科技部補助專題研究計畫成果報告
期末報告

相關誤差下的迴歸函數適應性估計

計 畫 類 別 ： 個別型計畫
計 畫 編 號 ： NSC 102-2118-M-004-007-
執 行 期 間 ： 102 年 08 月 01 日至 103 年 07 月 31 日
執 行 單 位 ： 國立政治大學統計學系

計 畫 主 持 人 ： 黃子銘

計畫參與人員： 碩士級-專任助理人員：鄭宇翔

處 理 方 式 ：
1.公開資訊：本計畫可公開查詢
2.「本研究」是否已有嚴重損及公共利益之發現：否
3.「本報告」是否建議提供政府單位施政參考：否

中 華 民 國 103 年 07 月 28 日

中 文 摘 要 ： 無母數迴歸估計問題中，若是迴歸函數平滑程度已知， 則容易得到收斂速度為最佳的估計量。 當迴歸函數平滑程度為未知時， 我們也希望迴歸函數估計量能達到一樣好的估計效果。 這樣的估計量稱為具有適應性。

當迴歸模型中誤差為相關時， 有人提出基於一種選模規則的迴歸函數估計量， 是具適應性的。 然而這方法使用上有困難， 因為選模規則牽涉到一些未知參數。 本計畫主要成果為提出一種迴歸函數估計量使用上不牽涉到一些未知參數，並證明其收斂速度達到幾乎最佳，可視為接近適應性的估計量。

中文關鍵詞： 迴歸, 適應性估計, 相關誤差

英 文 摘 要 ： In nonparametric regression, if the degree of smoothness of the regression function is known, it is often easy to obtain estimators that attain the optimal convergence rate. When the degree of smoothness of the regression function is unknown, it is desirable to have estimators for the regression function that can also achieve the optimal convergence rate. Estimators that have this property are called adaptive.

When the errors in a regression model are dependent, an adaptive estimator based on a model selection criterion has been proposed, but it is difficult to implement because the criterion involves unknown parameters for the error dependence structure and the error variance. In this project, it is proposed to consider an estimator that is also based on model selection approach without invovling unknown parameters. It is shown that the proposed estimator attends a nearly-optimal convergence rate.

study the possibility of replacing the unknown parameters by their consistent estimators to make the adaptive estimator implementable.

英文關鍵詞： regression, adaptive estimation, dependent error

An adaptive knot selection method for regression splines via penalized minimum contrast estimation

Tzee-Ming Huang

Abstract

In this report, a knot selection method for regression splines is proposed. This method yields a least square spline estimator that adapts to the smoothness of the regression function, and the knots are allowed to be not equally spaced. If the true regression function s belongs to a Sobolev space $W_m^2[0,1]$, then for a sequence $\{a_n\}$ such that $\lim_{n\to\infty} a_n = \infty$ and a constant $\gamma_0 > 0$, the proposed estimator (depending on a_n and γ_0) can converge to s at the rate $O(\sqrt{a_n(\log n)^{1+\gamma_0} n^{-2m/(1+2m)}})$ in terms of L^2 norm in probability.

1 Introduction

One of the most popular methods in non-parametric regression is B-spline estimation. B-splines are piecewise polynomials joined smoothly at points called knots. For implementation, one has to choose the number of knots and the degrees of polynomials. The choice of knots is especially crucial. For functions that are m times continuously differentiable with the m-th derivatives bounded by a constant, Zhou, Shen and Wolfe [9] showed that the number of knots should grow at the rate $n^{1/(1+2m)}$ for the spline estimator for the regression function to achieve the optimal convergence rate $n^{-2m/(1+2m)}$ in integrated mean squared error.

It is possible to construct estimators for regression functions in the Sobolev space $W_m^2[0,1]$ that can achieve the rate $n^{-m/(1+2m)}$ with respect to the L^2 norm without knowing m. These estimators are known as adaptive estimators. Barron, Birgé and Massart [2] derived risk bounds for penalized minimum contrast estimators, which can be used to construct adaptive estimators for regression functions. Huang [5] applied an inequality in Yang and Barron [8], which is obtained using a similar approach in [2], to construct an adaptive estimator for regression function using B-splines with equally spaced knots assuming the errors are normally distributed.

The objective of this study is to construct an adaptive estimator for regression function using B-splines without requiring that the knots are equally spaced or that the errors are normally distributed. This objective is achieved by first establishing an exponential inequality to control the error of minimum contrast estimators and then applying the result to derive convergence rate for a spline estimator obtained via model selection. The exponential inequality is given in Section 2. The application to adaptive B-spline estimation is given in Section 3.

2 Error Control for Minimum Contrast Estimators

In this section, the problem of minimum contrast estimation is introduced, which includes least square regression as a special case. Then, a theorem that gives error bounds for penalized minimum contrast estimators is presented, which will be used for establishing the rate of convergence of the proposed regression estimator.

The problem set-up for minimum contrast estimation is this: Consider the problem of estimating an unknown function s based on observations $Z_1, \ldots, Z_n$, where there exists a function γ such that $E\left(\frac{1}{n}\sum_{i=1}^{n}\gamma(Z_i,t)\right)$ is minimized when $t = s$. Suppose that s is estimated by

$$\tilde{s} = \arg\min_{t \in S} \frac{1}{n}\sum_{i=1}^{n}\gamma(Z_i,t).$$

Then $\tilde{s}$ is called the minimum contrast estimator over S with respect to γ, and the function γ is called the contrast function.

Least square regression fits into the framework of minimum contrast estimation. Consider the regression model

$$Y_i = s(X_i) + W_i, \qquad i = 1, \ldots, n, \tag{1}$$

where the regression function s is defined on an interval I_0, W_i are errors of mean zero that are independent of the X_i's. If s is estimated by some function in a parametric family S, then the least square estimator $\tilde{s}$ is a minimum contrast estimator with $Z_i = (X_i, Y_i)$ with respect to the contrast function $\gamma(z,t) = (y - t(x))^2$ for $z = (x, y)$.

In minimum contrast estimation, a key requirement is that the function s can be approximated well by some function in S. An approach to achieve the approximation requirement is to take $S = \cup_{j\in\Lambda} S_j$, where S_j's are various families of functions and the collection $\{S_j\}$ is rich enough to include some family that is close to s. This approach is known as the method of sieve (cf:

Barron?). To prevent overfitting, a penalty term $\eta_{1,j}$ is often considered in minimum contrast estimation to obtain the following estimator

$$\hat{s} = \arg\min_{j\in\Lambda}\left(\eta_{1,j} + \arg\min_{t\in S_j}\frac{1}{n}\sum_{i=1}^{n}\gamma(Z_i,t)\right). \tag{2}$$

To control the estimation error of the estimator penalized minimum contrast estimator $\hat{s}$, several assumptions are made. Let

$$\nu_n(\cdot) = \frac{1}{n}\sum_{i=1}^{n}[\gamma(Z_i,\cdot) - E\gamma(Z_i,\cdot)],$$

and let $\|\cdot\|$ and $\|\cdot\|_\infty$ denote the L^2 norm and sup norm on I_0. The assumptions are given below.

Assumption M1. Suppose that for $i = 1, \ldots, n$, $Z_i = f_0(s, X_i, W_i)$, where f_0 is known. Suppose that there exists a non-negative constant k_0 and for each $j \in \Lambda$, there exist non-negative constants A_j, B_j, $A_{2,j}$, $B_{2,j}$, non-negative functions M_j, Δ_j, $M_{2,j}$, $\Delta_{2,j}$, such that for $u \in S_j$,

$$|\gamma(z,u) - \gamma(z,v)| \le \begin{cases} M_j(w)\Delta_j(x,u,v), & \text{if } v \in S_j; \\ M_{2,j}(w)\Delta_{2,j}(x,u,s), & \text{if } v = s, \end{cases}$$

and, for all $m \ge 2$, $i = 1, \ldots, n$,

$$E_s[M_j^m(W_i)] \le a_m A_j^m,$$

$$E_s[\Delta_j^m(X_i,u,v)] \le b_m k_0 \|u-v\|^2 B_j^{m-2},$$

$$E_s[M_{2,j}^m(W_i)] \le a_m A_{2,j}^m,$$

$$E_s[\Delta_{2,j}^m(X_i,u,s)] \le b_m k_0 \|u-s\|^2 B_{2,j}^{m-2},$$

where $a_m = 1$, $b_m = m!/2$, for all $m \ge 2$, or $b_m = 1$, $a_m = m!/2$, for all $m \ge 2$.

Assumption M2. For each $j \in \Lambda$, there exist constants $B_j' \ge 1$, $r_j > 0$ and $D_j \ge 1$ such that, for $\sigma > 0$ and $0 < \delta < \sigma/5$, for any ball $\mathcal{B} \subset S_j$ of radius σ with respect to $\|\cdot\|$, one can find T: a subset of $\mathcal{B}$ such that $|T| \le (B_j'\sigma/\delta)^{D_j}$ and for every $u \in \mathcal{B}$, there exists $v \in T$ such that

$$\|u - v\| \le \delta$$

and

$$\|\Delta_j(\cdot,u,v)\|_\infty \le r_j\delta.$$

Assumption M3. There exist positive constants k_1 and k_2 such that

$$k_1\|u - s\|^2 \le E[\gamma(Z_i, u) - \gamma(Z_i, s)] \le k_2\|u - s\|^2 \tag{3}$$

for all $u \in S_j$ for all $j \in \Lambda$.

Under Assumptions M1 – M3, if (X_1, W_1), …, (X_n, W_n) are β-mixing, one can derive an error bound for $\hat{s}$ using bounds for $\sup_{u\in S}[\nu_n(s) - \nu_n(u)]$ and $\nu_n(u) - \nu_n(s)$ for $u \in S_j$. This result is stated in Theorem 1 below.

Theorem 1 *Suppose that Assumptions M1 – M3 hold and* (X_1, W_1), …, (X_n, W_n) *are β-mixing. Suppose that* $B_j \ge 1$, $r_j B_j^2 \le n/4$, $r_j \ge 1$, *and there exist constants* $A_0 > 0$, $B_0 > 0$ *such that* $0 < A_j/B_j \le A_0 + 2$ *and* $0 < A_{2,j}/B_{2,j} \le 1/B_0$. *Suppose that* $\|u\|_\infty \le B_j$ *for all* $u \in S_j$. *Suppose that q_n is a positive integer and* $\tilde{q}_n = \lfloor q_n/2 \rfloor$. *Let* $\ell_n = \lfloor n/q_n \rfloor$. *Suppose that* $\delta_n > 0$, $\ell_n \ge A_0^2$ *and* $r_j B_j^2 \le (\ell_n/4) \wedge \delta_n$, $\sigma > 0$, $\tau > 0$,

$$\theta \ge \left(\frac{A_0 + 2}{c(8\tau)}\right) \vee 5, \tag{4}$$

and the penalty $\eta_{1,j}$ is chosen so that

$$\frac{\ell_n \eta_{1,j}}{2} > 24 B_j^2 r_j (1 + 2D_j \log 2) \text{ for each } j \in \Lambda. \tag{5}$$

Let

$$c_1(t) = \frac{t}{(1 + \sqrt{1+t})^2} \text{ and } c(t) = c_1\left(\frac{t}{k_0(A_0 + 2)}\right) \tag{6}$$

for $t > 0$ and take

$$\eta_j = \frac{1}{2}\left(\eta_{1,j} + \frac{B_j^2 r_j \xi}{\ell_n}\right).$$

Then there exists a set Ω_n with

$$P(\Omega_n^c) \le (\ell_n + 1)(\beta_{\tilde{q}_n} + \beta_{q_n - \tilde{q}_n})$$

such that for $\xi > 0$, $j^* \in \Lambda$ *and* $s^* \in S_{j^*}$, *we have*

$$\begin{aligned}(k_1 - 9\tau)\|\hat{s} - s\|^2 I_{\Omega_n} - 9\tau\left(\frac{\sigma^2}{4}\right) \\ \le 1.5\eta_{1,j^*} + (k_2 + \tau)\|s^* - s\|^2 + \frac{1.5\delta_n \xi}{\ell_n}\end{aligned} \tag{7}$$

except on a set of probability at most

$$q_n\left(p_{j^*}^*(\ell_n, \eta_{j^*}) + \sum_{j\in\Lambda} p_j(\ell_n, \eta_j)\right),$$

where β_k denotes the k-th β-mixing coefficient for $k \geq 1$,

$$\begin{aligned} p_j(\ell_n, \eta_j) &= \exp\left(-\frac{c_1(2\tau B_0/k_0)\ell_n\eta_j}{A_{2,j}B_{2,j}}\right) \\ &+1.6\left(1-\exp\left(-\frac{\tau\sigma^2}{2}\left[c(\tau)\wedge\frac{1}{3}\right]\right)\right)^{-1}\left((B_j'\theta)\vee 2\right)^{3D_j}\times \\ &\left[\exp\left(-\frac{\ell_n c(\tau)\eta_j}{2A_jB_j}\right)+\exp\left(-\frac{\ell_n\eta_j}{24B_j^2r_j}\right)\right] \end{aligned} \tag{8}$$

and

$$p^*_{j^*}(\ell_n, \eta_{j^*}) = \exp\left(-\frac{c_1(2\tau B_0/k_0)\ell_n\eta_{j^*}}{A_{2,j^*}B_{2,j^*}}\right). \tag{9}$$

The proof of Theorem 1 is given in Section 5.1.

Remark. In the least regression framework, while the inequality in Lemma 0 in [8] can also be applied to give control of $\nu_n(s)-\nu_n(u)$ (see [5] for example), the exponential inequality given in (21) is more direct. Lemma 0 in [8] gives control of the likelihood function, and in [5], the control of $\nu_n(s)-\nu_n(u)$ is achieved by assuming the W_i's are normally distributed. In contrast, the inequality in (21) is derived following the proof of Theorem 5 in Birgé and Massart [3], which only requires moment conditions of the W_i's.

3 Application to Adaptive B-spline Estimation

In this section, we will apply Theorem 1 to obtain adaptive B-spline estimators for s in (1). Here the regression function s is assumed to be in the Sobolev space $W_m^2[0,1]$ for some $m \geq 1$ and each S_j is taken to be a collection of B-splines that are on $[0,1]$ with order q for some integer $q \geq 1$ and boundary knots at 0 and 1 and distinct internal knots $\xi_1, \ldots, \xi_k$ in $\{1/2^J, \ldots, (2^J-1)/2^J\}$ for some positive integer J. Also, the coefficients for B-splines in S_j are bounded by b in absolute value for some positive integer b, and the index j is $(b, q, \xi_1, \ldots, \xi_k)$. Let

$$\tilde{\Delta}_{2,j} = \max_{1\leq i\leq k+1}(\xi_i - \xi_{i-1}), \tag{10}$$

$$\tilde{\Delta}_{1,j} = \min_{1\leq i\leq k+1}(\xi_i - \xi_{i-1}), \tag{11}$$

$$B_j' = \sqrt{2\pi e}\left(0.5 + \frac{q\sqrt{q}(2q+1)9^{q-1}\sqrt{\tilde{\Delta}_{2,j}}}{\sqrt{\tilde{\Delta}_{1,j}}}\right), \tag{12}$$

$$r_j = 1, \tag{13}$$

and

$$J_j = \min\{J \geq 1 : \ \xi_1, \ldots, \xi_k \text{ are in } \{1/2^J, \ldots, (2^J-1)/2^J\}\ \},$$

where $\xi_0 = 0$ and $\xi_{k+1} = 1$. Let Λ be the set of all $j = (b, q, \xi_1, \ldots, \xi_k)$'s such that b and q are positive integers, $2^{J_j} + q + b \leq n$ and $r_j(2b)^2 \leq \delta_n$, where $\{\delta_n\}$ is chosen such that

$$\lim_{n\to\infty} \delta_n = \infty = \lim_{n\to\infty} \frac{\ell_n}{\delta_n} \text{ and } \lim_{n\to\infty} \frac{\delta_n \log(n)}{n^\alpha} = 0 \text{ for all } \alpha > 0. \tag{14}$$

Then the estimator for s considered here is the penalized least square estimator

$$\hat{s} = \text{argmin}_{j\in\Lambda, u\in S_j} \left(\frac{1}{n} \sum_{i=1}^{n} (Y_i - u(X_i))^2 + \eta_{1,j} \right), \tag{15}$$

where

$$\eta_{1,j} = \frac{a_n r_j (2b)^2}{\ell_n} \Big((k+q)\log(B_j') + \lambda[(\log 2)2^{J_j} + q + b] \Big), \tag{16}$$

$\lambda \geq 1$ is a constant, and $\{a_n\}$ is a sequence of positive numbers such that $\lim_{n\to\infty} a_n = \infty$. The L^2 convergence rate for $\hat{s}$ is given in Theorem 2.

Theorem 2 *Suppose that the regression model in (1) holds, where the regression function s is in $W_m^2[0,1]$, $\{(X_i, W_i)\}_{i\geq 1}$ is a β-mixing series, and its ℓ-th β-mixing coefficient, denoted by β_ℓ, satisfies*

$$\beta_\ell \leq \gamma_1 e^{-\ell\gamma_2} \textit{ for } \ell \geq 1, \tag{17}$$

for some positive constants γ_1 and γ_2, and the W_i's are of mean zero and are independent of the X_i's. Suppose that $Ee^{\alpha|W_i|} < \Gamma$ for some $\alpha > 0$ and $\Gamma \geq 1$ and X_i has a Lebesgue density that is bounded above and bounded below from zero. Suppose that $\{a_n\}$ and $\{\ell_n\}$ are two sequences of positive numbers such that $\lim_{n\to\infty} a_n = \infty$ and $\ell_n = O(n(\log n)^{-(1+\gamma_0)})$ for some $\gamma_0 > 0$. Suppose that $\{\delta_n\}$ is chosen so that (14) holds. Then, for the estimator $\hat{s}$ given in (15) with S_j defined above in this section and $\eta_{1,j}$ defined in (16), we have $E\|\hat{s} - s\|^2 = O(a_n (\log n)^{1+\gamma_0} n^{-2m/(1+2m)})$.

The proof of Theorem 2 is given in Section 5.3.

4 Conclusion

An adaptive estimator for regression function using B-splines that allows non-equally spaced knots is successfully constructed. The estimator is a penalized least square estimator and it achieves the L^2 convergence rate $O(a_n(\log n)^{1+\gamma_0} n^{-2m/(1+2m)})$ for any sequence $\{a_n\}$ such that $\lim_{n\to\infty} a_n = \infty$ and constant $\gamma_0 > 0$ when the regression function is in $W_m^2[0,1]$ for some $m \geq 1$.

5 Proofs

5.1 Proof of Theorem 1

The proof of Theorem 1 is based on the following result:

Lemma 1 *Suppose that Assumptions M1 and M2 hold and* (X_1, W_1), …, (X_n, W_n) *are* β*-mixing. Suppose that* $B_j \geq 1$, $r_j B_j^2 \leq n/4$, $r_j \geq 1$, *and there exist constants* $A_0 > 0$ *and* $B_0 > 0$ *such that* $0 < A_j/B_j \leq A_0 + 2$ *and* $0 < A_{2,j}/B_{2,j} \leq 1/B_0$. *Suppose that* $\|u\|_\infty \leq B_j$ *for all* $u \in S_j$. *Suppose that* q_n *is a positive integer and* $\tilde{q}_n = \lfloor q_n/2 \rfloor$. *Let* $\ell_n = \lfloor n/q_n \rfloor$. *Suppose that* $\ell_n \geq A_0^2$ *and* $r_j B_j^2 \leq \ell_n/4$ *for each* $j \in \Lambda$, $\sigma > 0$, $\tau > 0$, *(4) holds. Then there exists a set* Ω_n *such that* $P(\Omega_n^c) \leq \lceil \ell_n \rceil (\beta_{\tilde{q}_n} I(\tilde{q}_n \geq 1) + \beta_{q_n - \tilde{q}_n})$ *and for* η_j *such that*

$$\ell_n \eta_j > 24 B_j^2 r_j (1 + 2D_j \log 2) \text{ for each } j \in \Lambda, \tag{18}$$

for $j^* \in \Lambda$ *and* $s^* \in S_{j^*}$,

$$P_s(\Omega_n \cap (S_{2,1} \cup S_{2,1}^*)) \quad \leq \quad q_n \left(p_{j^*}^*(\ell_n, \eta_{j^*}) + \sum_{j\in\Lambda} p_j(\ell_n, \eta_j) \right), \tag{19}$$

where $S_{2,1}$ *is the event that*

$$\nu_n[\gamma(\cdot, s) - \gamma(\cdot, u)] > 9\tau \left(\frac{\sigma^2}{4} \vee \|s - u\|^2 \right) + 2\eta_j \text{ for some } u \in S_j$$

for some $j \in \Lambda$,

$$S_{2,1}^* = \left\{ \nu_n[\gamma(\cdot, s^*) - \gamma(\cdot, s)] > \tau \|s - s^*\|^2 + \eta_{j^*} \right\},$$

β_k *is the* k*-th* β*-mixing coefficient for* $k \geq 1$ *and* $p_j(\ell_n, \eta_j)$ *and* $p_j^*(\ell_n, \eta_{j^*})$ *are given in (8) and (9) respectively.*

Lemma 1 can be derived by first establishing the special case where (X_1, W_1), …, (X_n, W_n) are independent and then applying a corollary of Berbee's Lemma taken from Claim 2 in [1]. The independent case of Lemma 1 is stated in Lemma 2 below, whose proof is given in Section 5.2. The corollary of Berbee's Lemma is stated in Fact 1 below.

Lemma 2 *Suppose that Assumptions M1 and M2 hold and* (X_1, W_1), …, (X_n, W_n) *are independent. Suppose that* $B_j \geq 1$, $r_j B_j^2 \leq n/4$, $r_j \geq 1$, *and there exist constants* $A_0 > 0$ *and* $B_0 > 0$ *such that* $0 < A_j/B_j \leq A_0 + 2$ *and* $0 < A_{2,j}/B_{2,j} \leq 1/B_0$. *Suppose that* $\|u\|_\infty \leq B_j$ *for all* $u \in S_j$. *Suppose that* $n \geq A_0^2$, $\sigma > 0$, $\tau > 0$, *(4) and and*

$$n\eta_j > 24B_j^2 r_j(1 + 2D_j \log 2) \text{ for each } j \in \Lambda. \tag{20}$$

Then we have

$$P_s\left[\nu_n[\gamma(\cdot, s) - \gamma(\cdot, u)] > 9\tau\left(\frac{\sigma^2}{4} \vee \|s-u\|^2\right) + 2\eta \text{ for some } u \in S_j\right]$$

$$\begin{aligned}\leq\ & \exp\left(-\frac{c_1(2\tau B_0/k_0)n\eta_j}{A_{2,j}B_{2,j}}\right) \\ & +1.6\left(1 - \exp\left(-\frac{\tau\sigma^2}{2}\left[c(\tau) \wedge \frac{1}{3}\right]\right)\right)^{-1}\left((B_j'\theta) \vee 2\right)^{3D_j} \times \\ & \left[\exp\left(-\frac{nc(\tau)\eta_j}{2A_jB_j}\right) + \exp\left(-\frac{n\eta_j}{24B_j^2 r_j}\right)\right],\end{aligned} \tag{21}$$

where c_1 *and* c *are defined in (6). In addition, for* $u \in S_j$ *and* $\eta_j \geq 0$, *we have*

$$\begin{aligned}& P_s\left[\nu_n[\gamma(\cdot, u) - \gamma(\cdot, s)] > \tau\|u-s\|^2 + \eta_j\right] \\ & \leq \exp\left(-\frac{c_1(2\tau B_0/k_0)n\eta_j}{A_{2,j}B_j 2, j}\right)\end{aligned} \tag{22}$$

Fact 1 *Suppose that a sequence* $\{u_i\}_{i=1}^\infty$ *is* β*-mixing and for* $n \geq 1$, $\tilde{q}_n$ *and* q_n *are integers such that* $0 \leq \tilde{q}_n \leq q_n/2$, $q_n \geq 1$. *Let* u_i *for* $i \geq 1$. *Let* $\ell_n = \lfloor n/q_n \rfloor$. *Then there exists* u_i^*: $i = 1$, …, $\lceil \ell_n \rceil q_n$ *such that (i)–(iii) hold.*

(i) For $\ell = 1$, …, $\lceil \ell_n \rceil$, *let* $U_{\ell,1} = (u_{(\ell-1)q_n+1}, \ldots, u_{(\ell-1)q_n+\tilde{q}_n})^T$, $U_{\ell,1}^* = (u^*_{(\ell-1)q_n+1}, \ldots, u^*_{(\ell-1)q_n+\tilde{q}_n})^T$, $U_{\ell,2} = (u_{(\ell-1)q_n+\tilde{q}_n+1}, \ldots, u_{\ell q_n})^T$, *and* $U_{\ell,2}^* = (u^*_{(\ell-1)q_n+\tilde{q}_n+1}, \ldots, u^*_{\ell q_n})^T$, *then for each* $\delta \in \{1,2\}$, $U_{\ell,\delta}$ *and* $U_{\ell,\delta}^*$ *have the same distribution.*

(ii) For $\ell = 1, \ldots, \lceil \ell_n \rceil$, $P(U_{\ell,1} \neq U^*_{\ell,1}) \leq \beta_{q_n - \tilde{q}_n}$ and $P(U_{\ell,2} \neq U^*_{\ell,2}) \leq \beta_{\tilde{q}_n}$, where β_k denotes the k-th β-mixing coefficient for $k \geq 1$.

(iii) For each $\delta \in \{1, 2\}$, $U^*_{1,\delta}, \ldots, U^*_{\lceil \ell_n \rceil,\delta}$ are independent.

Next, we will prove Lemma 1 using Lemma 2 and Fact 1. To find an upper bound for $P_s(S_{2,1})$ when the series $\{(X_i, W_i)\}_{i \geq 1}$ is β-mixing, we will apply Fact 1 with $u_i = (X_i, W_i)$. We will only prove the case $q_n > 1$, which implies that $\tilde{q}_n = \lfloor q_n/2 \rfloor \geq 1$, since the proof for the case $q_n = 1$ is similar. Let $(X^*_i, W^*_i) = u^*_i$, and

$$\Omega_n = \{(X_i, W_i) = (X^*_i, W^*_i) \text{ for } i = 1, \ldots, n\}.$$

Then $P_s(S_{2,1}) \leq P(\Omega_n^c) + P_s(S_{2,1} \cap \Omega_n)$. For $k = 1, \ldots, q_n$, let $\Gamma_k = \{i : i = k + \ell q_n$ for some integer ℓ and $1 \leq i \leq n\}$ and

$$\nu_{n,k}[\gamma(\cdot, u)] = \left(\frac{1}{|\Gamma_k|} \sum_{i \in \Gamma_k} \gamma(Z_i, u) \right) - E\gamma(Z_1, u)$$

for all u, where $|\Gamma_k|$ denotes the number of elements in Γ_k, which is at most $\lceil \ell_n \rceil$. Let $S_{2,1,k}$ be the event that

$$\nu_{n,k}[\gamma(\cdot, s) - \gamma(\cdot, u)] > 9\tau \left(\frac{\sigma^2}{4} \vee \|s - u\|^2 \right) + 2\eta_j$$

for some $u \in S_j$ for some $j \in \Lambda$ and let

$$S^*_{2,1,k} = \left\{ \nu_{n,k}[\gamma(\cdot, s^*) - \gamma(\cdot, s)] > \tau \|s^* - s\|^2 + \eta_{j^*} \right\}.$$

Then $S_{2,1} \subset \cup_{k=1}^{q_n} S_{2,1,k}$ and $S^*_{2,1} \subset \cup_{k=1}^{q_n} S^*_{2,1,k}$ and it follows from (iii) in Claim 2 and Theorem 2 that $P(\Omega_n^c) \leq \lceil \ell_n \rceil (\beta_{\tilde{q}_n} + \beta_{q_n - \tilde{q}_n})$ and

$$\begin{aligned}
P_s(\Omega_n \cap (S_{2,1} \cup S^*_{2,1})) &\leq \sum_{k=1}^{q_n} (P_s(S_{2,1,k} \cap \Omega_n) + P_s(S^*_{2,1_k} \cap \Omega_n)) \\
&\leq \sum_{k=1}^{q_n} \sum_{j \in \Lambda} p_j(|\Gamma_k|, \eta_j) + \sum_{k=1}^{q_n} p^*_{j^*}(|\Gamma_k|, \eta_{j^*}) \\
&\leq \left(q_n \sum_{j \in \Lambda} p_j(\ell_n, \eta_j) \right) + q_n p^*_{j^*}(\ell_n, \eta_{j^*})
\end{aligned}$$

if (4) holds, $\ell_n \geq A_0^2$ and for each $j \in \Lambda$, $r_j B_j^2 \leq \ell_n/4$, and $\ell_n \eta_j > 24 B_j^2 r_j (1 + 2D_j \log 2)$. The proof of Lemma 1 is complete.

To give an error bound for $\|\hat{s}-s\|$ using Lemma 1, for $\xi>0$, take

$$\eta_j = \frac{1}{2}\left(\eta_{1,j} + \frac{B_j^2 r_j \xi}{\ell_n}\right),$$

then (18) holds and on $S_{2,1}^c$, we have

$$\begin{aligned}
& k_1\|u-s\|^2 - 9\tau\left(\frac{\sigma^2}{4}\vee\|s-u\|^2\right) \\
& \le E(\gamma(Z_i,u)) - E(\gamma(Z_i,s)) - 9\tau\left(\frac{\sigma^2}{4}\vee\|s-u\|^2\right) \\
& \le \frac{1}{n}\sum_{i=1}^n \gamma(Z_i,u) + \eta_{1,j} + \frac{B_j^2 r_j\xi}{\ell_n} - \frac{1}{n}\sum_{i=1}^n \gamma(Z_i,s) \\
& \le \frac{1}{n}\sum_{i=1}^n \gamma(Z_i,u) + \eta_{1,j} + \frac{\delta_n\xi}{\ell_n} - \frac{1}{n}\sum_{i=1}^n \gamma(Z_i,s)
\end{aligned}$$

for all $u\in S_j$ for all $j\in\Lambda$, and on $(S_{2,1}^*)^c$, we have

$$\begin{aligned}
& \frac{1}{n}\sum_{i=1}^n \gamma(Z_i,s^*) - \frac{1}{n}\sum_{i=1}^n \gamma(Z_i,s) \\
& \le E[\gamma(Z_i,s^*) - \gamma(Z_i,s)] + \tau\|s^*-s\|^2 + \eta_{j^*} \\
& \le E[\gamma(Z_i,s^*) - \gamma(Z_i,s)] + \tau\|s^*-s\|^2 + \frac{\eta_{1,j^*}}{2} + \frac{\delta_n\xi}{2\ell_n}.
\end{aligned}$$

Thus on $\Omega_n\cap(S_{2,1}^c\cap(S_{2,1}^*)^c)$, we have the error bound

$$\begin{aligned}
& (k_1-9\tau)\|\hat{s}-s\|^2 - 9\tau\left(\frac{\sigma^2}{4}\right) \\
\le\ & \frac{1}{n}\sum_{i=1}^n \gamma(Z_i,s^*) + \eta_{1,j^*} + \frac{\delta_n\xi}{\ell_n} - \frac{1}{n}\sum_{i=1}^n \gamma(Z_i,s) \\
\le\ & 1.5\left(\eta_{1,j^*} + \frac{\delta_n\xi}{\ell_n}\right) + E[\gamma(Z_i,s^*) - \gamma(Z_i,s)] + \tau\|s^*-s\|^2 \\
\le\ & 1.5\eta_{1,j^*} + (k_2+\tau)\|s^*-s\|^2 + \frac{1.5\delta_n\xi}{\ell_n}.
\end{aligned}$$

Let

$$U = (k_1-9\tau)\|\hat{s}-s\|^2 I_{\Omega_n} - \left(9\tau\left(\frac{\sigma^2}{4}\right) + 1.5\eta_{1,j^*} + (k_2+\tau)\|s^*-s\|^2\right),$$

then the above result can be expressed as $P(U>0)\le P(\Omega_n\cap(S_{2,1}\cup S_{2,1}^*))$, where $P(\Omega_n^c)\le(\ell_n+1)(\beta_{\tilde{q}_n}+\beta_{q_n-\tilde{q}_n})$ and an upper bounds for $P(\Omega_n\cap(S_{2,1}\cup S_{2,1}^*))$ is given in (19).

5.2 Proof of Lemma 2

To prove Lemma 2, we will first establish the following result:

Fact 2 *Suppose that the conditions in Lemma 2 hold. Suppose that θ and η_j satisfy (4) and (18) for some $\tau > 0$. Then for $\sigma > 0$, $x > 0$, $x_k > 0$, we have*

$$P_s\left[\sup_{u\in B(s^*,\sigma)} \nu_n[\gamma(\cdot,s^*)-\gamma(\cdot,u)] > \tau\sigma^2+\eta_j\right]$$

$$\leq \exp(-nx+H_0)+\sum_k \exp(-nx_k+H_k+H_{k+1}) \qquad (23)$$

if

$$\sigma\sqrt{2k_0x}+B_jx+(\sigma/\theta)\sum_{k=0}^{\infty}2^{-k}(\sqrt{5k_0x_k}+1.5rx_k)\leq\frac{\tau\sigma^2+\eta_j}{A_j}, \qquad (24)$$

where $H_k = D_j\log(B'_j 2^k\theta) = D_j\log(B'_j\theta)+kD_j\log(2)$.

The proof of Fact 2 relies on a version of Bernstein's inequality given Lemma 8 in [3], which is given below.

Fact 3 *Suppose that $U_1, \ldots, U_n$ are independent random variables such that*

$$\frac{1}{n}\sum_{i=1}^{n}E[|U_i|^m]\leq\frac{m!}{2}v^2c^{m-2} \text{ for all } m\geq 2$$

for some positive constants v and c. Then, for $x \geq 0$,

$$P\left[\sum_{i=1}^{n}U_i - E\left(\sum_{i=1}^{n}U_i\right)\geq n(v\sqrt{2x}+cx)\right]\leq\exp(-nx).$$

Fact 2 follows from Fact 3 and Assumption M2. To prove Fact 2, let $\delta_k = 2^{-k}\sigma/\theta$. Let T_k be a subset of $B(s^*,\sigma)$ such that $|T_k| \leq (B'2^k\theta)^D$ and for every $u \in B(s^*,\sigma)$, there exists $v \in T_k$ such that $\|u-v\| \leq \delta_k$ and $\|\Delta(\cdot,u,v)\|_\infty \leq r\delta_k$. Let V_k be the set of (u_k,u_{k+1})'s such that $u_k \in T_k$, $u_{k+1}\in T_{k+1}$ and there exists $u \in S_j$ such that $\|u_{k+1}-u\| \leq \delta_{k+1}$, $\|u_k - u\| \leq \delta_k$, $\|\Delta(\cdot,u,u_{k+1})\|_\infty \leq r\delta_{k+1}$, and $\|\Delta(\cdot,u,u_k)\|_\infty \leq r\delta_k$. For $u \in B(s^*,\sigma)$, since

$$\begin{aligned}&\nu_n[\gamma(\cdot,s^*)-\gamma(\cdot,u)]\\ &= \nu_n[\gamma(\cdot,s^*)-\gamma(\cdot,u_0)]+\sum_{k=0}^{\infty}\left(\nu_n[\gamma(\cdot,u)-\gamma(\cdot,u_{k+1})]-\nu_n[\gamma(\cdot,u)-\gamma(\cdot,u_k)]\right),\end{aligned}$$

for some $u_k \in T_k$ for $k = 0, 1, \ldots$, we have

$$\begin{aligned}
&P_s\left[\sup_{u\in B(s^*,\sigma)} \nu_n[\gamma(\cdot,s^*)-\gamma(\cdot,u)] > \tau\sigma^2+\eta_j\right] \\
&\leq \underbrace{P_s\left[\nu_n[\gamma(\cdot,s^*)-\gamma(\cdot,u_0)] > y \text{ for some } u_0 \in T_0\right]}_{I} \\
&+ \sum_k \underbrace{P_s\left[\nu_n[\gamma(\cdot,u_k)-\gamma(\cdot,u_{k+1})] > y_k \text{ for some } (u_k,u_{k+1}) \in V_k\right]}_{II_k}
\end{aligned}$$

if

$$y + \sum_k y_k \leq \tau\sigma^2 + \eta_j.$$

Below we will use Fact 3 to find bounds for I and II_k. To derive an upper bound for I, note that

$$\begin{aligned}
E_s|\gamma(Z_i,s^*)-\gamma(Z_i,u_0)|^m &\leq E_s[M_j^m(W_i)]E_s[\Delta_j^m(X_i,s^*,u_0)] \\
&\leq a_m A_j^m E_s[\Delta^m(X_i,s^*,u_0)]
\end{aligned}$$

and

$$\frac{1}{n}\sum_{i=1}^n E_s|\gamma(Z_i,s^*)-\gamma(Z_i,u_0)|^m \leq a_m A_j^m b_m k_0 \sigma^2 B_j^{m-2}.$$

Take $y = A_j(\sigma\sqrt{2k_0x} + B_jx)$, then $I \leq (B_j'\theta)_j^D \exp(-ny) = \exp(-ny + H_0)$.

To derive an upper bound for II_k, note that

$$E_s|\gamma(Z_i,u_{k+1})-\gamma(Z_i,u_k)|^m \leq E_s[M_j^m(W_i)]E_s(\Delta_j(X_i,u,u_{k+1})+\Delta_j(X_i,u,u_k))^m,$$

where

$$\begin{aligned}
&E_s(\Delta(X_i,u,u_{k+1})+\Delta(X_i,u,u_k))^m \\
&\leq E_s(\Delta_j(X_i,u,u_{k+1})+\Delta_j(X_i,u,u_k))^2(\|\Delta_j(\cdot,u,u_{k+1})\|_\infty + \|\Delta_j(\cdot,u,u_{k+1})\|_\infty)^{m-2},
\end{aligned}$$

so

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n E_s|\gamma(Z_i,u_{k+1})-\gamma(Z_i,u_k)|^m &\leq a_m A_j^m b_m k_0[2(\delta_k^2+\delta_{k+1}^2)(r_j\delta_k + r_j\delta_{k+1})^{m-2}] \\
&\leq \frac{m!}{2}\left(\frac{5}{2}\delta_k^2 A_j^2 k_0\right)\left(A_j\frac{3r_j\delta_k}{2}\right)^{m-2}.
\end{aligned}$$

Therefore, if $y_k = A_j\delta_k(\sqrt{5k_0x_k} + 1.5r_jx_k)$, then

$$II_k \leq \left(B_j'2^k\theta\right)^{D_j}\left(B_j'2^{k+1}\theta\right)^{D_j}\exp(-nx_k) = \exp(-nx_k + H_k + H_{k+1}).$$

From the upper bounds for I and II_k given above, (23) holds true if

$$A_j(\sigma\sqrt{2k_0x} + B_jx) + \sum_k A_j\delta_k(\sqrt{5k_0x_k} + 1.5r_jx_k) \le \tau\sigma^2 + \eta_j,$$

so (24) implies (23) and we have Fact 2.

Next we will apply Fact 2 with specific x and x_k's for case $0 < \sigma \le 2B_j$. In such case, (24) is implied by

$$\sigma\sqrt{2k_0x} + B_jx + \frac{\sigma}{\theta}\sum_{k=0}^{\infty} 2^{-k}(\sqrt{5k_0x_k}) + \frac{2B_j}{\theta}\sum_{k=0}^{\infty} 2^{-k}(1.5r_jx_k) \le \frac{\tau\sigma^2 + \eta_j}{A_j}. \tag{25}$$

Let $g(x) = (x/(1+\sqrt{1+x}))^2$ for $x > 0$, then $g(x)/x < 1$ is increasing on $(0, \infty)$. Let $x_k = (k+1)\tilde{y}$, then (25) holds if

$$0 < x \le \frac{k_0\sigma^2}{2B_j^2} g\left(\frac{B_j}{A_jk_0\sigma^2}(\tau\sigma^2 + \eta_j)\right)$$

and

$$0 < \tilde{y} \le \frac{c_1^2k_0\sigma^2}{16B_j^2c_0^2r_j^2} g\left(\frac{4c_0r_j\theta B_j}{c_1^2A_jk_0\sigma^2}(\tau\sigma^2 + \eta_j)\right),$$

where $c_1 = \sqrt{5}\sum_{k=0}^{\infty}(\sqrt{k+1})2^{-k} \approx 3.789034$ and $c_0 = 1.5\sum_{k=0}^{\infty}(k+1)2^{-k} = 6$. If x and $\tilde{y}$ satisfy the above constraints and $\tilde{y} > 2D_j\log(2)/n$, then

$$P_s\left[\sup_{u\in B(s^*,\sigma)} \nu_n[\gamma(\cdot,s^*) - \gamma(\cdot,u)] > \tau\sigma^2 + \eta_j\right]$$
$$\le \exp(-nx + D_j\log(B_j'\theta)) + \frac{\exp\left(-n\tilde{y} + D_j\log(2) + 2D_j\log(B_j'\theta)\right)}{1 - \exp(-[n\tilde{y} - 2D_j\log(2)])}.$$

From the assumptions that $r_j \ge 1$, $B_j \ge 1$, $B_j/A_j \ge A_0 + 2$ and $\theta > 5$, we have

$$\frac{B_j}{A_j\sigma^2}(\tau\sigma^2) \ge \frac{\tau}{A_0 + 2}$$

and

$$\frac{4c_0r_j\theta B_j}{c_1^2A_j\sigma^2}(\tau\sigma^2) \ge \frac{20c_0B_j\tau}{c_1^2A_j} \ge \frac{8\tau}{A_0 + 2}.$$

Since

$$c(\tau) = g\left(\frac{\tau}{k_0(A_0 + 2)}\right)\frac{k_0(A_0 + 2)}{\tau},$$

take

$$x = c(\tau)\frac{\tau\sigma^2 + \eta_j}{2A_jB_j} \text{ and } \tilde{y} = c(8\tau)\frac{\theta(\tau\sigma^2 + \eta_j)}{4A_jB_jc_0r_j},$$

then by (4),

$$n\tilde{y} \geq \frac{n\eta_j c(8\tau)\theta}{4A_j B_j c_0 r_j} \geq \frac{n\eta_j c(8\tau)\theta}{4(A_0+2)B_j^2 c_0 r_j} \geq \frac{n\eta_j}{4B_j^2 c_0 r_j}.$$

Thus $n\tilde{y} \geq 1 + 2D_j \log 2$ if (20) holds. In such case,

$$\begin{aligned}
&P_s\left[\sup_{u\in B(s^*,\sigma)} \nu_n[\gamma(\cdot,s^*)-\gamma(\cdot,u)] > \tau\sigma^2+\eta_j\right]\\
&\leq \exp\left(-\frac{nc(\tau)\tau\sigma^2}{2A_jB_j} - \frac{nc(\tau)\eta_j}{2A_jB_j} + D_j\log(B_j'\theta)\right)\\
&+\exp\left(-\frac{nc(8\tau)\theta\tau\sigma^2}{4A_jB_jc_0r_j} - \frac{nc(8\tau)\theta\eta_j}{4A_jB_jc_0r_j} + D_j\log(2) + 2D_j\log(B_j'\theta)\right)\\
&\times\left(1-\exp(-[\frac{nc(8\tau)\theta\tau\sigma^2}{4A_jB_jc_0r_j} + \frac{nc(8\tau)\theta\eta_j}{4A_jB_jc_0r_j} - 2D_j\log(2)])\right)^{-1}\\
&\leq \exp\left(-\frac{nc(\tau)\tau\sigma^2}{2A_jB_j} - \frac{nc(\tau)\eta_j}{2A_jB_j} + D_j\log(B_j'\theta)\right)\\
&+\exp\left(-\frac{n\tau\sigma^2}{4B_j^2c_0r_j} - \frac{n\eta_j}{4B_j^2c_0r_j} + D_j\log(2) + 2D_j\log(B_j'\theta)\right)\left(1-e^{-1}\right)^{-1}.
\end{aligned}$$

In summary, we have prove the following fact assuming $\sigma \leq 2B_j$:

Fact 4 *Under the conditions in Lemma 2, for $\tau > 0$, $\sigma > 0$, if θ and η_j satisfy (4) and (18), then*

$$\begin{aligned}
&P_s\left[\sup_{u\in B(s^*,\sigma)} \nu_n[\gamma(\cdot,s^*)-\gamma(\cdot,u)] > \tau\sigma^2+\eta_j\right]\\
&\leq \exp\left(-\frac{nc(\tau)\tau\sigma^2}{2A_jB_j} - \frac{nc(\tau)\eta_j}{2A_jB_j} + D_j\log(B_j'\theta)\right)\\
&+\exp\left(-\frac{n\tau\sigma^2}{4B_j^2c_0r_j} - \frac{n\eta_j}{4B_j^2c_0r_j} + D_j\log(2) + 2D_j\log(B_j'\theta)\right)\left(1-e^{-1}\right)^{-1}.
\end{aligned}$$

Note that Fact 4 also holds for $\sigma > 2B_j$. To see this, note that $\|u\|_\infty \leq B_j$ for all $u \in S_j$, so

$$\begin{aligned}
&P_s\left[\sup_{u\in B(s^*,\sigma)} \nu_n[\gamma(\cdot,s^*)-\gamma(\cdot,u)] > \tau\sigma^2+\eta_j\right]\\
&= P_s\left[\sup_{u\in B(s^*,2B_j)} \nu_n[\gamma(\cdot,s^*)-\gamma(\cdot,u)] > \tau(2B_j)^2 + \tau[\sigma^2-(2B_j)^2]+\eta_j\right].
\end{aligned}$$

Apply Fact 4 with $\sigma = 2B_j$ and replace η_j with $\tau[\sigma^2 - (2B_j)^2] + \eta_j$, then it is clear that Fact 4 also holds for $\sigma > 2B_j$.

To prove Lemma 2 using Fact 4, for $\varepsilon > 0$, choose $s^* \in S_j$ so that

$$\|s - s^*\| \leq \|s - u\| + \varepsilon \text{ for all } u \in S_j,$$

and then we will derive upper bounds for

$$II = P_s\left[\nu_n[\gamma(\cdot, s^*) - \gamma(\cdot, u)] > 2\tau(\sigma^2 \vee \|s^* - u\|^2) + \eta_j \text{ for some } u \text{ in } S_j\right]$$

and

$$III = P_s\left[\nu_n[\gamma(\cdot, s) - \gamma(\cdot, s^*)] > \tau\|s^* - s\|^2 + \eta_j\right]$$

to control $\nu_n[\gamma(\cdot, s) - \gamma(\cdot, u)]$.

To find an upper bound for II, note that

$$\begin{aligned}
&P_s\left[\nu_n[\gamma(\cdot, s^*) - \gamma(\cdot, u)] > 2\tau(\sigma^2 \vee \|s^* - u\|^2) + \eta_j \text{ for some } u \text{ in } S_j\right] \\
&\leq \sum_{k=1}^{\infty} P_s\left[\nu_n[\gamma(\cdot, s^*) - \gamma(\cdot, u)] > \tau(k+1)\sigma^2 + \eta_j \text{ for some } u \in S_{1,k}\right] \\
&+P_s\left[\sup_{u \in B(s^*, \sigma)} \nu_n[\gamma(\cdot, s^*) - \gamma(\cdot, u)] > 2\tau\sigma^2 + \eta_j\right],
\end{aligned}$$

where $S_{1,k} = S_j \cap B(s^*, \sigma\sqrt{(k+1)}) \cap B(s^*, \sigma\sqrt{k})^c$. Apply Fact 4 with σ replaced by $\sigma\sqrt{k+1}$, we have

$$\begin{aligned}
&II = P_s\left[\nu_n[\gamma(\cdot, s^*) - \gamma(\cdot, u)] > 2\tau(\sigma^2 \vee \|s^* - u\|^2) + \eta_j \text{ for some } u \text{ in } S_j\right] \\
&\leq \sum_{k=0}^{\infty} \exp\left(-\frac{nc(\tau)\tau(k+1)\sigma^2}{2A_jB_j} - \frac{nc(\tau)\eta_j}{2A_jB_j} + D_j\log(B_j'\theta)\right) \\
&+\sum_{k=0}^{\infty} \exp\left(-\frac{n\tau(k+1)\sigma^2}{4B_j^2c_0r_j} - \frac{n\eta_j}{4B_j^2c_0r_j} + D_j\log(2) + 2D_j\log(B_j'\theta)\right)\left(1 - e^{-1}\right)^{-1} \\
&\leq \left(1 - \exp\left(-\frac{nc(\tau)\tau\sigma^2}{2A_jB_j}\right)\right)^{-1} \exp\left(-\frac{nc(\tau)\tau\sigma^2}{2A_jB_j} - \frac{nc(\tau)\eta_j}{2A_jB_j} + D_j\log(B_j'\theta)\right) \\
&+1.6\left(1 - \exp\left(-\frac{n\tau\sigma^2}{4B_j^2c_0r_j}\right)\right)^{-1} \exp\left(-\frac{n\tau\sigma^2}{4B_j^2c_0r_j} - \frac{n\eta_j}{4B_j^2c_0r_j} + D_j\log(2) + 2D_j\log(B_j'\theta)\right).
\end{aligned}$$

Since $2B_j^2 \leq 2r_jB_j^2 \leq n/2$ and $n \geq A_0^2$, we have $A_jB_j \leq n$ and $4r_jB_j^2 \leq n$. Thus the above upper bound for II is at most

$$1.6\left(1 - \exp\left(-\frac{\tau\sigma^2}{2}\left[c(\tau) \wedge \frac{1}{3}\right]\right)\right)^{-1}\left((B_j'\theta) \vee 2\right)^{3D_j}\left[\exp\left(-\frac{nc(\tau)\eta_j}{2A_jB_j}\right) + \exp\left(-\frac{n\eta_j}{24B_j^2r_j}\right)\right].$$

To derive an upper bound for III, note that

$$E_s|\gamma(Z_i, s^*) - \gamma(Z_i, s)|^m \leq E_s[M_{2,j}^m(W_i)]E_s[\Delta_{2,j}^m(X_i, s^*, s)],$$

so

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} E_s|\gamma(Z_i, s^*) - \gamma(Z_i, s)|^m &\leq a_m A_{2,j}^m b_m k_0 \|s^* - s\|^2 B_{2,j}^{m-2} \\ &\leq \frac{m!}{2}\left(k_0\|s^* - s\|^2 A_{2,j}^2\right)(A_{2,j}B_{2,j})^{m-2}.\end{aligned}$$

For $x > 0$ and

$$y_2 \geq \|s^* - s\|A_{2,j}\sqrt{2k_0 x} + A_{2,j}B_{2,j}x = A_{2,j}(\|s^* - s\|\sqrt{2k_0 x} + B_{2,j}x), \quad (26)$$

we have

$$P_s\left[\nu_n[\gamma(\cdot, s^*) - \gamma(\cdot, s)] > y_2\right] \leq \exp(-nx).$$

(26) is equivalent to

$$x \leq g\left(\frac{2B_{2,j}y_2}{k_0 A_{2,j}\|s - s^*\|^2}\right)\frac{k_0\|s - s^*\|^2}{2B_{2,j}^2}, \quad (27)$$

where $g(x) = (x/(1 + \sqrt{1 + x}))^2$. Since $g(x)/x$ is increasing on $(0, \infty)$ and $B_{2,j}/A_{2,j} \geq B_0$, for $y_2 \geq \tau\|s - s^*\|^2$, (27) holds if

$$x \leq \frac{g(2\tau B_0/k_0)}{2\tau B_0/k_0}\frac{y_2}{A_{2,j}B_{2,j}} = \frac{c_1(2\tau B_0/k_0)y_2}{A_{2,j}B_{2,j}}.$$

Therefore,

$$\begin{aligned} III &= P_s\left[\nu_n[\gamma(\cdot, s^*) - \gamma(\cdot, s)] > \tau\|s - s^*\|^2 + \eta_j\right] \\ &\leq \exp\left(-\frac{c_1(2\tau B_0/k_0)n\eta_j}{A_{2,j}B_{2,j}}\right). \quad (28)\end{aligned}$$

From the above bounds for II and III and the fact that

$$\begin{aligned}&2\tau(\sigma^2 \vee \|s^* - u\|^2) + \eta_j + \tau\|s - s^*\|^2 + \eta_j \\ &\leq 2\tau(\sigma^2 \vee (2\|s - u\| + \varepsilon)^2) + \tau(\|s - u\| + \varepsilon)^2 + 2\eta_j \\ &\leq 9\tau\left(\frac{\sigma^2}{4} \vee (\|s - u\| + \varepsilon)^2\right) + 2\eta_j,\end{aligned}$$

we have

$$P_s\left[\nu_n[\gamma(\cdot, s) - \gamma(\cdot, u)] > 9\tau\left(\frac{\sigma^2}{4} \vee (\|s - u\| + \varepsilon)^2\right) + 2\eta_j \text{ for some } u \in S\right]$$

$$
\begin{aligned}
&\leq \exp\left(-\frac{c_1(2\tau B_0/k_0)n\eta_j}{A_{2,j}B_{2,j}}\right) \\
&+1.6\left(1-\exp\left(-\frac{\tau\sigma^2}{2}\left[c(\tau)\wedge\frac{1}{3}\right]\right)\right)^{-1}\left((B_j'\theta)\vee 2\right)^{3D_j}\times \\
&\left[\exp\left(-\frac{nc(\tau)\eta_j}{2A_jB_j}\right)+\exp\left(-\frac{n\eta_j}{24B_j^2 r_j}\right)\right].
\end{aligned}
$$

Since $\varepsilon > 0$ is arbitrary, we can obtain (21) by letting $\varepsilon \to 0$.

It is clear that (22) holds since (28) still holds if s^* is replaced by any $u \in S_j$. The proof of Lemma 2 is complete.

5.3 Proof of Theorem 2

Proof of Theorem 2. Theorem 2 is an application of Theorem 1 with $Z_i = (X_i, Y_i)$ and $\gamma(z,t) = \gamma((x,y),t) = (y - t(x))^2$. To apply Theorem 1, Assumptions M1 – M3 will be verified first.

Verification of Assumption M1. Note that

$$
\begin{aligned}
|\gamma(z,u)-\gamma(z,v)| &= |u(x)-v(x)|\cdot|2y-u(x)-v(x)| \\
&= |u(x)-v(x)|\cdot|2(s(x)+w)-u(x)-v(x)|,
\end{aligned}
$$

where

$$|2(s(x)+w)-u(x)-v(x)| \leq 2(|w|+\|s\|_\infty+b)$$

for u, $v \in S_j$ since functions in S_j are bounded by b in sup-norm. Take $M_j(w) = 2(|w|+\|s\|_\infty+b)$ and $\Delta_j(x,u,v) = |u(x)-v(x)|$, then

$$|\gamma(z,u)-\gamma(z,v)| \leq M_j(w)\Delta_j(x,u,v).$$

An upper bound for $E[M_j^m(W_i)]$ can be obtained by controlling $E|W_i|^m$:

$$
\begin{aligned}
E[M_j^m(W_i)] &= E[2^m(|W_i|+\|s\|_\infty+b)^m] \\
&\leq 4^m\left(\frac{1}{2}E(|W_i|^m)+\frac{1}{2}(\|s\|_\infty+b)^m\right) \\
&\leq \frac{m!}{2}4^m\left(\frac{\Gamma}{\alpha^m}+\frac{(\|s\|_\infty+b)^m}{m!}\right) \\
&\leq \frac{m!}{2}4^m\left(\frac{\Gamma}{\alpha}+\|s\|_\infty+b\right)^m. \qquad (29)
\end{aligned}
$$

Here the inequality $E|W_i|^m \leq \dfrac{m!\Gamma}{\alpha^m}$ follows from the assumption that $Ee^{\alpha|W_i|} < \Gamma$ for some $\alpha > 0$ and $\Gamma \geq 1$.

To control $E_s[\Delta_j^m(X_i,u,v)] = E_s|u(X_i)-v(X_i)|^m$, note that for $u \in S_j$ and $v \in S_j$,

$$E_s|u(X_i)-v(X_i)|^m \le E_s[u(X_1)-v(X_1)]^2\|u-v\|_\infty^{m-2},$$

where $E_s[u(X_1)-v(X_1)]^2 \le k_0\|u-v\|^2$ for some constant k_0 that does not depend on j since the density of X_i is bounded above. Thus for $u \in S_j$ and $v \in S_j$,

$$\begin{aligned} E_s[\Delta_j^m(X_i,u,v)] &\le k_0\|u-v\|^2\|u-v\|_\infty^{m-2} & (30)\\ &\le k_0\|u-v\|^2\,(2b)^{m-2}. & (31)\end{aligned}$$

To control $|\gamma(z,u)-\gamma(z,v)|$ for $u \in S_j$ and $v = s$, take $M_{2,j}(w) = 2|w| + \|s\|_\infty + b$ and $\Delta_{2,j}(x,u,s) = |u(x)-s(x)| = \Delta_j(x,u,s)$, then

$$|\gamma(z,u)-\gamma(z,s)| \le M_{2,j}(w)\Delta_{2,j}(x,u,s)$$

since(29) still holds and

$$|2(s(x)+w)-u(x)-s(x)| \le 2|w| + \|s\|_\infty + b.$$

Modify slightly the equations for deriving (29) by replacing $\|s\|_\infty + b$ with $(\|s\|_\infty + b)/2$ and we have

$$E[M_{2,j}^m(W_i)] \le \frac{m!}{2}4^m\left(\frac{\Gamma}{\alpha} + \frac{\|s\|_\infty + b}{2}\right)^m. \qquad (32)$$

Also, from (30), we have

$$E_s[\Delta_{2,j}^m(X_i,u,s)] \le k_0\|u-v\|^2(\|s\|_\infty + b)^m \qquad (33)$$

Let $A_0 = 4(\Gamma/\alpha + \|s\|_\infty)$ and $B_0 = 0.5$. From (29), (31), (32), (33), Assumption M1 holds with $b_m = 1$, $a_m = m!/2$, $B_j = 2b$,

$$A_j = 4\left(\frac{\Gamma}{\alpha} + \|s\|_\infty + b\right) = A_0 + 2B_j,$$

$A_{2,j} = A_0 + B_j$, and $B_{2,j} = (A_0+B_j)/2$. It is clear that $B_j \ge 1$ and $\|u\|_\infty \le B_j$ for all $u \in S_j$. Also, $0 < A_j/B_j \le A_0 + 2$ and $A_{2,j}/B_{2,j} = 2 = 1/B_0$.

Verification of Assumption M2. To identify the constants in Assumption M2, Facts 5 and 6 will be applied. These two facts are first stated and proved below.

Fact 5 *Let $\bar{S}$ be a D-dimensional subspace of $L_2 \cap L_\infty(\mu)$ spanned by some basis $\{\phi_i : i \in \{1,\ldots,D\}\}$. Let $\|\cdot\|_2$ and $\|\cdot\|_\infty$ denotes the L_2-norm and* *the L_∞-norm with respect to μ. Let $|\cdot|_2$ and $|\cdot|_\infty$ denote the l_2-norm and the l_∞-norm in R^D.*

Suppose that there exist constants T_1 and T_2 such that for $(\theta_1, \ldots, \theta_D) \in R^D$,

$$\left\| \sum_{i=1}^{D} \theta_i \phi_i \right\|_\infty \leq T_1 |\theta|_\infty \tag{34}$$

and

$$\frac{T_2}{\sqrt{D}} |\theta|_2 \leq \left\| \sum_{i=1}^{D} \theta_i \phi_i \right\|_2 \leq \frac{T_3}{\sqrt{D}} |\theta|_2. \tag{35}$$

Take $r' \geq T_1/T_3$ and

$$B' = \sqrt{2\pi e} \left(0.5 + \max\left(\frac{T_3}{T_2}, 1 \right) \right) \tag{36}$$

Then for $\mathcal{B}$: an L_2 ball of radius σ in $\bar{S}$ with $0 < \delta < \sigma/5$, there exists a finite set $T \subset \mathcal{B}$ such that T is a δ-net for $\mathcal{B}$ with respect to the L_2-norm and a $r'\delta$-net with respect to the L_∞ norm, and the number of elements in T is at most $(B'\sigma/\delta)^D$.

Proof. Suppose that the center of $\mathcal{B}$ is $\sum_{i=1}^{D} \theta_i^* \phi_i$. Let $\theta^* = (\theta_1^*, \ldots, \theta_D^*)$. Then it follows from the first inequality in (35) that $\mathcal{B}$ is contained $\{\sum_{i=1}^{D} \theta_i \phi_i : (\theta_1, \ldots, \theta_D) \in B_0\}$, where B_0 is the l_2 ball of center θ^* and radius $\sqrt{D}\sigma/T_2$. Since the volume for an l_2 ball in R^D with radius σ is bounded by $c_{00}(D)\sigma^D$ (cf. Proof of Lemma 2 in [3]), where

$$c_{00}(D) = (2\pi e/D)^{D/2} (\pi D)^{-1/2},$$

we can cover B_0 with cubes of edge length δ/T_3 such that the number of cubes is at most

$$\frac{c_{00}(D)(\sqrt{D}\sigma/T_2 + \sqrt{D}\delta/T_3)^D}{(\delta/T_3)^D} \leq (1 + (T_3\sigma)/(T_2\delta))^D (2\pi e)^{D/2} \leq (B'\sigma/\delta)^D$$

for the B' in (36) if $\sigma/5 > \delta > 0$. Choosing one point from each cube to form a set T_0, and take $T = \mathcal{B} \cap \{\sum_{i=1}^{D} \theta_i \phi_i : (\theta_1, \ldots, \theta_D) \in T_0\}$, then from the second inequality in (35), T is a δ-net for $\mathcal{B}$ with respect to the L_2-norm. From (34), T is a $r'\delta$-net for $\mathcal{B}$ with respect to the L_∞-norm for $r' \geq T_1/T_3$. The proof of Fact 5 is complete.

Fact 6 *Suppose that μ is the Lebesgue measure on $[0, 1]$. Let $\bar{S}$ be the space of B-splines on $[0, 1]$ with order q and k knots $\xi_1, \ldots, \xi_k$ with multiplicities $m_1, \ldots, m_k$, where $0 < \xi_1 < \cdots < \xi_k < 1$. Then $\bar{S}$ is a sub-space of $L_2 \cap L_\infty(\mu)$. Suppose that $0 < \tilde{\Delta}_1 \leq \xi_i - \xi_{i-1} \leq \tilde{\Delta}_2$ for $i = 1, \ldots, k+1$,* *where $\xi_0 = 0$ and $\xi_{k+1} = 1$. Let $K = m_1 + \cdots + m_k$ and $D = K + q$. Then (34) holds with $T_1 = 1$ and (35) holds with $T_2 = \sqrt{\tilde{\Delta}_1 D}/(\sqrt{q}(2q+1)9^{q-1})$ and $T_3 = q\sqrt{\tilde{\Delta}_2 D}$.*

Proof. Let

$$(y_1, \ldots, y_{K+2q}) = (\underbrace{0, \ldots, 0}_{q \text{ times}}\ \underbrace{\xi_1, \ldots, \xi_1}_{m_1 \text{ times}}, \ldots, \underbrace{\xi_k, \ldots, \xi_k}_{m_k \text{ times}}\ \underbrace{1, \ldots, 1}_{q \text{ times}})$$

and let ϕ_i be the (normalized) B-spline basis of order q associated with knots $y_i, \ldots, y_{i+q}$ for $i = 1, \ldots, D$. Then $\phi_1, \ldots, \phi_D$ spans $\bar{S}$. It follows from Equation (4.80) in Schumaker [6] that (34) holds with $T_1 = 1$, so it remains to check (35).

To check that the first inequality in (35) holds with the T_2 specified above, note that from (4.79) and (4.86) in [6], we have that for $f = \sum_{i=1}^{K+q} \theta_i \phi_i$,

$$|\theta_i| \leq (2q+1)^2 9^{2(q-1)} \tilde{\Delta}_1^{-1/2} \|f\|_{L_2[y_i, y_{i+q}]},$$

where ϕ_i is supported on $[y_i, y_{i+q}]$ which implies that

$$\begin{aligned} \sum_{i=1}^{K+q} \theta_i^2 &\leq (2q+1)^2 9^{2(q-1)} \tilde{\Delta}_1^{-1} \sum_{i=1}^{K+q} \|f\|_{L_2[y_i, y_{i+q}]}^2 \\ &\leq (2q+1)^2 9^{2(q-1)} \tilde{\Delta}_1^{-1} q \|f\|_2^2, \end{aligned}$$

which implies that the first inequality in (35) holds with $T_2 = \sqrt{\tilde{\Delta}_1 D}/(\sqrt{q}(2q+1)9^{q-1})$.

To check that the first inequality in (35) holds with the T_3 specified above, we follow the approach in the proof of Lemma 4.2 in Ghosal, Ghosh and Van der Vaart [4], which is originally given in Stone [7]. For $f = \sum_{j=1}^{K+q} \theta_j \phi_j$, we have that for $x \in [y_i, y_{i+1})$ and $q+1 \leq i \leq q+K$, $f(x) = \sum_{j=i+1-q}^{i} \theta_j \phi_j(x)$ (cf. [6], Equations (4.25) and (4.29)), so it follows from Schwartz inequality that for $x \in [y_i, y_{i+1})$,

$$f^2(x) \leq q \sum_{j=i+1-q}^{i} \theta_j^2 \phi_j^2(x) \leq q \sum_{j=i+1-q}^{i} \theta_j^2,$$

which gives

$$\int_0^1 f^2(x)dx = \sum_{i=q+1}^{q+K} \int_{y_i}^{y_{i+1}} f^2(x)dx \leq q \sum_{i=q+1}^{q+K} \sum_{j=i+1-q}^{i} \theta_j^2 (y_{i+1} - y_i) \leq q^2 \tilde{\Delta}_2 \sum_{j=1}^{D} \theta_j^2$$

and the second inequality in (35) holds with $T_3 = q\sqrt{\tilde{\Delta}_2 D}$. The proof of Fact 6 is complete.

From Facts 5 and 6, Assumption M2 holds with $D_j = q + k$, B'_j defined in (12) and $r_j = 1$ (defined in (13)) since

$$1 \geq \frac{1}{q\sqrt{\tilde{\Delta}_{2,j}(k+q)}},$$

where $\tilde{\Delta}_{2,j}$ and $\tilde{\Delta}_{2,j}$ are defined in (10) and (11) respectively. It is clear that $B'_j \geq 1$ and $D_j \geq 1$ for $j \in \Lambda$, as required in Theorem 1. In addition, (14) and $B_j^2 r_j \leq \delta_n$ together implies that $r_j B_j^2 \leq \ell_n/4$ for $j \in \Lambda$ if n is large enough.

Assumption M3 holds with constants k_1 and k_2 that do not depend on j since the density for the distribution of X_i is supported on $[0,1]$ and is bounded below from zero and bounded above on $[0,1]$ and

$$E\left(\frac{1}{n}\sum_{i=1}^{n}(Y_i - t(X_i))^2 - \frac{1}{n}\sum_{i=1}^{n}(Y_i - s(X_i))^2\right) = E(s(X_i) - t(X_i))^2.$$

Next, we will verify (5). Note that with $B_j = 2b$ and $D_j = k + q$, (16) implies that

$$\frac{\ell_n \eta_{1,j}}{2r_j B_j^2} > \frac{a_n(k+q)\log(B'_j)}{2} \geq \frac{a_n D_j \log(3.5\sqrt{2\pi e})}{2}$$

and $24(1 + 2D_j \log 2) \leq 24(1 + 2\log 2)D_j$, so if n is large enough so that

$$a_n \log(3.5\sqrt{2\pi e}) \geq 48(1 + 2\log 2), \tag{37}$$

then (5) holds.

Now we have verified that the conditions in Theorem 1 hold true. Let $\eta_j = (\eta_{1,j} + r_j(2b)^2\xi/\ell_n)/2$. Then by Theorem 1, the error bound in (7) holds with $\theta = \left(\frac{A_0 + 2}{c(8\tau)}\right) \vee 5$ except on a set of probability at most

$$q_n\left(p^*_{j^*}(\ell_n, \eta_{j^*}) + \sum_{j\in\Lambda} p_j(\ell_n, \eta_j)\right), \tag{38}$$

for a given $j^* \in \Lambda$.

Below we will calculate the probability upper bound in (38) with

$$j^* = \left(b^*, q^*, \frac{1}{2^{J^*}}, \ldots, \frac{2^{J^*} - 1}{2^{J^*}}\right) \stackrel{\text{def}}{=} (b^*, q^*, \xi^*),$$

where $q^* = m+1$ and $J^* = \lfloor \log_2(n^{1/(1+2m)}) \rfloor$, and b^* is a constant large enough so that for any spline with knot vector ξ^*, order q^*, and sup-norm bounded by $\lfloor \|s\|_\infty \rfloor + 1$, the spline coefficients are bounded by b^*. Note that

$$\begin{aligned}
\eta_j &= \frac{1}{2}\left(\eta_{1,j} + \frac{r_j(2b)^2\xi}{\ell_n}\right) \\
&= \frac{1}{2}\left(\frac{a_n r_j(2b)^2}{\ell_n}\Big((k+q)\log(B'_j) + \lambda[(\log 2)2^{J_j} + q + b]\Big) + \frac{r_j(2b)^2\xi}{\ell_n}\right),
\end{aligned}$$

so

$$\begin{aligned}
p^*_{j^*}(\ell_n, \eta_{j^*}) &= \exp\left(-\frac{2c_1(\tau/k_0)\ell_n\eta_{j^*}}{(A_0+2b^*)^2}\right) \\
&\le \exp\Big(-a_n c_9[(\log 2)2^{J^*} + q^* + b^*] - c_9\xi\Big), \qquad (39)
\end{aligned}$$

and for $0 < \sigma < 1$,

$$\begin{aligned}
p_j(\ell_n, \eta_j) &= \exp\left(-\frac{2c_1(\tau/k_0)\ell_n\eta_j}{(A_0+2b)^2}\right) \\
&\quad +1.6\left(1 - \exp\left(-\frac{\tau\sigma^2}{2}\left[c(\tau) \wedge \frac{1}{3}\right]\right)\right)^{-1}\Big((B'_j\theta) \vee 2\Big)^{3(k+q)} \times \\
&\quad \left[\exp\left(-\frac{\ell_n c(\tau)\eta_j}{2(A_0+2b)(2b)}\right) + \exp\left(-\frac{\ell_n\eta_j}{24 r_j(2b)^2}\right)\right] \\
&\le \exp\left(-\frac{2c_3\ell_n\eta_j}{(2b)^2}\right) + \frac{c_4}{\sigma^2}\exp\left(-\frac{2c_5\ell_n\eta_j}{(2b)^2}\right)\Big(B'_j\theta\Big)^{3(k+q)}
\end{aligned}$$

for some constants c_9, c_3, c_4, c_5. For $0 < \sigma < 1$, if n is large enough so that

$$c_5 a_n \ge 3 + \frac{3\log\theta}{\log(3.5\sqrt{2\pi e})}, \qquad (40)$$

then for $\xi > 0$ and $c_6 = c_3 \wedge c_5$,

$$p_j(\ell_n, \eta_j) \le 2\left(1 \vee \frac{c_4}{\sigma^2}\right)\exp\Big(-c_6\left[a_n[(\log 2)2^{J_j} + q + b] + \xi\right]\Big).$$

For n large enough so that $a_n c_6 > 2$,

$$\begin{aligned}
&\sum_{j\in\Lambda} \exp\Big(-c_6 a_n[(\log 2)2^{J_j} + q + b]\Big) \\
&\le \sum_b \sum_q \sum_J \sum_{k=1}^{2^J-1} \frac{2^J-1}{k!(2^J-1-k)!}\exp\Big(-2[(\log 2)2^J + q + b]\Big) \\
&\le \sum_b \sum_q \sum_J 2^{-2^J} e^{-2b} e^{-2q} \stackrel{\text{def}}{=} c_7 < \infty,
\end{aligned}$$

so for $c_8 = 2c_7$,

$$\sum_{j\in\Lambda} p_j(\ell_n, \eta_j) \le c_8 \left(1 \vee \frac{c_4}{\sigma^2}\right) \exp\left(-c_6\xi\right).$$

Therefore, for $0 < \sigma < 1$ and for n large enough so that $a_n c_6 > 2$ and (40) and (37) hold, we have

$$(k_1 - 9\tau)\|\hat{s} - s\|^2 I_{\Omega_n} > 9\tau\left(\frac{\sigma^2}{4}\right) + 1.5\eta_{1,j^*} + (k_2+\tau)\|s^* - s\|^2 + \frac{1.5\delta_n\xi}{\ell_n}$$

with probability at most

$$q_n\left(\exp(-c_9\xi) + c_8\left(1 \vee \frac{c_4}{\sigma^2}\right)\exp\left(-c_6\xi\right)\right)$$

for $\xi > 0$. Let

$$U = (k_1 - 9\tau)\|\hat{s} - s\|^2 I_{\Omega_n} - \left(9\tau\left(\frac{\sigma^2}{4}\right) + 1.5\eta_{1,j^*} + (k_2+\tau)\|s^* - s\|^2\right),$$

then for $\xi_0 > 0$, $0 < \sigma < 1 \wedge \sqrt{c_4}$, $c_{10} = c_6 \wedge c_9$ and $c_{11} = c_4(1 + c_8)$, we have

$$\begin{aligned}
E\left(\frac{\ell_n U}{1.5\delta_n}\right) &\le \xi_0 + \int_{\xi_0}^{\infty} q_n\left[\exp(-c_9\xi) + c_8\left(1 \vee \frac{c_4}{\sigma^2}\right)\exp\left(-c_6\xi\right)\right] d\xi \\
&\le \xi_0 + \frac{c_{11}q_n}{\sigma^2}\int_{\xi_0}^{\infty} \exp\left(-c_{10}\xi\right) d\xi \\
&\le \xi_0 + \frac{c_{11}q_n}{c_{10}\sigma^2}\exp\left(-c_{10}\xi_0\right).
\end{aligned}$$

Take $\xi_0 = \delta_n^{-1}\ell_n n^{-2m/(1+2m)}$ and $\sigma = n^{-m/(1+2m)}$, then $\xi_0 = O(\delta_n^{-1}(\log n)^{-1-\gamma_0} n^{1/(1+2m)})$ and

$$\limsup_n n^{2m/(1+2m)} E(U) \le \limsup_n \frac{1.5\delta_n n^{2m/(1+2m)}}{\ell_n}\left[\xi_0 + \frac{c_{11}}{c_{10}\sigma^2}\exp\left(-c_{10}\xi_0\right)\right] < \infty,$$

so

$$(k_1 - 9\tau)E\|\hat{s} - s\|^2 I_{\Omega_n} \le 1.5\eta_{1,j^*} + (k_2+\tau)\|s^* - s\|^2 + Cn^{-2m/(1+2m)}$$

for some constant $C > 0$. Recall that for the splines in S_{j^*}, the knots are equally spaced and the number of knots is $2^{J^*} - 1 = O(n^{1/(1+2m)})$, so we can choose $s^* \in S_{j^*}$ such that $\|s^* - s\|^2 = O(n^{-2m/(1+2m)})$ (cf. Theorem 6.25 in [6]) and

$$(k_1 - 9\tau)E\|\hat{s} - s\|^2 I_{\Omega_n} \le 1.5\eta_{1,j^*} + O(n^{-2m/(1+2m)}). \tag{41}$$

Since $r_{j^*} = 1$ and

$$B'_{j^*} = \sqrt{2\pi e}\left(0.5 + q^*\sqrt{q^*}(2q^*+1)9^{q^*-1}\right),$$

we have

$$\begin{aligned}\eta_{1,j^*} &= \frac{a_n r_{j^*}(2b^*)^2}{\ell_n}\left((2^{J^*}-1+q^*)\log(B'_{j^*}) + \lambda[(\log 2)2^{J^*}+q^*+b^*]\right) \\ &= O(a_n(\log n)^{1+\gamma_0}n^{-2m/(1+2m)}).\end{aligned}$$

Choose $\tau < k_1/9$, then (41) implies that

$$E\|\hat{s}-s\|^2 I_{\Omega_n} = O(a_n(\log n)^{1+\gamma_0}n^{-2m/(1+2m)}). \tag{42}$$

It remains to establish an upper bound for $E\|\hat{s}-s\|^2 I_{\Omega_n^c}$. Note that $\|\hat{s}\| \le \delta_n$, $\|s\|_\infty < \infty$ and $P(\Omega_n^c) \le (\ell_n+1)(\beta_{\tilde{q}_n} + \beta_{q_n-\tilde{q}_n})$, where both $\tilde{q}_n$ and $q_n - \tilde{q}_n$ are $O((\log n)^{1+\gamma_0})$ from the choice that $\ell_n = O(n(\log n)^{-1-\gamma_0})$. Thus there exists a constant $c_{12} > 0$ such that

$$\begin{aligned}n^{2m/(1+2m)}E\|\hat{s}-s\|^2 I_{\Omega_n^c} &= O(n^{2m/(1+2m)}\delta_n^2(\ell_n+1)\exp(-c_{12}(\log n)^{1+\gamma_0}) \\ &= o(1). \end{aligned} \tag{43}$$

The proof of Theorem 2 is complete by combining (42) and (43).

References

[1] Y Baraud, F. Comte, and G. Viennet. Adaptive estimation in autoregression or β-mixing regression via model selection. *The Annals of Statistics*, 29(3):839–875, 2001.

[2] Andrew Barron, Lucien Birgé, and Pascal Massart. Risk bounds for model selection via penalization (Disc: P415-419). *Probability Theory and Related Fields*, 113:301–413, 1999.

[3] Lucien Birgé and Pascal Massart. Minimum contrast estimators on sieves: Exponential bounds and rates of convergence. *Bernoulli*, 4:329–375, 1998.

[4] Subhashis Ghosal, Jayanta K. Ghosh, and Aad W. van der Vaart. Convergence rates of posterior distributions. *The Annals of Statistics*, 28(2):500–531, 2000.

[5] Tzee-Ming Huang. Convergence rates for posterior distributions and adaptive estimation. *The Annals of Statistics*, 32(4):1556–1593, 2004.

[6] Larry L. Schumaker. *Spline Functions: Basic Theory.* Wiley-Interscience, 1981.

[7] Charles J. Stone. The dimensionality reduction principle for generalized additive models. *The Annals of Statistics*, 14:590–606, 1986.

[8] Yuhong Yang and Andrew R. Barron. An asymptotic property of model selection criteria. *IEEE Transactions on Information Theory*, 44(1):95–116, 1998.

[9] S. Zhou, X. Shen, and D. A. Wolfe. Local asymptotics for regression splines and confidence regions. *The Annals of Statistics*, 26(5):1760–1782, 1998.

科技部補助計畫衍生研發成果推廣資料表

日期:2014/07/28

<table>
<tr><td rowspan="3">科技部補助計畫</td><td>計畫名稱: 相關誤差下的迴歸函數適應性估計</td></tr>
<tr><td>計畫主持人: 黃子銘</td></tr>
<tr><td>計畫編號: 102-2118-M-004-007-　　學門領域: 數理統計與機率</td></tr>
<tr><td colspan="2">無研發成果推廣資料</td></tr>
</table>

102 年度專題研究計畫研究成果彙整表

<table>
<tr><td colspan="3">計畫主持人：黃子銘</td><td colspan="5">計畫編號：102-2118-M-004-007-</td></tr>
<tr><td colspan="8">計畫名稱：相關誤差下的迴歸函數適應性估計</td></tr>
<tr><td colspan="3" rowspan="2">成果項目</td><td colspan="3">量化</td><td rowspan="2">單位</td><td rowspan="2">備註（質化說明：如數個計畫共同成果、成果列為該期刊之封面故事...等）</td></tr>
<tr><td>實際已達成數（被接受或已發表）</td><td>預期總達成數(含實際已達成數)</td><td>本計畫實際貢獻百分比</td></tr>
<tr><td rowspan="12">國內</td><td rowspan="4">論文著作</td><td>期刊論文</td><td>0</td><td>0</td><td>100%</td><td rowspan="3">篇</td><td></td></tr>
<tr><td>研究報告/技術報告</td><td>0</td><td>0</td><td>100%</td><td></td></tr>
<tr><td>研討會論文</td><td>0</td><td>0</td><td>100%</td><td></td></tr>
<tr><td>專書</td><td>0</td><td>0</td><td>100%</td><td></td><td></td></tr>
<tr><td rowspan="2">專利</td><td>申請中件數</td><td>0</td><td>0</td><td>100%</td><td rowspan="2">件</td><td></td></tr>
<tr><td>已獲得件數</td><td>0</td><td>0</td><td>100%</td><td></td></tr>
<tr><td rowspan="2">技術移轉</td><td>件數</td><td>0</td><td>0</td><td>100%</td><td>件</td><td></td></tr>
<tr><td>權利金</td><td>0</td><td>0</td><td>100%</td><td>千元</td><td></td></tr>
<tr><td rowspan="4">參與計畫人力（本國籍）</td><td>碩士生</td><td>0</td><td>0</td><td>100%</td><td rowspan="4">人次</td><td></td></tr>
<tr><td>博士生</td><td>0</td><td>0</td><td>100%</td><td></td></tr>
<tr><td>博士後研究員</td><td>0</td><td>0</td><td>100%</td><td></td></tr>
<tr><td>專任助理</td><td>0</td><td>0</td><td>100%</td><td></td></tr>
<tr><td rowspan="12">國外</td><td rowspan="4">論文著作</td><td>期刊論文</td><td>0</td><td>0</td><td>100%</td><td rowspan="3">篇</td><td></td></tr>
<tr><td>研究報告/技術報告</td><td>0</td><td>0</td><td>100%</td><td></td></tr>
<tr><td>研討會論文</td><td>1</td><td>1</td><td>100%</td><td>論文名稱：’’An adaptive knot selection method for regression splines via penalized minimum contrast estimation’’. 發表於 2014 IMS APRM 國際研討會.</td></tr>
<tr><td>專書</td><td>0</td><td>0</td><td>100%</td><td>章/本</td><td></td></tr>
<tr><td rowspan="2">專利</td><td>申請中件數</td><td>0</td><td>0</td><td>100%</td><td rowspan="2">件</td><td></td></tr>
<tr><td>已獲得件數</td><td>0</td><td>0</td><td>100%</td><td></td></tr>
<tr><td rowspan="2">技術移轉</td><td>件數</td><td>0</td><td>0</td><td>100%</td><td>件</td><td></td></tr>
<tr><td>權利金</td><td>0</td><td>0</td><td>100%</td><td>千元</td><td></td></tr>
<tr><td rowspan="4">參與計畫人力（外國籍）</td><td>碩士生</td><td>0</td><td>0</td><td>100%</td><td rowspan="4">人次</td><td></td></tr>
<tr><td>博士生</td><td>0</td><td>0</td><td>100%</td><td></td></tr>
<tr><td>博士後研究員</td><td>0</td><td>0</td><td>100%</td><td></td></tr>
<tr><td>專任助理</td><td>0</td><td>0</td><td>100%</td><td></td></tr>
</table>

其他成果 (無法以量化表達之成果如辦理學術活動、獲得獎項、重要國際合作、研究成果國際影響力及其他協助產業技術發展之具體效益事項等，請以文字敘述填列。)	無

	成果項目	量化	名稱或內容性質簡述
科教處計畫加填項目	測驗工具(含質性與量性)	0	
	課程/模組	0	
	電腦及網路系統或工具	0	
	教材	0	
	舉辦之活動/競賽	0	
	研討會/工作坊	0	
	電子報、網站	0	
	計畫成果推廣之參與（閱聽）人數	0	

科技部補助專題研究計畫成果報告自評表

請就研究內容與原計畫相符程度、達成預期目標情況、研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）、是否適合在學術期刊發表或申請專利、主要發現或其他有關價值等，作一綜合評估。

1. 請就研究內容與原計畫相符程度、達成預期目標情況作一綜合評估 ■達成目標 □未達成目標（請說明，以 100 字為限） □實驗失敗 □因故實驗中斷 □其他原因 說明：
2. 研究成果在學術期刊發表或申請專利等情形： 論文：□已發表 □未發表之文稿 □撰寫中 ■無 專利：□已獲得 □申請中 ■無 技轉：□已技轉 □洽談中 ■無 其他：（以 100 字為限） 發表於 2014 IMS APRM 國際研討會. 論文名稱: ’’An adaptive knot selection method for regression splines via penalized minimum contrast estimation’’
3. 請依學術成就、技術創新、社會影響等方面，評估研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）（以 500 字為限） 目前已針對解釋變數觀察值為時間數列的相依資料, 提出適應性估計量. 未來可以考慮針對解釋變數觀察值為 random field 的資料, 提出適應性估計量.