Non-stationary variance estimation and kriging prediction
Proof of Theorems

Shu Yang and Zhengyuan Zhu

Department of Statistics, Iowa State University

Supplementary Material

S1 Proof of Theorem 1

To prove Theorem 1, we need the following lemma to simplify the calculation.

Lemma1: If $(X, Y) \sim N\left(\left(\begin{array}{c} 0 \\ 0 \end{array}\right), \left(\begin{array}{cc} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{array}\right)\right)$, then $Cov[X^2, Y^2] = 2\big(Cov[X, Y]\big)^2$.

Proof: Since $X|Y \sim N(\frac{\rho\sigma_1}{\sigma_2}Y, (1-\rho^2))$, $E[X^2|Y] = (1-\rho^2)\sigma_1^2 + (\frac{\rho\sigma_1}{\sigma_2}Y)^2$

$$\begin{aligned}
Cov[X^2, Y^2] &= E[X^2Y^2] - E[X^2]E[Y^2] \\
&= E\{E[X^2Y^2|Y]\} - \sigma_1^2\sigma_2^2 \\
&= E[Y^2\{(1-\rho^2)\sigma_1^2 + (\frac{\rho\sigma_1}{\sigma_2}Y)^2\}] - \sigma_1^2\sigma_2^2 \\
&= (1-\rho^2)\sigma_1^2\sigma_2^2 + (\frac{\rho\sigma_1}{\sigma_2})^2 3\sigma_2^4 - \sigma_1^2\sigma_2^2 \\
&= 2\rho^2\sigma_1^2\sigma_2^2 \\
&= 2\big(Cov[X, Y]\big)^2.
\end{aligned}$$

To establish the asymptotic results for the local polynomial regression estimator, which will be in Section 3, we rely on the following analytical properties of $K_n(\frac{x-x_i}{\lambda})$,

1. $\sum_{i=1}^{n-1} K_n(\frac{x-x_i}{\lambda}) = 1$,

2. $\sum_{i=1}^{n-1}(x-x_i)^j K_n(\frac{x-x_i}{\lambda}) = 0$, for any $j = 1, \ldots, p$,
3. $K_n(\cdot) = 0$ for all $|x - x_i| > \lambda$,
4. $K_n(\frac{x-x_i}{\lambda}) = O((n\lambda)^{-1})$ uniformly for all $x \in [0, 1]$,
5. $\sum_{i=1}^{n-1} \left(K_n(\frac{x-x_i}{\lambda})\right)^2 = O((n\lambda)^{-1})$.

Property 5 can be derived from property 4 and the Cauchy-Schwartz inequality.

Proof of Theorem 1, recall $\Delta_i = h^{-1}(D_{h,i}^2 - 2\sigma_\epsilon^2)$, where $D_{h,i} = Z(x_i) - Z(x_i + h)$. So

$$E[D_{h,i}^2] = \sigma_i^2 h + \{\sigma_i^{(1)2} x_i + \sigma_i^{2(1)}\} h^2 + 2\sigma_\epsilon^2 + o(h^2),$$

Thus

$$\begin{aligned} E[\Delta_i] &= E[h^{-1}(D_{h,i}^2 - 2\sigma_\epsilon^2)] \\ &= h^{-1}[\sigma_i^2 h + \{\sigma_i^{(1)2} x_i + \sigma_i^{2(1)}\} h^2 + o(h^2)] \\ &= \sigma_i^2 + \{\sigma_i^{(1)2} x_i + \sigma_i^{2(1)}\} h + o(h). \end{aligned}$$

The bias of $\hat\sigma_\lambda^2(x;\sigma_\epsilon^2) = \sum_{i=1}^{n-1} K_n(\frac{x-x_i}{\lambda})\Delta_i$ can be calculated as

$$\begin{aligned} E[\hat\sigma_\lambda^2(x;\sigma_\epsilon^2) - \sigma^2(x)] &= \sum_{i=1}^{n-1} K_n(\frac{x-x_i}{\lambda})\{E[\Delta_i] - \sigma^2(x)\} \\ &= \sum_{i=1}^{n-1} K_n(\frac{x-x_i}{\lambda})\{\sigma_i^2 - \sigma^2(x) + (\sigma_i^{(1)2} x_i + \sigma_i^{2(1)})h + O(h^2)\} \\ &= \sum_{i=1}^{n-1} K_n(\frac{x-x_i}{\lambda})\{\sum_{j=1}^{\lfloor\beta\rfloor} \frac{\sigma^{2(j)}(x)}{j!}(x_i - x)^j + O(|x_i - x|^\beta) \\ &\quad +(\sigma_i^{(1)2} x_i + \sigma_i^{2(1)})h + O(h^2)\} \\ &= \sum_{i=1}^{n-1} K_n(\frac{x-x_i}{\lambda})\{O(|x_i - x|^\beta) + (\sigma_i^{(1)2} x_i + \sigma_i^{2(1)})h + O(h^2)\}, \end{aligned}$$

where the third equality is obtained by Taylor expansion of $\sigma_i^2 = \sigma^2(x_i)$ at x and the assumption that $\sigma^2 \in C_\beta^+$, and the last equality follows by Property 2 of $K_n(\frac{x-x_i}{\lambda})$. Note that

$$\begin{aligned} \Big|\sum_{i=1}^{n-1} K_n(\frac{x-x_i}{\lambda})|x_i - x|^\beta\Big| &\le \sum_{i=1}^{n-1} |K_n(\frac{x-x_i}{\lambda})||x_i - x|^\beta \\ &\le \sum_{i;|x-x_i|<\lambda} |K_n(\frac{x-x_i}{\lambda})|\lambda^\beta \\ &= O(\lambda^\beta) \end{aligned}$$

where the second inequality comes from Property 3 of $K_n(\frac{x-x_i}{\lambda})$. So Bias term is $O(\max(h, \lambda^\beta))$.

To simplify the notation, let $\sigma_{i,h} = \sigma(x_i + h)$, $\sigma_{i,2h} = \sigma(x_i + 2h)$,

$$\begin{aligned} Var[D_{h,i}^2] &= Var[(Z(x_i+h) - Z(x_i))^2] \\ &= 2\{\sigma_i^2 x_i + \sigma_{i,h}^2 (x_i + h) - 2\sigma_i \sigma_{i,h} x_i + 2\sigma_\epsilon^2\}^2 \\ &= 2\{\sigma_i^2 h + (\sigma_i^{(1)2} x_i + \sigma_i^{2(1)}) h^2 + O(h^3) + 2\sigma_\epsilon^2\}^2, \end{aligned}$$

where the second equality follows from Lemma 1. For $j = i + 1$,

$$\begin{aligned} Cov[\Delta_i, \Delta_j] &= h^{-2} Cov[D_{h,i}^2, D_{h,j}^2] \\ &= 2h^{-2} [Cov\{D_{h,i}, D_{h,j}\}]^2 \\ &= 2h^{-2} \sigma_\epsilon^4, \end{aligned}$$

For $j \geq i + 2$, $Cov[\Delta_i, \Delta_j] = 0$. Thus

$$\begin{aligned} Var[\hat{\sigma}_\lambda^2(x; \sigma_\epsilon^2)] &= \sum_{i=1}^{n-1} \sum_{j=1}^{n-1} K_n(\frac{x - x_i}{\lambda}) K_n(\frac{x - x_j}{\lambda}) Cov[\Delta_i, \Delta_j] \\ &= \sum_{i=1}^{n-1} K_n(\frac{x - x_i}{\lambda})^2 Var[\Delta_i] \\ &\quad + 2 \sum_{i=1}^{n-1} \sum_{j>i}^{n-1} K_n(\frac{x - x_i}{\lambda}) K_n(\frac{x - x_j}{\lambda}) Cov[\Delta_i, \Delta_j] \\ &= \sum_{i=1}^{n-1} K_n(\frac{x - x_i}{\lambda})^2 h^{-2} 2\{\sigma_i^2 h + (\sigma_i^{(1)2} x_i + \sigma_i^{2(1)}) h^2 + O(h^3) + 2\sigma_\epsilon^2\}^2 \\ &\quad + 2 \sum_{i=1}^{n-2} K_n(\frac{x - x_i}{\lambda}) K_n(\frac{x - x_{i+1}}{\lambda}) 2h^{-2} \sigma_\epsilon^4 \\ &= O\left((n\lambda)^{-1} \cdot \max\{1, n^{2-2\alpha}\}\right), \end{aligned}$$

where the last equality is obtained by Property 5 of $K_n(\frac{x-x_i}{\lambda})$. If $\alpha \geq 1$, the bias term is $O(\max(h, \lambda^\beta))$ and the variance term is $O((n\lambda)^{-1})$, the optimal bandwidth is $\lambda = O(n^{-1/(1+2\beta)})$, under which the mean squared error is $O(n^{-\beta/(1+2\beta)})$.

If $\frac{1}{2} < \alpha < 1$, the bias term is $O(\max(h, \lambda^\beta))$ and the variance term is $O((n\lambda)^{-1} n^{2-2\alpha})$, the optimal bandwidth is $\lambda = O(n^{-(2\alpha-1)/(1+2\beta)})$, under which the mean squared error is $O(n^{-2(1+\beta-\alpha)/(1+2\beta)})$. Theorem 1 follows.

S2 Proof of Theorem 2

Proof: Let $a_{ni} = K_n(\frac{x-x_i}{\lambda})$ and $\xi_i = \Delta_i$. Check the following conditions as in Theorem 2.2 in Peligrad and Utev (1997).

1. $\max_{1 \leq i \leq n} |a_{ni}| \to 0$ as $n \to \infty$ and this condition holds since

$$K_n(\frac{x - x_i}{\lambda}) = O((n\lambda)^{-1}),$$

2. $\sup_n \sum_{i=1}^n a_{ni}^2 < \infty$ and this condition holds since
$$\sum_{i=1}^{n-1} \left(K_n(\frac{x-x_i}{\lambda})\right)^2 = O((n\lambda)^{-1}),$$

3. For a certain $\delta > 0$, $\{|\xi_i|^{2+\delta}\}$ is uniformly integrable and this condition can be easily verified by the fact that
$$\Delta_i = h^{-1}(D_{h,i}^2 - 2\sigma_\epsilon^2)$$
and
$$D_{h,i} = Z_i - Z_{i+1} \sim N(0, u_i),$$
where $u_i = \sigma_i^2 h + (\sigma_i^{(1)2} x_i + \sigma_i^{2(1)})h^2 + O(h^3) + 2\sigma_\epsilon^2$. So $E\{D_{h,i}^6\} = 15u_i^3$, then it is easy to check that
$$\sup_{1\le i\le n} E\{|\xi_i|^3\} < \infty,$$
which guarantees that $\{|\xi_i|^{2+\delta}\}$ is uniformly integrable.

The CLT in Theorem 2 follows.

S3 Proof of Theorem 3

To prove Theorem 3, we can write
$$\begin{aligned}\hat{U}(\sigma_\epsilon^2) &= \frac{\partial}{\partial \sigma_\epsilon^2}(-\frac{1}{2}\log|\hat{\Sigma}| - \frac{1}{2}\mathbf{d}^T\hat{\Sigma}^{-1}\mathbf{d}) \\ &= -\frac{1}{2}tr(\hat{\Sigma}^{-1}B) + \frac{1}{2}\mathbf{d}^T\hat{\Sigma}^{-1}B\hat{\Sigma}^{-1}\mathbf{d}, \qquad \text{(S3.1)}\end{aligned}$$
where $B = \begin{pmatrix} 0 & -1 & \cdots & 0 & 0 \\ -1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & -1 \\ 0 & 0 & \cdots & -1 & 0 \end{pmatrix}$, and
$$\hat{\Sigma} = \begin{pmatrix} \hat{D}_{h,\lambda}(x_1) & -\sigma_\epsilon^2 & \cdots & 0 & 0 \\ -\sigma_\epsilon^2 & \hat{D}_{h,\lambda}(x_2) & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \hat{D}_{h,\lambda}(x_{n-2}) & -\sigma_\epsilon^2 \\ 0 & 0 & \cdots & -\sigma_\epsilon^2 & \hat{D}_{h,\lambda}(x_{n-1}) \end{pmatrix}.$$

Since $\hat{\sigma}_\epsilon^2$ satisfies $\hat{U}(\hat{\sigma}_\epsilon^2) = 0$, by the mean value theorem, we have
$$\begin{aligned} 0 &= \hat{U}(\hat{\sigma}_\epsilon^2) \\ &= \hat{U}(\sigma_{\epsilon,0}^2) + \dot{U}(\sigma_\epsilon^{2*})(\hat{\sigma}_\epsilon^2 - \sigma_{\epsilon,0}^2),\end{aligned}$$

where σ_ϵ^{2*} is the value between $\sigma_{\epsilon,0}^2$ and $\hat{\sigma}_\epsilon^2$, and

$$\begin{aligned}\dot{U}(\sigma_\epsilon^2) &= \partial\hat{U}(\sigma_\epsilon^2)/\partial\sigma_\epsilon^2 \\ &= \frac{1}{2}tr\{(\hat{\Sigma}^{-1}B)^2\} - \mathbf{d}^T\hat{\Sigma}^{-1}B\hat{\Sigma}^{-1}B\hat{\Sigma}^{-1}\mathbf{d}. \qquad (S3.2)\end{aligned}$$

So

$$\hat{\sigma}_\epsilon^2 - \sigma_{\epsilon,0}^2 = -\{\dot{U}(\sigma_\epsilon^{2*})\}^{-1}\hat{U}(\sigma_{\epsilon,0}^2). \qquad (S3.3)$$

Follow the same argument in proof of Theorem 1, we can establish

$$\hat{D}_{h,\lambda}(x) = D_{h,\lambda}(x) + O_p\left(n^{-1-\beta/(1+2\beta)}\right).$$

So we have

$$\hat{\Sigma} = \Sigma + O_p\left(n^{-1-\beta/(1+2\beta)}\right)D,$$

where D is some constant diagonal matrix. Using the property of matrix inverse, if ϵ is a small number then

$$(V+\epsilon F)^{-1} = V^{-1} - \epsilon V^{-1}FV^{-1} + O(\epsilon^2).$$

We have

$$\hat{\Sigma}^{-1} = \Sigma^{-1} + O_p(n^{-1-\beta/(1+2\beta)})\Sigma^{-1}D\Sigma^{-1}.$$

Replacing $\hat{\Sigma}^{-1}$ by the above expression in terms of Σ^{-1} in (S3.1) and (S3.2), we obtain

$$\begin{aligned}\hat{U}(\sigma_\epsilon^2) &= -\frac{1}{2}tr\left(\{\Sigma^{-1} + O_p(n^{-1-\beta/(1+2\beta)})\Sigma^{-1}D\Sigma^{-1}\}B\right) \\ &+ \frac{1}{2}\mathbf{d}^T\{\Sigma^{-1} + O_p(n^{-1-\beta/(1+2\beta)})\Sigma^{-1}D\Sigma^{-1}\}B\{\Sigma^{-1} + O_p(n^{-1-\beta/(1+2\beta)})\Sigma^{-1}D\Sigma^{-1}\}\mathbf{d},\end{aligned}$$

and

$$\begin{aligned}\dot{U}(\sigma_\epsilon^2) &= \frac{1}{2}tr\left([\Sigma^{-1} + O_p(n^{-1-\beta/(1+2\beta)})\Sigma^{-1}D\Sigma^{-1}\}B]^2\right) \\ &- \mathbf{d}^T\{\Sigma^{-1} + O_p(n^{-1-\beta/(1+2\beta)})\Sigma^{-1}D\Sigma^{-1}\}B\{\Sigma^{-1} + O_p(n^{-1-\beta/(1+2\beta)})\Sigma^{-1}D\Sigma^{-1}\} \\ &\cdot B\{\Sigma^{-1} + O_p(n^{-1-\beta/(1+2\beta)})\Sigma^{-1}D\Sigma^{-1}\}\mathbf{d}.\end{aligned}$$

Now we will discuss the order of (S3.3) case by case.

1. $\alpha > 1$: $\Sigma^{-1} = O(n)$ and $O_p(n^{-1-\beta/(1+2\beta)})\Sigma^{-1}D\Sigma^{-1} = O_p(n^{1-\beta/(1+2\beta)})$, thus $\hat{U}(\sigma_\epsilon^2) \cong -\frac{1}{2}tr\left(\Sigma^{-1}B\right) + \frac{1}{2}\mathbf{d}^T\Sigma^{-1}B\Sigma^{-1}\mathbf{d}$, and $\dot{U}(\sigma_\epsilon^2) \cong \frac{1}{2}tr\{(\Sigma^{-1}B)^2\} - \mathbf{d}^T\Sigma^{-1}B\Sigma^{-1}B\Sigma^{-1}\mathbf{d}$. So

$$\begin{aligned}\hat{\sigma}_\epsilon^2 - \sigma_{\epsilon,0}^2 &\cong -E\left\{\dot{U}(\sigma_\epsilon^2)\right\}^{-1}\{-\frac{1}{2}tr\left(\Sigma^{-1}B\right) + \frac{1}{2}\mathbf{d}^T\Sigma^{-1}B\Sigma^{-1}\mathbf{d}\} \\ &= \left\{\frac{1}{2}tr\{(\Sigma^{-1}B)^2\}\right\}^{-1}\{-\frac{1}{2}tr\left(\Sigma^{-1}B\right) + \frac{1}{2}\mathbf{d}^T\Sigma^{-1}B\Sigma^{-1}\mathbf{d}\}.\end{aligned}$$

We claim that

$$[tr\{(\Sigma^{-1}B)^2\}]^{-1}\{tr(\Sigma^{-1}B) - \mathbf{d}^T\Sigma^{-1}B\Sigma^{-1}\mathbf{d}\} = O_p(n^{-3/2})$$

in probability as $n \to \infty$. Since

$$E\left([tr\{(\Sigma^{-1}B)^2\}]^{-1}\{tr(\Sigma^{-1}B) - \mathbf{d}^T\Sigma^{-1}B\Sigma^{-1}\mathbf{d}\}\right) = 0$$

and

$$\begin{aligned}
&Var\left([tr\{(\Sigma^{-1}B)^2\}]^{-1}\{tr(\Sigma^{-1}B) - \mathbf{d}^T\Sigma^{-1}B\Sigma^{-1}\mathbf{d}\}\right)\\
=&[tr\{(\Sigma^{-1}B)^2\}]^{-2}Var\left(\mathbf{d}\Sigma^{-1}B\Sigma^{-1}\mathbf{d}\right)\\
=&[tr\{(\Sigma^{-1}B)^2\}]^{-2}tr\left\{(\Sigma^{-1}B)^2\right\}\\
=&[tr\{(\Sigma^{-1}B)^2\}]^{-1}\\
=&O(n^{-3})
\end{aligned}$$

2. $\alpha < 1$: $\Sigma^{-1} = O(n^{\alpha})$ and $O_p(n^{-1-\beta/(1+2\beta)})\Sigma^{-1}D\Sigma^{-1} = O_p(n^{2\alpha-1-\beta/(1+2\beta)})$. The inequality $2\alpha - 1 - \beta/(1+2\beta) < \alpha$ always holds in case of $\alpha < 1$, regardless of $\beta > 0$. So

$$\begin{aligned}
\hat{\sigma}_\epsilon^2 - \sigma_{\epsilon,0}^2 &\cong -E\left\{\dot{U}(\sigma_\epsilon^2)\right\}^{-1}\{-\frac{1}{2}tr\left(\Sigma^{-1}B\right) + \frac{1}{2}\mathbf{d}^T\Sigma^{-1}B\Sigma^{-1}\mathbf{d}\}\\
&= \left\{\frac{1}{2}tr\{(\Sigma^{-1}B)^2\}\right\}^{-1}\{-\frac{1}{2}tr\left(\Sigma^{-1}B\right) + \frac{1}{2}\mathbf{d}^T\Sigma^{-1}B\Sigma^{-1}\mathbf{d}\}.
\end{aligned}$$

We claim that

$$[tr\{(\Sigma^{-1}B)^2\}]^{-1}\{tr(\Sigma^{-1}B) - \mathbf{d}^T\Sigma^{-1}B\Sigma^{-1}\mathbf{d}\} = O_p(n^{-(1+2\alpha)/2})$$

in probability as $n \to \infty$. Since

$$E\left([tr\{(\Sigma^{-1}B)^2\}]^{-1}\{tr(\Sigma^{-1}B) - \mathbf{d}^T\Sigma^{-1}B\Sigma^{-1}\mathbf{d}\}\right) = 0$$

and

$$\begin{aligned}
&Var\left([tr\{(\Sigma^{-1}B)^2\}]^{-1}\{tr(\Sigma^{-1}B) - \mathbf{d}^T\Sigma^{-1}B\Sigma^{-1}\mathbf{d}\}\right)\\
=&[tr\{(\Sigma^{-1}B)^2\}]^{-2}Var\left(\mathbf{d}\Sigma^{-1}B\Sigma^{-1}\mathbf{d}\right)\\
=&[tr\{(\Sigma^{-1}B)^2\}]^{-2}tr\left\{(\Sigma^{-1}B)^2\right\}\\
=&[tr\{(\Sigma^{-1}B)^2\}]^{-1}\\
=&O(n^{-(1+2\alpha)}).
\end{aligned}$$

Thus

$$\hat{\sigma}_\epsilon^2 - \sigma_{\epsilon,0}^2 = O_p(n^{-(1+2\alpha)/2}).$$

S4 Proof of Theorem 4

Recall

$$p(x_0) \quad = \quad C(\underset{\sim}{\sigma})^T\{V(\underset{\sim}{\sigma})\}^{-1}\mathbf{z},$$

where $\underset{\sim}{\sigma} = (\sigma_1, \ldots, \sigma_n, \sigma_0, \sigma_\epsilon)$, $\sigma_0 = \sigma(x_0)$,

$$C(\underset{\sim}{\sigma}) = \begin{pmatrix} \sigma_1\sigma_0 \min(x_1, x_0) \\ \sigma_2\sigma_0 \min(x_2, x_0) \\ \vdots \\ \sigma_n\sigma_0 \min(x_n, x_0) \end{pmatrix}, \ \mathbf{z} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix},$$

and

$$V(\underset{\sim}{\sigma}) = \begin{pmatrix} \sigma_1^2 x_1 + \sigma_\epsilon^2 & \sigma_1\sigma_2 x_1 & \cdots & \sigma_1\sigma_n x_1 \\ \sigma_1\sigma_2 x_1 & \sigma_2^2 x_2 + \sigma_\epsilon^2 & \cdots & \sigma_2\sigma_n x_2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_1\sigma_n x_1 & \sigma_2\sigma_n x_2 & \cdots & \sigma_n^2 x_n + \sigma_\epsilon^2 \end{pmatrix}.$$

The plug-in kriging predictor is given by

$$\hat{p}(x_0) \quad = \quad C(\underset{\sim}{\hat{\sigma}})^T\{V(\underset{\sim}{\hat{\sigma}})\}^{-1}\mathbf{z},$$

where $\underset{\sim}{\hat{\sigma}} = (\hat{\sigma}_1, \ldots, \hat{\sigma}_n, \hat{\sigma}_0, \hat{\sigma}_\epsilon)$. Using Taylor expansion technique, we have

$$\begin{aligned} C(\underset{\sim}{\hat{\sigma}}) \quad &\cong \quad C(\underset{\sim}{\sigma}) + \sigma_0 \begin{pmatrix} \min(x_1, x_0)(\hat{\sigma}_1 - \sigma_1) \\ \min(x_2, x_0)(\hat{\sigma}_2 - \sigma_2) \\ \vdots \\ \min(x_n, x_0)(\hat{\sigma}_n - \sigma_n) \end{pmatrix} + \begin{pmatrix} \sigma_1 \min(x_1, x_0) \\ \sigma_2 \min(x_2, x_0) \\ \vdots \\ \sigma_n \min(x_n, x_0) \end{pmatrix} (\hat{\sigma}_0 - \sigma_0) \\ &= \quad C(\underset{\sim}{\sigma}) + O(n^{-p})E, \end{aligned}$$

Similarly, we can have

$$V(\underset{\sim}{\hat{\sigma}}) \quad \cong \quad V(\underset{\sim}{\sigma}) + O(n^{-q})F,$$

where p is the convergent rate of $\hat{\sigma}$, in the case when $\alpha \geqslant 1$, $p = \beta/(1+2\beta)$, and in the case when $1/2 < \alpha < 1$, $p = (2\alpha - 1)/(1+2\beta)$, q is the minimum convergent rate of $\hat{\sigma}$ and $\hat{\sigma}_\epsilon$, i.e. $q = \min\{p, 2\alpha\} = p$ considering $2\alpha \geqslant p$ always holds, E and F are constant matrices. Using the property of matrix inverse, if ϵ is a small number then

$$(V + \epsilon F)^{-1} = V^{-1} - \epsilon V^{-1} F V^{-1} + O(\epsilon^2),$$

We have

$$(V(\underset{\sim}{\hat{\sigma}}))^{-1} = V(\underset{\sim}{\sigma})^{-1} + O(n^{-q}) V(\underset{\sim}{\sigma})^{-1} F V(\underset{\sim}{\sigma})^{-1}.$$

So

$$\begin{aligned}\hat{p}(x_0) &= \{C(\underset{\sim}{\sigma}) + O(n^{-p})E\}^{-T}\{V(\underset{\sim}{\sigma}) + O(n^{-q})F\}^{-1}\mathbf{z} \\ &= C(\underset{\sim}{\sigma})^T V(\underset{\sim}{\sigma})^{-1}\mathbf{z} + O(\max(n^{-p}, n^{-q})) \\ &= p(x_0) + O(n^{-p}).\end{aligned}$$

Furthermore, we know the simply kriging predictor $p(x_0)$ is consistent to $\sigma(x_0)W(x_0)$ with convergence rate $O(n^{-1/2})$. Combine the above two facts, we have

$$\hat{p}(x_0) = \sigma(x_0)W(x_0) + O(\max\{n^{-p}, n^{-1/2}\}).$$

In our consideration, $p < 1/2$ always holds. So the convergence rate is $O(n^{-p})$. Theorem 4 follows.

S5 Figures from the Simulation Study

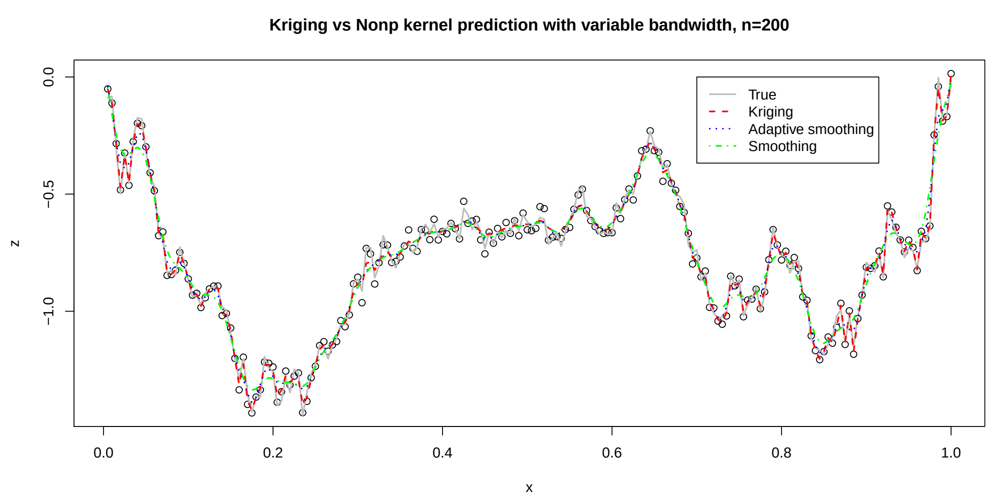

Figure 1: Plot of prediction and smoothing with $\sigma_\epsilon^2 = 0.1/n$; grey= true process $\sigma(x)W(x)$, red=kriging, blue=ALPRE, and green=LPRE.

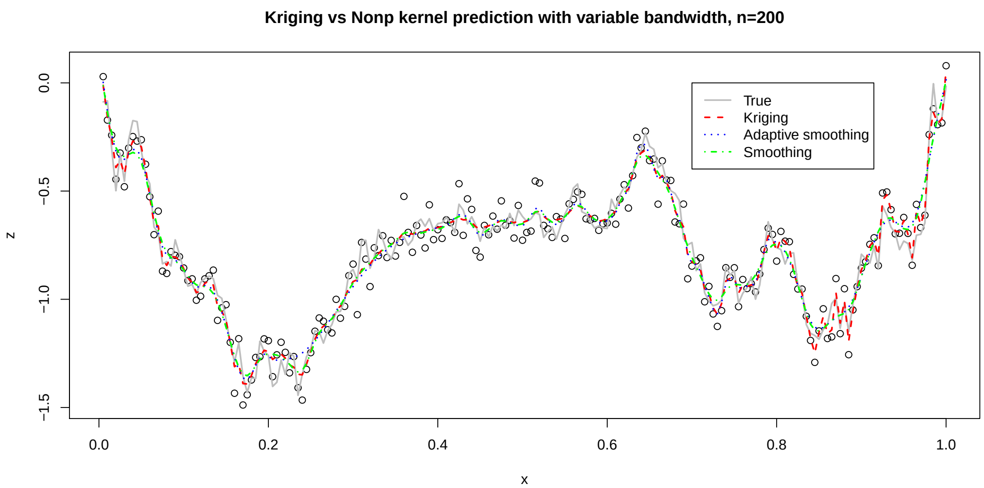

Figure 2: Plot of prediction and smoothing with $\sigma_\epsilon^2 = 1/n$, grey= true process $\sigma(x)W(x)$, red=kriging, blue=ALPRE, and green=LPRE.

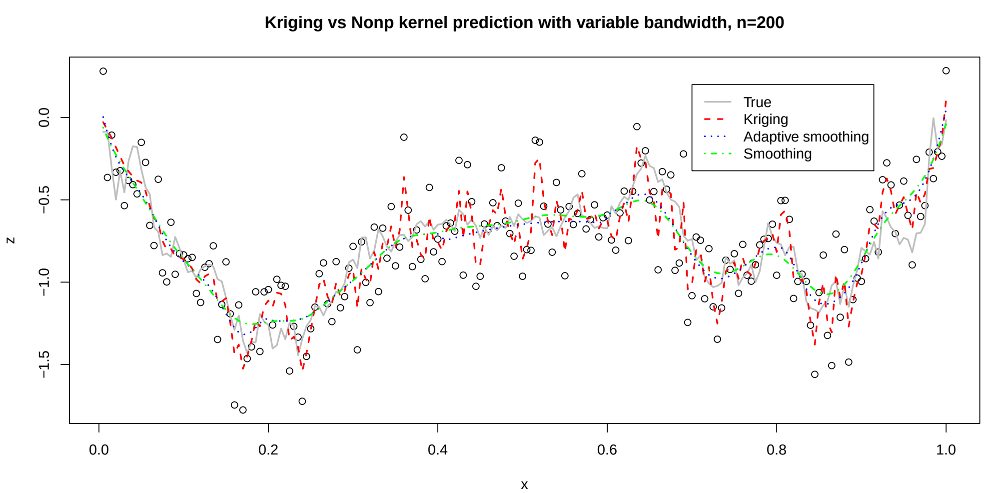

Figure 3: Plot of prediction and smoothing with $\sigma_\epsilon^2 = 10/n$, grey= true process $\sigma(x)W(x)$, red=kriging, blue=ALPRE, and green=LPRE.