X_p^T W X_p)^{-1}$. From Fan and Gijbels (1996), $s_{j,k}(x)$ involves $f^{-1}(x)$ that relates to x. With X_i's equally spaced, (S0.1) is a function of $(X_i - X_l)$ and hence H_p^* is asymptotically a Toeplitz matrix. The matrix is banded because the summand in (S0.1) equals 0 if $|X_i - X_l| > h$. This completes the proof.

Proofs of Theorem 1 and 2:

The proofs for $p = 2$ and $p = 3$ in Theorem 1(a) are similar to the case of $p = 1$ in Huang and Davidson (2010) but with more tedious derivations. In short, it involves the following steps: (1) deriving an asymptotic form of $(X_p^T W X_p)^{-1}$ as in Section 3.7 in Fan and Gijbels (1996); (2) deriving the equivalent kernels for $\hat\beta_0, \ldots, \hat\beta_p$ (some are already given in Table 3.1 in Fan and Gijbels (1996)); and (3) the weighted integration in (2.7) implies the convolution operations between the equivalent kernel of $\hat\beta_j$ in (2) and $u^j K(u)$, $j = 0, \ldots, p$. Then the moments of the equivalent kernels are calculated to show the order of $(0, 2(p+1))$ in Theorem 1(b) and the order of the asymptotic variance in Theorem 1(c) is proved based on the definition of the equivalent kernels.

For Theorem 2(a)(i), it is similar to Theorem 1(a) but the convolution operations are with $u^j L(u)$, $j = 0, \ldots, p$. To show Theorem 2(a)(ii) for $t = (1 + c)h$, the weighted integration in (2.7) is split into two parts, $[h, (2+c)h]$ and $[ch, h]$. The latter part involves $\hat\beta_j(x)$ with $x \in [0, h)$ and their equivalent kernels $K_{j,d}^e$ are different from the first part. Similarly for Theorem 2(a)(iii), the integration in (2.7) is split into two parts, $[h, (1+c)h]$ and $[0, h]$, yielding the two-part expression in (3.6).

We next show Theorem 2(b)(i) for $p = 1$ while the cases of $p = 2$ and 3 are analogous. First note that

$$\begin{aligned} & \int \left(\beta_0(x) + \hat\beta_1(x)(t - x)\right) L_g(t - x) dx - \int m(t) L_g(t - x) dx \\ = & \int \left((\hat\beta_0(x) - \beta_0(x)) + (\hat\beta_1(x) - \beta_1(x))(t - x)\right) L_g(t - x) dx \\ & - \int (\beta_2(x)(t - x)^2 + r(x,t)) L_g(t - x) dx, \end{aligned} \qquad \text{(S0.2)}$$

where $r(x,t)$ denotes the remainder terms after a Taylor expansion. Then the bias of (3.3) with $p = 1$ is the expectation of (S0.2) and by using the bias expressions of $\hat\beta_j(x)$, $j = 0, 1$ from Fan and Gijbels (1996), (3.7) is obtained.

Proof of Theorem 3:

By straightforward linear algebra arguments, it is easy to see that H_p^* and ${H_p^*}^{-1}$ share the same eigenvectors. From Lemma 1, the eigenvalues of H_p^* are positive and less than or equal to 1. Denote the eigenvectors of H_p^* as $\mathbf{z}_1, \ldots, \mathbf{z}_{k_1}, \ldots, \mathbf{z}_{k_1+k_2}$, where $\mathbf{z}_1, \ldots, \mathbf{z}_{k_1}$ correspond to the same eigenvalue 1, and $\mathbf{z}_{k_1+1}, \ldots, \mathbf{z}_{k_1+k_2}$ correspond to non-zero eigenvalues $1 > \lambda_{k_1+1} \geq \ldots \geq \lambda_{k_1+k_2} > 0$. Without loss of generality, assume that $\{\mathbf{z}_{k_1+1}, \ldots, \mathbf{z}_{k_1+k_2}\}$ are orthonormal and that the space spanned by $\{\mathbf{z}_1, \ldots, \mathbf{z}_{k_1}\}$ is of dimension k_1. Let $\mathbf{X} = (\mathbf{z}_1, \ldots, \mathbf{z}_{k_1})$ and $\mathbf{Z} = (\mathbf{z}_{k_1+1}, \ldots, \mathbf{z}_{k_1+k_2})$. Then $H_p^* \mathbf{y}$ can be expressed as $\mathbf{X}\hat{\mathbf{b}} + \mathbf{Z}\hat{\mathbf{u}}$ for some coefficient vectors $\hat{\mathbf{b}}$ and $\hat{\mathbf{u}}$. Since $\mathbf{X}^T(H_p^{*^{-1}} - I)\mathbf{X} = 0$ and $\mathbf{Z}^T(H_p^{*^{-1}} - I)\mathbf{X} = 0$, the penalty term in (4.2) is $(\mathbf{X}\hat{\mathbf{b}} + \mathbf{Z}\hat{\mathbf{u}})^T(H_p^{*^{-1}} - I)(\mathbf{X}\hat{\mathbf{b}} + \mathbf{Z}\hat{\mathbf{u}}) = \hat{\mathbf{u}}^T D\hat{\mathbf{u}}$ with D a diagonal matrix with entries $(1/\lambda_k - 1)$, $k = k_1 + 1, \ldots, k_1 + k_2$. Hence the connection between local polynomial regression and mixed models is established.

Proof of Theorem 4:

We provide the proof in the case of local linear regression in detail. The other cases $p = 0, 2, 3$ can be established analogously by using their respective equivalent kernels. Note that H_p^* is a symmetric matrix and therefore its eigenvectors corresponding to distinct eigenvalues must be orthogonal. Moreover $\cos(2k\pi\mathbf{x})$ and $\sin(2k\pi\mathbf{x})$ form an orthogonal basis for functions defined on $[0, 1]$.

Let us first explore a special case of a finite k (and so $kh \to 0$). From (2.8), the i-th element of $H_1^*(\cos 2k\pi\mathbf{x})$ is

$$\begin{aligned}
&\frac{1}{nf(X_i)} \sum_j h^{-1} \left(K_0^* \left(\frac{X_j - X_i}{h} \right) - \mu_2^{-1} K_1^* \left(\frac{X_j - X_i}{h} \right) \right) \cos(2k\pi X_j) \\
\approx &\frac{1}{f(X_i)} h^{-1} \int \left(K_0^* \left(\frac{x - X_i}{h} \right) - \mu_2^{-1} K_1^* \left(\frac{x - X_i}{h} \right) \right) \cos(2k\pi x) f(x) dx \\
= &\frac{1}{f(X_i)} \int \left(K_0^*(u) - \mu_2^{-1} K_1^*(u) \right) \cos(2k\pi(X_i + hu)) f(X_i + hu) du. \qquad \text{(S0.3)}
\end{aligned}$$

Then for a finite k,

$$\cos(2k\pi(X_i + hu)) = \cos(2k\pi X_i)\cos(2k\pi hu) - \sin(2k\pi X_i)\sin(2k\pi hu),$$

with $\cos(2k\pi hu) = 1 - (2k\pi hu)^2/2! + \ldots$ and $\sin(2k\pi hu) = 2k\pi hu - (2k\pi hu)^3/3! + \ldots$. Since $(K_0^*(u) - \mu_2^{-1}K_1^*(u))$ is a kernel of order $(0, 4)$, $H_1^* \cos(2k\pi\mathbf{x}) = \cos(2k\pi\mathbf{x})(1 + O(h^4))$ after expanding both $\cos(2k\pi(X_i + hu))$ and $f(X_i + hu)$ in (S0.3). For $H_1^* \sin(2k\pi\mathbf{x})$, the arguments are analogous except that

$$\sin(2k\pi(X_i + hu)) = \sin(2k\pi X_i)\cos(2k\pi hu) - \cos(2k\pi X_i)\sin(2k\pi hu).$$

Then $H_1^* \sin(2k\pi\mathbf{x}) = \sin(2k\pi\mathbf{x})(1 + O(h^4))$.

The above discussion applies for finite k. For those k such that $k \to \infty$ and $kh \to 0$, the above arguments continue to hold but with the term $O(h^4)$ replaced by $O(k^4h^4)$. Hence $\cos(2k\pi\mathbf{x})$ and $\sin(2k\pi\mathbf{x})$ are asymptotic eigenvectors of H_1^* when k is finite and when $k \to \infty$ and $kh \to 0$ as stated in Theorem 4(a).

We now consider the cases in (c): when $kh \to \infty$ and $kh \to c$, where c is a constant. When k is of larger order than h so that $kh \to \infty$, $H_1^* \cos(2k\pi\mathbf{x}) \to 0$ and $H_1^* \sin(2k\pi\mathbf{x}) \to$

0 by the Riemann-Lebesgue Lemma. When kh converges to a constant,

$$\cos(2k\pi(X_i + hu)) = \cos(2k\pi X_i)\cos(2c\pi u) - \sin(2k\pi X_i)\sin(2c\pi u).$$

Since $\cos(2c\pi u)$ and $\sin(2c\pi u)$ are finite functions, by the Riemann-Lebesgue Lemma, again $H_1^* \cos(2k\pi\mathbf{x}) \to 0$ and $H^* \sin(2k\pi\mathbf{x}) \to 0$.

To show Theorem 4(b), in the case of $\cos(2k\pi\mathbf{x})$, it can be seen that the corresponding eigenvalue is $1 + O(k^4h^4)$, and the second order term $O(k^4h^4)$ is negative since the 4-th moment of $(K_0^*(u) - \mu_2^{-1}K_1^*(u))$ is $6(\mu_2^2 - \mu_4) < 0$. Hence the larger the k, the smaller the corresponding eigenvalue λ_k and the larger the penalty weight $(1/\lambda_k - 1)$. For $\sin(2k\pi\mathbf{x})$, analogous arguments applies.

For $p = 0, 2, 3$, the arguments are similar except that the equivalent kernel in (S0.3) is replaced by the corresponding equivalent kernels stated in Theorem 1 and the second order terms $O(h^4)$ replaced by $O(h^{(2p+2)})$ and $O(k^4h^4)$ replaced by $O((kh)^{(2p+2)})$. For $p = 0$, the second moment of $K_0^*(\cdot)$ is positive but the h^2-term in the geometric expansion of $cos(2k\pi hu)$ is negative. Hence the claim that "the larger the k, the more the penalty weight" still holds. Similar arguments work for $p = 2$. For $p = 3$, the 8-th moment of the equivalent kernel is not always negative unless $(\mu_8 - \mu_4^2)(\mu_4 - \mu_2^2) > (\mu_6 - \mu_4\mu_2)^2$. It is negative for the Gaussian and Epanechnikov kennels but not for the Uniform kernel.