Biometrika (20XX), **XX**, 1, *pp*. 1–20

Printed in Great Britain

Supplementary Material for Partially Functional Linear Regression in High Dimensions

By Dehan Kong

Department of Biostatistics, University of North Carolina, Chapel Hill, North Carolina 27599, U.S.A.

kongdehanstat@gmail.com

Kaijie Xue and Fang Yao

Department of Statistical Sciences, University of Toronto, Ontario M5S 3G3, Canada

kaijie@utstat.toronto.edu fyao@utstat.toronto.edu

Hao H. Zhang

Department of Mathematics, University of Arizona, 617 North Santa Rita Avenue, Tucson, Arizona 85721, U.S.A.

hzhang@math.arizona.edu

1. Additional Simulation Results

We present additional simulation results based on 200 Monte Carlo runs. In particular, Design III illustrates the situation that the response does not necessarily depend on the leading principal components and the regression coefficients may decay more slowly than theoretically required. Specifically, we use $b_{\ 1} = 0$, $b_{\ 2} = 1$, $b_{\ 3} = 0{\cdot}5$, $b_{\ \ } = (k-2)^{-3}$ for $k = 4, \ldots, 50$, $j = 1, 2$, and other settings are the same as Design I. The results across s_n follow a similar pattern, while the automated ABIC captures the regression relationship adaptively and resembles the optimal estimation and prediction. The refitting step using least squares with jointly tuned s_n behaves similarly as in Design I. Designs IV contains ultra-high numbers of scalar covariates $\gamma = (1_5^{\mathrm{T}}, 0_{995}^{\mathrm{T}})^{\mathrm{T}}$ with $p_n = 1000$, and other settings the same as Design II. The results exhibit similar phenomenon as those in Design II. Moreover, Table 2 includes the results obtained from the same settings as Design I–IV except for a larger sample size $n = 400$. For Table 3, the only setting difference is that the regression error ϵ is generated from $N(0, 2)$ instead of $N(0, 1)$. As expected, a larger sample size reduced the estimation and prediction errors, while a higher noise level increased such errors.

2. Regularity Conditions

Without loss of generality, we assume that $\{X\ , j = 1, \ldots, d\}$, Y and Z have been centred to have mean zero. With $W\ \ = x\ (t\ \ ) + \epsilon\ $, for definiteness, we consider the local linear smoother for each set of subjects using bandwidths $\{h\ \ , j = 1, \ldots, d\}$, and denote the smoothed trajectories by $\hat{x}\ $. Denote the minimum and maximum eigenvalues of a symmetric matrix A by $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$. Recall that the first g functional predictors are significant, while the rest are not. Define the $(gs_n + q_n) \times 1$ vector $\tilde{N} =$

Design	s_n	FZ$_f$	FN$_f$	MSE$_f$	FZ	FN	MSE	PE
	1	2·0	0	2·5 (0)	1·5	1·1	4·2 (0·15)	26·0 (0·1)
	2	0·70	0	1·7 (0·06)	0·02	1·2	0·49 (0·028)	3·9 (0·1)
	3	0	0	0·076 (0·004)	0	0·38	0·069 (0·004)	0·21 (0·009)
	4	0	0	0·069 (0·003)	0	0·38	0·065 (0·004)	0·13 (0·005)
	5	0	0	0·11 (0·005)	0	0·40	0·067 (0·004)	0·14 (0·005)
III	6	0	0	0·15 (0·006)	0	0·41	0·068 (0·004)	0·15 (0·005)
$p_n = 20$	10	0	0·01	0·60 (0·02)	0	0·37	0·071 (0·004)	0·19 (0·006)
	16	0·17	0·21	3·4 (0·1)	0	0·05	0·089 (0·007)	0·65 (0·05)
	ABIC			$\hat{s}_n$=3·73 (0·048)				
		0	0	0·072 (0·004)	0	0·38	0·066 (0·004)	0·14 (0·005)
	TUNE s_n			$\hat{s}_{n1}$=3·99 (0·083), $\hat{s}_{n2}$=3·98 (0·079)				
		0	0	0·056 (0·003)	0	0·32	0·063 (0·004)	0·13 (0·005)
				$\hat{s}_n$=3·50 (0·040)				
	STEP 1	0	0·09	0·10 (0·003)	0	6·5	0·41 (0·02)	0·81 (0·021)
IV	STEP 2	0	0	0·059 (0·003)	0	0·13	0·048 (0·003)	0·13 (0·005)
$p_n = 1000$				$\hat{s}_{n1}$=4·05 (0·067), $\hat{s}_{n2}$ =4·11 (0·065)				
	TUNE s_n	0	0	0·042 (0·002)	0	0·09	0·045 (0·003)	0·11 (0·004)

Table 1: Simulation results with sample size $n = 200$ based on 200 Monte Carlo replicates for Designs III and IV. Shown are the Monte Carlo averages (standard errors in parentheses) for the number of false zero functional predictors (FZ$_f$), the number of the false nonzero functional predictors (FN$_f$), the functional mean squared error (MSE$_f$), the number of false zero scalar covariates (FZ), the number of false nonzero scalar covariates (FN), the scalar mean squared error (MSE), and the prediction error (PE). We first use ABIC to choose the tuning parameter λ_n and a common truncation s_n, then tune s_n jointly with AIC by refitting the selected model using ordinary least squares. In Design IV, Step 1 results are based on the original sample in each Monte Carlo r

Design	s_n	FZ$_f$	FN$_f$	MSE$_f$	FZ	FN	MSE	PE
	1	0·90	0	4·1 (0·04)	0·15	5·1	2·3 (0·08)	25·8 (0·2)
	2	0	0	1·3 (0·009)	0	1·4	0·34 (0·01)	4·6 (0·04)
	3	0	0	0·58 (0·008)	0	0·23	0·11 (0·004)	1·5 (0·02)
	4	0	0	0·063 (0·002)	0	0·23	0·027 (0·002)	0·13 (0·004)
	5	0	0	0·071 (0·003)	0	0·27	0·028 (0·002)	0·12 (0·004)
I	6	0	0	0·095 (0·003)	0	0·24	0·028 (0·002)	0·12 (0·004)
$p_n = 20$	10	0	0	0·32 (0·01)	0	0·31	0·029 (0·002)	0·13 (0·004)
	16	0	0	1·2 (0·04)	0	0·10	0·026 (0·002)	0·15 (0·004)
	ABIC	$\hat{s}_n$=4·22 (0·047)						
		0	0	0·067 (0·003)	0	0·23	0·027 (0·002)	0·13 (0·004)
	TUNE s_n	$\hat{s}_{n1}$=4·67 (0·058), $\hat{s}_{n2}$=4·68 (0·061)						
		0	0	0·049 (0·002)	0	0·20	0·026 (0·002)	0·12 (0·004)
		$\hat{s}_n$=4·10 (0·027)						
	STEP 1	0	0	0·11 (0·004)	0	4·6	0·13(0·007)	0·51 (0·01)
II	STEP 2	0	0	0·052 (0·002)	0	0·07	0·025 (0·001)	0·11 (0·004)
$p_n = 1000$		$\hat{s}_{n1}$=4·72 (0·056), $\hat{s}_{n2}$=4·60 (0·053)						
	TUNE s_n	0	0	0·041 (0·001)	0	0·06	0·025 (0·001)	0·11 (0·004)
	1	2·0	0	2·5 (0)	1·0	0·43	2·7 (0·1)	25·2 (0·09)
	2	0·10	0	0·72 (0·04)	0	0·43	0·19 (0·01)	2·8 (0·06)
	3	0	0	0·059 (0·002)	0	0·25	0·032 (0·002)	0·16 (0·006)
	4	0	0	0·033 (0·001)	0	0·23	0·026 (0·001)	0·073 (0·004)
	5	0	0	0·048 (0·002)	0	0·26	0·027 (0·002)	0·075 (0·004)
III	6	0	0	0·072 (0·003)	0	0·25	0·027 (0·002)	0·078 (0·004)
$p_n = 20$	10	0	0	0·28 (0·01)	0	0·25	0·028 (0·002)	0·099 (0·004)
	16	0	0·03	1·2 (0·04)	0	0·02	0·024 (0·001)	0·13 (0·004)
	ABIC	$\hat{s}_n$=3·91 (0·036)						
		0	0	0·034 (0·002)	0	0·23	0·026 (0·001)	0·075 (0·004)
	TUNE s_n	$\hat{s}_{n1}$=4·37 (0·064), $\hat{s}_{n2}$=4·38 (0·066)						
		0	0	0·026 (0·001)	0	0·22	0·026 (·001)	0·070 (0·004)
		$\hat{s}_n$=3·81 (0·043)						
	STEP 1	0	0·01	0·047 (0·001)	0	3·8	0·17 (0·009)	0·35 (0·008)
IV	STEP 2	0	0	0·031 (0·002)	0	0·08	0·024 (0·001)	0·073 (0·004)
$p_n = 1000$		$\hat{s}_{n1}$=4·28 (0·061), $\hat{s}_{n2}$=4·37 (0·067)						
	TUNE s_n	0	0	0·023 (0·001)	0	0·07	0·024 (0·001)	0·058 (0·004)

Table 2: Simulation results using the same settings as Design I & II in Section 4 of the paper and

Design	s_n	FZ$_f$	FN$_f$	MSE$_f$	FZ	FN	MSE	PE
I	1	0·94	0	4·1 (0·03)	0·41	7·3	4·9 (0·2)	28·0 (0·6)
$p_n = 20$	2	0·56	0	2·8 (0·1)	0·08	3·3	1·4 (0·07)	9·9 (0·3)
	3	0·01	0	0·63 (0·02)	0	1·7	0·38 (0·03)	1·7 (0·07)
	4	0	0	0·17 (0·007)	0	0·58	0·15 (0·009)	0·31 (0·01)
	5	0	0	0·23 (0·1)	0	0·58	0·15 (0·009)	0·30 (0·01)
	6	0	0	0·33 (0·01)	0	0·63	0·15 (0·009)	0·32 (0·01)
	10	0·01	0·01	1·3 (0·05)	0	0·49	0·15 (0·009)	0·44 (0·05)
	16	0·31	0·18	7·7 (0·3)	0·08	0·20	0·36 (0·03)	3·6 (0·3)
	ABIC			$\hat s_n$=4·12 (0·026)				
		0	0	0·18 (0·009)	0	0·56	0·15 (0·009)	0·30 (0·01)
	TUNE s_n			$\hat s_{n1}$=4·59 (0·069), $\hat s_{n2}$=4·61 (0·060)				
		0	0	0·14 (0·006)	0	0·48	0·14 (0·008)	0·28 (0·01)
II				$\hat s_n$=4·04 (0·021)				
$p_n = 1000$	STEP 1	0	0·04	0·29 (0·008)	0	7·3	0·55 (0·03)	1·9 (0·07)
	STEP 2	0	0	0·15 (0·008)	0	0·16	0·096 (0·005)	0·26 (0·02)
				$\hat s_{n1}$=4·53 (0·056), $\hat s_{n2}$=4·46 (0·050)				
	TUNE s_n	0	0	0·12 (0·006)	0	0·15	0·095 (0·005)	0·24 (0·008)
III	1	2·0	0	2·5 (0)	1·5	1·2	4·3 (0·15)	26·1 (0·1)
$p_n = 20$	2	0·33	0	1·1 (0·06)	0·02	0·69	0·43 (0·03)	3·3 (0·1)
	3	0	0	0·10 (0·005)	0	0·39	0·13 (0·008)	0·29 (0·01)
	4	0	0	0·12 (0·006)	0	0·39	0·13 (0·008)	0·22 (0·01)
	5	0	0	0·20 (0·009)	0	0·41	0·13 (0·008)	0·24 (0·01)
	6	0	0	0·29 (0·01)	0	0·41	0·14 (0·008)	0·26 (0·01)
	10	0·03	0·07	1·4 (0·06)	0	0·17	0·12 (0·007)	0·41 (0·03)
	16	0·02	1·3	11·8(0·4)	0·02	0·39	0·18 (0·02)	1·0 (0·03)
	ABIC			$\hat s_n$=3·76 (0·065)				
		0	0	0·15 (0·02)	0	0·36	0·13 (0·008)	0·24 (0·01)
	TUNE s_n			$\hat s_{n1}$=3·76 (0·064), $\hat s_{n2}$=3·79 (0·066)				
		0	0	0·070 (0·004)	0	0·29	0·12 (0·007)	0·20 (0·01)
IV				$\hat s_n$=3·33 (0·040)				
$p_n = 1000$	STEP 1	0	0·06	0·18 (0·006)	0	6·4	0·63 (0·03)	1·5 (0·04)
	STEP 2	0	0	0·091 (0·005)	0	0·12	0·099 (0·005)	0·23 (0·009)
				$\hat s_{n1}$=3·96 (0·060), $\hat s_{n2}$=3·73 (0·048)				
	TUNE s_n	0	0	0·065 (0·003)	0	0·10	0·095 (0·005)	0·18 (0·009)

Table 3: Simulation results using the same settings as Design I & II in Section 4 of the paper and Deisgn III & IV as above, based on 200 Monte Carlo replicates, except the variance of the regression error $\sigma^2 = 2$.

and $0 < c_2 \leq \lambda_{\min}(U_1) \leq \lambda_{\max}(U_1) \leq c_3 < \infty$ for all n, where $\Sigma = E(ZZ^{\mathrm{T}})$ and $U_1 = E(\tilde N \tilde N^{\mathrm{T}})$.

3. AUXILIARY LEMMAS

For each $j = 1, \ldots, d$, given the estimated covariances $\hat K\ (s,t) = n^{-1}\sum_{=1}^{n} \hat x\ (s)\hat x\ (t)$, the eigenvalues/functions and functional principal component scores are estimated by

$$\int \hat K\ (s,t)\hat\phi\ (s)ds = \hat\lambda\ \hat\phi\ (t), \qquad \hat\xi\ = \int \hat x\ (t)\hat\phi\ (t)dt, \tag{1}$$

subject to $\int \hat\phi^2\ (t)dt = 1$ and $\int \hat\phi\ _1(t)\hat\phi\ _2(t)dt = 0$ for $k_1 \neq k_2$. Denote $\hat\Delta^2 = |||\hat K\ - K\ |||^2 = \int\int \{\hat K\ (s,t) - K\ (s,t)\}^2 ds dt, \quad \Re\ (s,t) = n^{1/2}\{\hat K\ (s,t) - K\ (s,t)\}$,

$(x \otimes x)(s,t) = x(s)x(t)$, the second derivative of x by $x^{(2)}$, the minimum and maximum eigenvalues of a symmetric matrix A by $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$. Let the $n \times p_n$ matrix $Z_M = (z_1, \ldots, z_n)^{\mathrm{T}} = (Z_M^{(1)}, Z_M^{(2)})$ with $Z_M^{(1)}$ containing the first q_n columns of Z_M. Define M as the $n \times s_n$ matrix with (i,k)th element ξ, and $M = (M_1, \ldots, M_d)$, $M^{(1)} = (M_1, \ldots, M)$ and $M^{(2)} = (M_{+1}, \ldots, M_d)$. Similarly, we define $\tilde{M}$, $\tilde{M}^{(1)}$ and $\tilde{M}^{(2)}$, where ξ is replaced by $\xi\, w^{-1/2}$. Moreover, we denote $\hat{M}$ as the $n \times s_n$ matrix with (i,k)th element $\hat{\xi}$, and $\hat{M} = (\hat{M}_1, \ldots, \hat{M}_d)$, $\hat{M}^{(1)} = (\hat{M}_1, \ldots, \hat{M})$ and $\hat{M}^{(2)} = (\hat{M}_{+1}, \ldots, \hat{M}_d)$. Similarly, we define $\check{M}$, $\check{M}^{(1)}$ and $\check{M}^{(2)}$, where $\hat{\xi}$ is replaced by $\hat{\xi}\, w^{-1/2}$. Combining the functional and scalar covariates, we have $\tilde{N}_1 = (\tilde{M}^{(1)}, Z_M^{(1)}) = (\tilde{N}_1, \ldots, \tilde{N}_n)^{\mathrm{T}}$, where $\{\tilde{N} = (\tilde{N}_1, \ldots, \tilde{N}_{n+q_n})^{\mathrm{T}}, i = 1, \ldots, n\}$ are independently and identically distributed as $\tilde{N}$. We further denote $N_1 = (M^{(1)}, Z_M^{(1)}) = (N_1, \ldots, N_n)^{\mathrm{T}}$, $\check{N}^{(1)} = (\check{M}^{(1)}, Z_M^{(1)})$, $\hat{N}^{(1)} = (\hat{M}^{(1)}, Z_M^{(1)})$, and let $E(\tilde{N}\,\tilde{N}^{\mathrm{T}}) = U_1$. Recall that $\check{b}^{(1)}$ denotes the estimate of $\tilde{b}^{(1)}$, $\check{\eta} = (\check{b}^{(1)\mathrm{T}}, \hat{\gamma}^{\mathrm{T}})^{\mathrm{T}}$, and we further denote $\check{b}^{(1)}$ the estimate of $\tilde{b}^{(1)}$. Suppose $\alpha(\cdot)$, $\beta(\cdot)$ and $G(\cdot,\cdot)$ are square-integrable functions on T and $T \times T$. Write $||\alpha||$, $\int \alpha\beta$ (or $\langle \alpha, \beta \rangle$) and $\int G\alpha\beta$ for $\{\int \alpha^2(t)dt\}^{1/2}$, $\int \alpha(t)\beta(t)dt$ and $\int\int_{2} G(s,t)\alpha(s)\beta(t)dsdt$.

Lemma 1 provides results for the estimates obtained by functional principal component analysis. We first quantify the smoothing error of $\hat{x}$ and its influence carried over to the covariance and functional principal component estimates, $\hat{\xi} - \xi = \tilde{\xi} - \xi + \hat{\xi} - \tilde{\xi} = \int x(\hat{\phi} - \phi) + \int(\hat{x} - x)\hat{\phi}$. Then we obtain upper bounds for the differences between the estimated and true eigenfunctions in terms of covariance perturbation and eigenvalue spacings for quantifying the increased estimation error of the higher order terms. For each $j = 1, \ldots, d$, denote

$$\delta = \min_{=1} (w - w_{+1}), \quad \mathcal{J}_n = \{k = 1, \ldots, \infty : w - w_{+1} > 2\hat{\Delta}\},$$

where $\hat{\Delta} = |||\hat{K} - K|||$, that is, consider $k \in \mathcal{J}_n$ for which the distance of w to the nearest other eigenvalues does not fall below $2\hat{\Delta}$. It is known that ϕ can be consistently estimated for $k \in \mathcal{J}_n$ (Theorem 1, Hall & Hosseini-Nasab, 2006), $\|\hat{\phi} - \phi\| = O_p(\delta\ \hat{\Delta}) = O_p(k^{+1} n^{-1/2})$, implying $\sup \mathcal{J}_n = o\{n^{1/(2+2)}\}$. However, for our theoretical analysis, we need a sharper bound without compromising the number of eigenfunctions considered. Define the set of realizations such that, for sample size n, some C and any $\tau < 1$,

$$\mathcal{F}_n = \{(\hat{w}_1 - w_2)^{-2} \leq 2(w_1 - w_2)^{-2} \leq Cn, k_1, k_2 = 1, \ldots, s_n, k_1 \neq k_2\}.$$

LEMMA 1*(a) Under conditions (B1)–(B4), for each* $i = 1, \ldots, n$ *and each* $j = 1, \ldots, d$, *we have* $E(\|\hat{x} - x\|^2) = o(n^{-1})$, $E\left\{\int (\hat{x} - x)^4\right\} = o(n^{-2})$. *For each* $j = 1, \ldots, d$, *we have* $E(|||\hat{K} - K|||^2) = O(n^{-1})$.

(b) Under conditions

where $||\alpha \quad || = O_p(k^{\ +2}n^{-1})$, $\Re \ = n^{1/2}(\hat{K} \ - K \ )$, *and* $O_p(\cdot)$ *is uniform in* $k = 1, \ldots, s_n$.

(d) *Under conditions (A1), (B1)–(B4) and* $n^{-1/2}s_n = o_p(1)$, *for any* $j = 1, \ldots, d$, *we have* $c_1\lambda_n \leq \lambda \ _n \leq c_2\lambda_n$ *for some positive constants* c_1, c_2.

The next lemma quantifies the asymptotic orders of several important types of expressions that will be encountered in the proofs of our lemmas and main theorems. For convenience, define the following notations, $l, l_1, l_2 = 1, \ldots, p_n$, $k, k_1, k_2 = 1, \ldots, s_n$, $j, j_1, j_2 = 1, \ldots, d$,

$$\theta^{(1)} = \sum_{=1}^{n}(\hat{\xi} \quad - \xi \quad)^2 w^{-1}, \qquad \theta_{\ 1\ 2}^{\ 1\ 2\ (2)} = n^{-1}\sum_{=1}^{n}(\hat{\xi} \ _{1\ 1}\hat{\xi} \ _{2\ 2} - \xi \ _{1\ 1}\xi \ _{2\ 2})(w \ _{1\ 1}w \ _{2\ 2})^{-1/2},$$

$$\theta_{\ 1\ 2}^{\ 1\ 2\ (3)} = n^{-1}\sum_{=1}^{n}\{\xi \ _{1\ 1}\xi \ _{2\ 2} - E(\xi \ _{1\ 1}\xi \ _{2\ 2})\}(w \ _{1\ 1}w \ _{2\ 2})^{-1/2}, \quad \theta_{\ 1\ 2}^{\ 1\ 2\ (4)} = \theta_{\ 1\ 2}^{\ 1\ 2\ (2)} + \theta_{\ 1\ 2}^{\ 1\ 2\ (3)},$$

$$\theta^{(5)} = n^{-1}\sum_{=1}^{n}(\hat{\xi} \quad - \xi \quad)z \ \ w^{-1/2}, \qquad \theta^{(6)} = n^{-1}\sum_{=1}^{n}\{\xi \quad z \quad - E(\xi \quad z \ \ )\}w^{-1/2},$$

$$\theta^{(7)} = \theta^{(5)} + \theta^{(6)}, \qquad \theta_{\ 1\ 2}^{(8)} = n^{-1}\sum_{=1}^{n}\{z \ _1 z \ _2 - E(z \ _1 z \ _2)\},$$

$$\vartheta^{(1)} = \sum_{=1}^{n}\sum_{=\ n+1}^{\infty}\xi \quad b \quad _0 z \ \ , \qquad \vartheta^{(2)} = \sum_{=1}^{n}\sum_{=1}^{n}(\hat{\xi} \quad - \xi \quad)(\hat{b} \quad - b \ \ _0)z \ \ ,$$

$$\vartheta^{(3)} = \sum_{=1}^{n}\sum_{=1}^{n}(\hat{\xi} \quad - \tilde{\xi} \quad)b \ \ _0 z \ \ , \qquad \vartheta^{(4)} = \sum_{=1}^{n}\sum_{=1}^{n}(\tilde{\xi} \quad - \xi \quad)b \ \ _0 z \ \ ,$$

$$\vartheta^{(5)} = \sum_{=1}^{n} b \ \ _0 \int (\sum_{=1}^{n} x \quad z \quad - n\Xi \ \ )(\hat{\phi} \quad - \phi \ \ ), \quad \vartheta^{(6)} = n\sum_{=1}^{n} b \ \ _0\langle \Xi \ \ , \alpha \ \ \rangle,$$

$$\vartheta^{(7)} = n^{1/2}\sum_{=1}^{n}\sum_{\neq} b \ \ _0(w \quad - w \quad)^{-1}\langle \Xi \ \ , \phi \ \ \rangle \int (\Re \ - \Re^*)\phi \quad \phi \quad ,$$

$$\vartheta^{(8)} = n^{1/2}\sum_{=1}^{n}\sum_{\neq} b \ \ _0(w \quad - w \quad)^{-1}\langle \Xi \ \ , \phi \ \ \rangle \int \Re^*\phi \quad \phi \quad ,$$

where $\vartheta^{\ (\ \bullet)} = (\vartheta_1^{(\ \bullet)}, \ldots \vartheta_{q_n}^{(\ \bullet)})^{\mathrm{T}}$ denote corresponding vectors and $\Re^* = n^{1/2}(\tilde{K} \ - K \ )$, where $\tilde{K} \ = n^{-1}\sum_{=1}^{n} x \quad \otimes x \quad$.

LEMMA 2*(a) Under conditions (A1), (A3), (B1)–(B5), we have*

$$\theta^{(1)} = O_p(k^{\ +2}), \qquad \theta_{\ 1\ 2}^{\ 1\ 2\ (2)} = O_p\big(k_1^{\ /2+1}n^{-1/2} + k_2^{\ /2+1}n^{-1/2}\big),$$
$$\theta_{\ 1\ 2}^{\ 1\ 2\ (3)} = O_p(n^{-1/2}), \qquad \theta_{\ 1\ 2}^{\ 1\ 2\ (4)} = O_p\big(k_1^{\ /2+1}n^{-1/2} + k_2^{\ /2+1}n^{-1/2}\big),$$
$$\theta^{(5)} = O_p\big(k^{\ /2+1}n^{-1/2}\big), \qquad \theta^{(6)} = O_p(n^{-1/2}),$$
$$\theta^{(7)} = O_p\big(k^{\ /2+1}n^{-1/2}\big), \qquad \theta_{\ 1\ 2}^{(8)} = O_p(n^{-1/2}),$$

where the $O_p(\cdot)$ *and* $o_p(\cdot)$ *terms are uniform for* $k, k_1, k_2 = 1, \ldots, s_n$ *and* $l, l_1, l_2 = 1, \ldots, p_n$.

(b) Under conditions (A1)–(A7), (B1)–(B5), uniformly for $l = 1, \ldots, p_n$,

$$\vartheta^{(1)} = \vartheta^{(2)} = \vartheta^{(3)} = \vartheta^{(5)} = \vartheta^{(6)} = \vartheta^{(7)} = o_p(n^{1/2}),$$
$$\vartheta^{(4)} = \vartheta^{(5)} + \vartheta^{(6)} + \vartheta^{(7)} + \vartheta^{(8)}.$$

Lemma 3 characterizes the eigenvalues of the design matrices $\check{N}_1$, and Lemma 4 concerns the asymptotic order of $\Lambda_1 = P_{\check{N}_1}(Y_M - \check{N}_1\tilde{\eta}_{10})$, where $P_{\check{N}_1} = \check{N}_1(\check{N}_1^{\mathrm{T}}\check{N}_1)^{-1}\check{N}_1^{\mathrm{T}}$, $Y_M = (y_1, ..., y_n)^{\mathrm{T}}$, and $\tilde{\eta}_{10}$ is the true parameter of $\tilde{\eta}_1$.

LEMMA 3. *Under conditions (A1), (A3), (A5), (B1)–(B5), we have* $|\lambda_{\min}(\check{N}_1^{\mathrm{T}}\check{N}_1/n) - \lambda_{\min}(U_1)| = o_p(1)$, $|\lambda_{\max}(\check{N}_1^{\mathrm{T}}\check{N}_1/n) - \lambda_{\max}(U_1)| = o_p(1)$.

LEMMA 4. *Under conditions (A1)–(A5), (B1)–(B5), we have* $\|\Lambda_1\|_2^2 = O_p(r_n^2)$, *where* $r_n^2 = q_n + s_n$.

4. PROOF OF LEMMAS

Proof of Lemma 1. We shall show Lemma 1 for any fixed $j = 1, \ldots, d$, we suppress the subscript j in this proof for convenience. For part (a), recall that $W = x(t) + \varepsilon$, where the error ε are independent of x. Thus one can factor the probability space $\Omega = \Omega_1 \times \Omega_2$, where Ω_1 is for $\{x\}_{=1}^{n}$ and Ω_2 for $\{\varepsilon\}_{=1}$ $_{i}$. Given the data for a single subject, a specific realization x corresponds to fixing a value $\omega \in \Omega_1$, that is $x(\cdot) = x(\cdot, \omega)$. We write $E_{\cdot}$ for expectations with regard to the probability measure on Ω_2, and E_X with respect to Ω_1. For each fixed ω, the error $\{\varepsilon, i = 1, \ldots, n\}$ are independent and identically distributed for different l, with $E_{\cdot}(\varepsilon) = 0$ and $E_{\cdot}(\varepsilon^2) = \sigma^2$. Given condition (B3) together with a local linear smoother, one has $E(\|\hat{x} - x\|^2) = E_X\big[E_{\cdot}\big\{\int(\hat{x} - x)^2\big\}\big]$. Using standard Taylor expansion argument, together with the dominant convergence theorem given $E(\|x^{(2)}\|^4) < \infty$, it is easy to verify $E_X\big[E_{\cdot}\big\{\int(\hat{x} - x)^2\big\}\big] \le C\int\big[E_X\big\{(x^{(2)})^2\big\}h^4 + (mh)^{-1}\big]$ under (B3), yielding $E(\|\hat{x} - x\|^2) = o(n^{-1})$ by (B4). Similar arguments will also lead to $E_X\big[E_{\cdot}\{\int(\hat{x} - x)^4\}\big] \le C\int\big[E_X\{(x^{(2)})^4\}h^8 + (mh)^{-2}\big]$, thus $E\big\{\int(\hat{x} - x)^4\big\}\big] = o(n^{-2})$. Regarding $\hat{K} = n^{-1}\sum_{=1}^{n}\hat{x} \otimes \hat{x}$, notice that $\hat{K} - K = n^{-1}\sum_{=1}^{n}\big\{(\hat{x} - x) \otimes x + (\hat{x} - x) \otimes (\hat{x} - x)\big\} + \big(n^{-1}\sum_{=1}^{n} x \otimes x - K\big)$. It is easy to see that the first term on the right hand side is dominated by $n^{-1}\sum_{=1}^{n}|||(\hat{x} - x) \otimes x|||$. Since $\{\hat{x}, x\}$ is independent of $\{\hat{x}_{\prime}, x_{\prime}\}$ for $i \ne i'$, we have $E\Big[\Big\{n^{-1}\sum_{=1}^{n}|||(\hat{x} - x) \otimes x|||\Big\}^2\Big] = \big\{E|||(\hat{x} - x) \otimes x|||\big\}^2 + n^{-1}\mathrm{var}(|||(\hat{x} - x) \otimes x|||) \le E(\|\hat{x} - x\|^2)(E\|x\|^2) + n^{-1}\big\{E(\|(\hat{x} - x)\|^4)(E\|x\|^4)\big\}^{1/2} = o(n^{-1})$. The second term $E(|||\big(n^{-1}\sum_{=1}^{n} x \otimes x - K|||^2) = O(n^{-1})$ by Lemma 3.3 of Hall & Hosseini-Nasab (2009). This leads to $E(|||\hat{K} - K|||^2) = O(n^{-1})$.

We next obtain the bound in (b) for $k = 1, \ldots, s_n$. First notice that $\mathcal{F}_n$ can be implied by $\min_{=1}$ $_{n}(w - w_{+1}) \ge s_n^{-}$ $^{-1} \ge Cn^{-}$ $^{/2}$ for some C and any $\tau < 1$, which is fulfilled by $s_n^{+1}n^{-1/2} \to 0$ in (A3), leading to $pr(\mathcal{F}_n) \to 1$ as $n \to \infty$. It is easy to see that for $k = 1, \ldots, s_n$, $w - w_{+1} \ge s_n^{-}$ $^{-1} > 2\hat{\Delta}$ as $n \to \infty$, that is $s_n \in \mathcal{J}_n$ for large n. Now we will bound $||\hat{\phi} - \phi||^2$ for $k = 1, \ldots, s_n$. By (5.16) in Hall & Horowitz (2007), one has $\|\hat{\phi} - \phi\|^2 \le 2\hat{u}^2$, where $\hat{u}^2 = \sum_{:\ne}(\hat{w} - w)^{-2}\Big\{\int(\hat{K} - K)\hat{\phi}\,\phi\Big\}^2$. Also $pr(\mathcal{F}_n) \to 1$

implies that $(\hat{w} \ \ - w \ )^{-2} \leq Cn$ with probability tending to 1. Then we have

$$
\begin{aligned}
\hat{u}^2 \leq & \sum_{: \neq} (\hat{w} \ \ - w \ )^{-2} \Big[2\Big\{ \int (\hat{K} - K)\phi \ \phi \ \Big\}^2 \\
& +2\Big\{ \int (\hat{K} - K)(\hat{\phi} \ \ - \phi \ )\phi \ \Big\}^2 \Big] \\
\leq & \ 2 \sum_{: \neq} (\hat{w} \ \ - w \ )^{-2} \Big\{ \int (\hat{K} - K)\phi \ \phi \ \Big\}^2 \\
& +2Cn \ \sum_{\neq} \Big\{ \int (\hat{K} - K)(\hat{\phi} \ \ - \phi \ )\phi \ \Big\}^2 \\
\leq & \ 4 \sum_{: \neq} (w \ \ - w \ )^{-2} \Big\{ \int (\hat{K} - K)\phi \ \phi \ \Big\}^2 + 2Cn \ \hat{\Delta}^2 ||\hat{\phi} \ \ - \phi \ ||^2.
\end{aligned}
$$

Plugging this into $\|\hat{\phi} \ \ - \phi \ \|^2 \leq 2\hat{u}^2$, one has

$$
(1 - 2Cn \ \hat{\Delta}^2)\|\hat{\phi} \ \ - \phi \ \|^2 \leq 4 \sum_{: \neq} (w \ \ - w \ )^{-2} \Big\{ \int (\hat{K} - K)\phi \ \phi \ \Big\}^2.
$$

As $\hat{\Delta} = O_p(n^{-1/2})$ and $\tau < 1$, we have $n \ \hat{\Delta}^2 = o_p(1)$, and $\|\hat{\phi} \ \ - \phi \ \|^2 \leq C \sum_{\neq} (w \ \ - w \ )^{-2}\{\int (\hat{K} - K)\phi \ \phi \ \}^2$. Given the results in (a), by analogy to (5.22) in Hall & Horowitz (2007), $nE[\sum_{\neq} (w \ \ - w \ )^{-2}\{\int (\hat{K} - K)\phi \ \phi \ \}^2] = O(k^2)$ still holds, the result follows by Chebyshev's inequality.

To obtain (c), using Lemma 5.1 of Hall & Horowitz (2007) with $\psi \ \ = \hat{\phi} \ $, $\lambda \ \ = \hat{w}$ and $L = \hat{K}$, since $\inf_{\{ \ : \ \neq \ \}} |\hat{w} \ \ - w \ | > 0$, we have

$$
\begin{aligned}
\hat{\phi} \ \ - \phi \ \ = & \sum_{: \neq} (\hat{w} \ \ - w \ )^{-1}\phi \ \int (\hat{K} - K)\hat{\phi} \ \phi \ + \phi \ \int (\hat{\phi} \ \ - \phi \ )\phi \\
= & \sum_{: \neq} (w \ \ - w \ )^{-1}\phi \ \int (\hat{K} - K)\phi \ \phi \ + \phi \ \int (\hat{\phi} \ \ - \phi \ )\phi \\
& + \sum_{: \neq} (w \ \ - w \ )^{-1}\phi \ \int (\hat{K} - K)(\hat{\phi} \ \ - \phi \ )\phi \\
& - \sum_{: \neq} (\hat{w} \ \ - w \ )\{(w \ \ - w \ )(\hat{w} \ \ - w \ )\}^{-1}\phi \ \int (\hat{K} - K)\hat{\phi} \ \phi \ .
\end{aligned}
$$

Denote the last three terms by $\alpha \ \ \equiv \alpha_1 \ + \alpha_2 \ + \alpha_3$. For α_1 , one has

$$
\|\hat{\phi} \ \ - \phi \ \|^2 = 2 - 2\int \hat{\phi} \ \phi \ = 2(\int \phi^2 - \int \hat{\phi} \ \phi \ ) = -2\int (\hat{\phi} \ \ - \phi \ )\phi \ ,
$$

that is $\alpha_1 \ = -\|\hat{\phi} \ \ - \phi \ \|^2/2$. From part (b), one has $\|\hat{\phi} \ \ - \phi \ \| = O_p(kn^{-1/2})$ uniformly in $k = 1, \ldots, s_n$, then $\|\alpha_1 \ \| = O_p(k^2n^{-1})$. To bound α_2 , noticing $\| \sum \ c \ \phi \ \|^2 = \sum \ c^2$ due to orthonormal $\{\phi\ \}$,

$$\|\alpha_2\ \| \le \Big[\sum_{\neq}(w\ \ - w\ )^{-2}\Big\{\int(\hat K - K)(\hat\phi\ \ - \phi\ )\phi\ \Big\}^2\Big]^{1/2}$$
$$\le \delta^{-1}\hat\Delta\|\hat\phi\ \ - \phi\ \| = O_p(k\ ^{+2}n^{-1}).$$

For α_3 , noticing that $\sup_{\ =1\ \ \ \infty}|\hat w\ \ - w\ | = O_p(\hat\Delta)$, we can bound $\|\alpha_3\ \|$ as follows,

$$2|\hat w\ \ - w\ |\Big[\sum_{:\ \neq}(w\ \ - w\ )^{-4}\Big\{\int(\hat K - K)\hat\phi\ \ \phi\ \Big\}^2\Big]^{1/2}$$
$$\le C\hat\Delta\Big[\sum_{:\ \neq}(w\ \ - w\ )^{-4}\Big\{\int(\hat K - K)\phi\ \ \phi\ \Big\}^2$$
$$+ \sum_{:\ \neq}(w\ \ - w\ )^{-4}\Big\{\int(\hat K - K)(\hat\phi\ \ - \phi\ )\phi\ \Big\}^2\Big]^{1/2}$$
$$= C\hat\Delta(A_1 + A_2)^{1/2},$$

where $A_1 = \sum_{:\ \neq}(w\ \ - w\ )^{-4}\big\{\int(\hat K - K)\phi\ \ \phi\ \big\}^2$ and $A_2 = \sum_{:\ \neq}(w\ \ - w\ )^{-4}\big\{\int(\hat K - K)(\hat\phi\ \ - \phi\ )\phi\ \big\}^2$. By the derivations of part (b), we have

$$A_1 \le \delta^{-2}\sum_{:\ \neq}(w\ \ - w\ )^{-2}\Big\{\int(\hat K - K)\phi\ \ \phi\ \Big\}^2 = O_p(k^2\delta^{-2}n^{-1}),$$
$$A_2 \le \delta^{-4}\hat\Delta^2\|\hat\phi\ \ - \phi\ \|^2 = O_p(\delta^{-4}j^2\hat\Delta^4).$$

Notice that $\delta^{-2} = O\{k^{2(\ +1)}\} = o(n)$ by (A3), which indicates $A_2 = o_p(k^2\delta^{-2}n^{-1})$. Thus $\|\alpha_3\ \| = O_p(k\delta^{-1}n^{-1})$. It leads to $\|\alpha\ \| = O_p(k\ ^{+2}n^{-1})$.

For part (d), since $\sum_{\ =1}^{n} w\ \ < \infty$ and $\sup_{\ =1\ \ \ n}|\hat w\ \ - w\ | \le |||\hat K - K||| = O_p(n^{-1/2})$ by Theorem 1 of Hall & Hosseini-Nasab (2006) and part (a), we can see that $(\sum_{\ =1}^{n}\hat w\ )^{1/2} = O_p(1)$ given $n^{-1/2}s_n = o(1)$, which completes the proof.

Proof of Lemma 2. For $\theta^{(1)}$ and any fixed j, we have

$$\theta^{(1)} = \sum_{\ =1}^{n}(\hat\xi\ \ \ - \xi\ \ )^2w^{-1} = \sum_{\ =1}^{n}(\hat\xi\ \ \ - \tilde\xi\ \ \ + \tilde\xi\ \ \ - \xi\ \ )^2w^{-1}$$
$$\le 2\sum_{\ =1}^{n}(\hat\xi\ \ \ - \tilde\xi\ \ )^2w^{-1} + 2\sum_{\ =1}^{n}(\tilde\xi\ \ \ - \xi\ \ )^2w^{-1}$$
$$= 2\sum_{\ =1}^{n}\Big\{\int\hat x\ \ (\hat\phi\ \ \ - \phi\ \ )\Big\}^2w^{-1} + 2\sum_{\ =1}^{n}\Big\{\int(\hat x\ \ \ - x\ \ )\phi\ \ \Big\}^2w^{-1}$$
$$\le 2\sum_{\ =1}^{n}\big(\|\hat x\ \ \|^2\|\hat\phi\ \ \ - \phi\ \ \|^2 + \|\phi\ \ \|^2\|\hat x\ \ \ - x\ \ \|^2\big)w^{-1}.$$

Given Lemma 1 and (B1), we know

$$E(\|\hat x\ \ \|^2) \le 2E(\|\hat x\ \ \ - x\ \ \|^2) + 2E(\|x\ \ \|^2) = O(1),$$

hence, $\|\hat{x}\ \ \|^2 = O_p(1)$. From Lemma 1, we have $\theta^{(1)} = \sum_{=1}^{n} \big\{O_p(1)O_p(k^2 n^{-1}) + o_p(n^{-1})\big\}k \ \ = O_p(k^{\ +2})$, uniformly for $k = 1, \ldots, s_n$.

For $\theta_{\ 1\ 2}^{\ 1\ 2\ (2)}$, it is obvious that

$$
\begin{aligned}
|\theta_{\ 1\ 2}^{\ 1\ 2\ (2)}| &= \Big| n^{-1} \sum_{=1}^{n} (\hat{\xi}\ \ _{1\ 1}\hat{\xi}\ \ _{2\ 2} - \xi\ \ _{1\ 1}\xi\ \ _{2\ 2})(w\ _{1\ 1} w\ _{2\ 2})^{-1/2} \Big| \\
&= n^{-1} \Big| \sum_{=1}^{n} \hat{\xi}\ \ _{1\ 1}(\hat{\xi}\ \ _{2\ 2} - \xi\ \ _{2\ 2})(w\ _{1\ 1} w\ _{2\ 2})^{-1/2} + \sum_{=1}^{n} \xi\ \ _{2\ 2}(\hat{\xi}\ \ _{1\ 1} - \xi\ \ _{1\ 1})(w\ _{1\ 1} w\ _{2\ 2})^{-1/2} \Big| \\
&\le n^{-1} \big(\sum_{=1}^{n} \hat{\xi}^2\ _{1\ 1} w^{-1}_{\ 1\ 1} \big)^{1/2} \big\{ \sum_{=1}^{n} (\hat{\xi}\ \ _{2\ 2} - \xi\ \ _{2\ 2})^2 w^{-1}_{\ 2\ 2} \big\}^{1/2} + \\
&\quad n^{-1} \big(\sum_{=1}^{n} \xi^2\ _{2\ 2} w^{-1}_{\ 2\ 2} \big)^{1/2} \big\{ \sum_{=1}^{n} (\hat{\xi}\ \ _{1\ 1} - \xi\ \ _{1\ 1})^2 w^{-1}_{\ 1\ 1} \big\}^{1/2}.
\end{aligned}
$$

Since $E(\sum_{=1}^{n} \xi^2\ _{2\ 2} w^{-1}_{\ 2\ 2}) = n$ for any $k_2 = 1, \ldots, s_n$, we have $\sum_{=1}^{n} \xi^2\ _{2\ 2} w^{-1}_{\ 2\ 2} = O_p(n)$, uniformly for $k_2 = 1, \ldots, s_n$. Moreover,

$$
\begin{aligned}
\sum_{=1}^{n} \hat{\xi}^2\ _{1\ 1} w^{-1}_{\ 1\ 1} &\le 2 \sum_{=1}^{n} (\hat{\xi}\ \ _{1\ 1} - \xi\ \ _{1\ 1})^2 w^{-1}_{\ 1\ 1} + 2 \sum_{=1}^{n} \xi^2\ _{1\ 1} w^{-1}_{\ 1\ 1} \\
&= O_p(k_1^{\ +2} + n) = O_p(n),
\end{aligned}
$$

uniformly for $k_1 = 1, \ldots, s_n$. In conclusion, we have

$$
\begin{aligned}
|\theta_{\ 1\ 2}^{\ 1\ 2\ (2)}| &= n^{-1} O_p(n^{1/2}) O_p(k_2^{\ /2+1}) + n^{-1} O_p(n^{1/2}) O_p(k_1^{\ /2+1}) \\
&= O_p\Big(k_1^{\ /2+1} n^{-1/2} + k_2^{\ /2+1} n^{-1/2} \Big),
\end{aligned}
$$

uniformly for $k_1, k_2 = 1, \ldots, s_n$.

For $\theta_{\ 1\ 2}^{\ 1\ 2\ (3)}$, $E(\theta_{\ 1\ 2}^{\ 1\ 2\ (3)})^2 \le n^{-1}\big\{E(\xi^4\ _{1\ 1} w^{-2}_{\ 1\ 1}) E(\xi^4\ _{2\ 2} w^{-2}_{\ 2\ 2})\big\}^{1/2} \le c_1 n^{-1}$, uniformly for $k_1, k_2 = 1, \ldots, s_n$ by (B1). It follows that $\theta_{\ 1\ 2}^{\ 1\ 2\ (3)} = O_p(n^{-1/2})$, uniformly for $k_1, k_2 = 1, \ldots, s_n$. Moreover, it is trivial that $\theta_{\ 1\ 2}^{\ 1\ 2\ (4)} = \theta_{\ 1\ 2}^{\ 1\ 2\ (2)} + \theta_{\ 1\ 2}^{\ 1\ 2\ (3)} = O_p(n^{-1/2} k_1^{\ /2+1} + n^{-1/2} k_2^{\ /2+1})$, uniformly for $k_1, k_2 = 1, \ldots, s_n$.

Based on (B5), we conclude that $z^4 \ = O_p(1)$, uniformly for $l = 1, \ldots, p_n$. For $\theta^{(5)}$, we have

$$
\begin{aligned}
|\theta^{(5)}| &\le n^{-1} \Big\{ \sum_{=1}^{n} (\hat{\xi}\ \ \ \ - \xi\ \ \ \ )^2 w^{-1} \Big\}^{1/2} \big(\sum_{=1}^{n} z^2 \big)^{1/2} \\
&= n^{-1} O_p(k^{\ /2+1}) O_p(n^{1/2}) = O_p\Big(k^{\ /2+1} n^{-1/2} \Big),
\end{aligned}
$$

uniformly for $k = 1, \ldots, s_n$, $l = 1, \ldots, p_n$. Furthermore,

$$
E(\theta^{(6)})^2 \le n^{-1} \big\{ E(\xi^4\ \ w^{-2}) E(z^4) \big\}^{1/2} = O\Big(n^{-1} \Big),
$$

uniformly for $k = 1, \ldots, s_n$, $l = 1, \ldots, p_n$. Thus $\theta^{(6)} = O_p(n^{-1/2})$, uniformly for $k = 1, \ldots, s_n$, $l = 1, \ldots, p_n$. Then it follows that $\theta^{(7)} = \theta^{(5)} + \theta^{(6)} = O_p(k^{\ /2+1} n^{-1/2})$, uniformly for $k = 1, \ldots, s_n$, $l = 1, \ldots, p_n$. For $\theta^{(8)}_{\ 1\ 2}$, we have $E(\theta^{(8)}_{\ 1\ 2})^2 \le n^{-1}\{E(z^4_{\ \ 1})E(z^4_{\ \ 2})\}^{1/2} = O(n^{-1})$, uniformly for $l_1 = 1, \ldots, p_n$, $l_2 = 1, \ldots, p_n$, and this entails that $\theta^{(8)}_{\ 1\ 2} = O_p(n^{-1/2})$, uniformly for $l_1 = 1, \ldots, p_n$, $l_2 = 1, \ldots, p_n$.

For $\vartheta^{(1)}$, given (A4), we have

$$E|\vartheta^{(1)}| \le \sum_{=1}^{n} \sum_{= \ n+1}^{\infty} |b \ \ _0| \, E\,|\xi \quad z \ \ | \le \sum_{=1}^{n} \sum_{= \ n+1}^{\infty} |b \ \ _0| \, \{E(\xi^2 \ \ )E(z^2)\}^{1/2}$$
$$\le c_1 \sum_{=1}^{n} \sum_{= \ n+1}^{\infty} k^{-} \, k^{-1/2} = O(n s_n^{- \ +1/2}) = o(n^{1/2}).$$

Hence $\vartheta^{(1)} = o_p(n^{1/2})$, uniformly for $l = 1, \ldots, p_n$. For $\vartheta^{(2)}$, we have

$$\vartheta^{(2)} = \sum_{=1}^{n} \sum_{=1}^{n} \{ \int \hat{x} \ \ (\hat{\phi} \quad - \phi \ \ ) + \int (\hat{x} \quad - x \ \ )\phi \ \ \}(\hat{b} \quad - b \ \ _0)z$$
$$= \int (\sum_{=1}^{n} \hat{x} \ \ z \ \ )\{\sum_{=1}^{n} (\hat{\phi} \quad - \phi \ \ )(\hat{b} \quad - b \ \ _0)\}$$
$$+ \int \{\sum_{=1}^{n} (\hat{x} \quad - x \ \ )z \ \ \}\{\sum_{=1}^{n} \phi \ \ (\hat{b} \quad - b \ \ _0)\}.$$

It follows that

$$(\vartheta^{(2)})^2 \le 2\| \sum_{=1}^{n} \hat{x} \ \ z \ \ \|^2 \| \sum_{=1}^{n} (\hat{\phi} \quad - \phi \ \ )(\hat{b} \quad - b \ \ _0)\|^2$$
$$+2\| \sum_{=1}^{n} (\hat{x} \quad - x \ \ )z \ \ \|^2 \| \sum_{=1}^{n} \phi \ \ (\hat{b} \quad - b \ \ _0)\|^2.$$

Since

$$E\| \sum_{=1}^{n} \hat{x} \ \ z \ \ \|^2 \le E(\sum_{=1}^{n} \|\hat{x} \ \ \||z \ \ |)^2 \le n^2\{E(\|\hat{x}_1 \ \ \||z_1 \ |)\}^2 + n\{E(\|\hat{x}_1 \ \ \|^4)E(z_1^4 \ )\}^{1/2} = O(n^2),$$

we have $\| \sum_{=1}^{n} \hat{x} \ \ z \ \ \|^2 = O_p(n^2)$. Similarly,

$$E\| \sum_{=1}^{n} (\hat{x} \quad - x \ \ )z \ \ \|^2 \le n^2(E\|\hat{x} \quad - x \ \ \||z \ \ |)^2 + n\{E(\|\hat{x} \quad - x \ \ \|^4)E(z^4)\}^{1/2} = o(n),$$

lead to $\| \sum_{=1}^{n} (\hat{x} \quad - x \ \ )z \ \ \|^2 = o_p(n)$. Moreover,

$$\| \sum_{=1}^{n} (\hat{\phi} \quad - \phi \ \ )(\hat{b} \quad - b \ \ _0)\|^2 \le \|\breve{b}^{(1)} - \tilde{b}^{(1)}_0\|_2^2 \sum_{=1}^{n} \|\hat{\phi} \quad - \phi \ \ \|^2 w^{-1}$$
$$= O_p(r_n^2/n) O_p\{\sum_{=1}^{n} (k^2/n)k \ \ \}$$
$$= O_p(r_n^2 s_n^{\ +3}/n^2),$$

and

$$\|\sum_{=1}^{n}\phi\quad(\hat{b}\quad - b\quad_0)\|^2 \le (\sum_{=1}^{n}|\hat{b}\quad - b\quad_0|)^2 \le \|\check{b}^{(1)} - \tilde{b}_0^{(1)}\|_2^2 \sum_{=1}^{n} w^{-1} = O_p(r_n^2 s_n^{\ +1}/n),$$

by Theorem 1. In summary, we get $(\vartheta^{(2)})^2 = o_p(r_n^2 s_n^{\ +3}) = o_p(q_n s_n^{\ +3} + s_n^{\ +4})$, uniformly for $l = 1, \ldots, p_n$. Under conditions (A3) and (A5), $\vartheta^{(2)} = o_p(n^{1/2})$, uniformly for $l = 1, \ldots, p_n$.

For $\vartheta^{(3)}$, we have

$$(\vartheta^{(3)})^2 = [\int\{\sum_{=1}^{n}(\hat{x}\quad - x\quad)z\quad\}(\sum_{=1}^{n}\hat{\phi}\quad b\quad_0)]^2 \le \|\sum_{=1}^{n}(\hat{x}\quad - x\quad)z\quad\|^2\|\sum_{=1}^{n}\hat{\phi}\quad b\quad_0\|^2$$
$$\le (\sum_{=1}^{n}|b\quad_0|)^2 o_p(n) = o_p(n),$$

hence $\vartheta^{(3)} = o_p(n^{1/2})$ uniformly for $l = 1, \ldots, p_n$.

Next, we show that $\vartheta^{(4)} = \vartheta^{(5)} + \vartheta^{(6)} + \vartheta^{(7)} + \vartheta^{(8)}$. Recall that $\Xi\quad = (\Xi\quad_1, ..., \Xi\quad_{q_n})^{\mathrm{T}}$ with $E\{x\quad(t)z\quad\} = \Xi\quad(t)$. By lemma 1, we have the expression

$$\hat{\phi}\quad(t) - \phi\quad(t) = n^{-1/2}\sum_{:\ \neq}(w\quad - w\quad)^{-1}\phi\quad(t)\int \Re\ \phi\quad\phi\quad + \alpha\quad(t),$$

where $||\alpha\quad|| = O_p(k^{\ +2}n^{-1})$, $\Re\quad = n^{1/2}(\hat{K}\quad - K\quad)$, and $O_p(\cdot)$ is uniform in $k = 1, \ldots, s_n$. Since

$$\vartheta^{(4)} = \sum_{=1}^{n}\sum_{=1}^{n}(\tilde{\xi}\quad - \xi\quad)b\quad_0 z\quad = \int(\sum_{=1}^{n}x\quad z\quad)\{\sum_{=1}^{n}(\hat{\phi}\quad - \phi\quad)b\quad_0\},$$

substitute the expression for $\hat{\phi}\quad(t) - \phi\quad(t)$ into the above equation, we immediately get $\vartheta^{(4)} = \vartheta^{(5)} + \vartheta^{(6)} + \vartheta^{(7)} + \vartheta^{(8)}$.

For $\vartheta^{(5)}$, it is obvious that

$$(\vartheta^{(5)})^2 \le \|\sum_{=1}^{n}(\hat{\phi}\quad - \phi\quad)b\quad_0\|^2\|\sum_{=1}^{n}(x\quad z\quad - \Xi\quad)\|^2,$$

where

$$E\|\sum_{=1}^{n}(x\quad z\quad - \Xi\quad)\|^2 = \int E[\{\sum_{=1}^{n}(x\quad z\quad - \Xi\quad)\}^2] = n\int \mathrm{var}(x\quad z\quad)$$
$$\le n\int\{E(x^4\quad)E(z^4)\}^{1/2} = O(n),$$

and

$$\|\sum_{=1}^{n}(\hat{\phi}\quad - \phi\quad)b\quad_0\| \le \sum_{=1}^{n}|b\quad_0|\|\hat{\phi}\quad - \phi\quad\| = O_p\big(\sum_{=1}^{n}k^{-}\ kn^{-1/2}\big) = O_p(n^{-1/2}),$$

thus $(\vartheta^{(5)})^2 = O_p(1) = o_p(n)$, and it follows that $\vartheta^{(5)} = o_p(n^{1/2})$ uniformly for $l = 1, \ldots, p_n$. For $\vartheta^{(6)}$, we have

$$(\vartheta^{(6)})^2 \le n^2 \int (Ex \quad z \quad)^2 \| \sum_{=1}^{n} b \quad {}_0\alpha \quad \|^2 \le n^2 \int \{E(x^4\ )E(z^4)\}^{1/2} (\sum_{=1}^{n} |b \quad {}_0| \|\alpha \quad \|)^2$$
$$= n^2 O_p\{(\sum_{=1}^{n} k^{-} \ k \ ^{+2} n^{-1})^2\} = o_p(n),$$

hence $\vartheta^{(6)} = o_p(n^{1/2})$ uniformly for $l = 1, \ldots, p_n$.

From the proof of lemma 1, it is easy to see that $\Re \ - \Re^* = n^{1/2}(\hat{K} \ - \tilde{K}\ ) = o_p(1)$, which follows that $\vartheta^{(7)} = o_p(n^{1/2})$, uniformly for $l = 1, \ldots, p_n$.

Proof of Lemma 3. First, it is obvious that $|\lambda_{\min}(\check{N}_1^{\mathrm{T}}\check{N}_1/n) - \lambda_{\min}(U_1)| \le ||\check{N}_1^{\mathrm{T}}\check{N}_1/n - U_1||_1$, where $||.||_1$ is the L_1 norm for matrix. Since

$$||\check{N}_1^{\mathrm{T}}\check{N}_1/n - U_1||_1 \le O_p(\sum_{1=1}^{n} |\theta^{(4)}_{\ 1\ 2 \atop 1\ n}| + \sum_{=1}^{q_n} |\theta^{(7)}_{\ 1\ n}| + \sum_{1=1}^{n} |\theta^{(7)}_{\ 1\ 1q_n}| + \sum_{1=1}^{q_n} |\theta^{(8)}_{1q_n}|)$$
$$= O_p\{\sum_{1=1}^{n} (k_1^{\ /2+1} n^{-1/2} + s_n^{\ /2+1} n^{-1/2}) + \sum_{=1}^{q_n} (s_n^{\ /2+1} n^{-1/2})$$
$$+ \sum_{1=1}^{n} (k_1^{\ /2+1} n^{-1/2}) + \sum_{1=1}^{q_n} n^{-1/2}\}$$
$$= O_p(s_n^{\ /2+2} n^{-1/2} + q_n s_n^{\ /2+1} n^{-1/2}),$$

by Lemma 2 (a), hence we have $|\lambda_{\min}(\check{N}_1^{\mathrm{T}}\check{N}_1/n) - \lambda_{\min}(U_1)| = O_p(s_n^{\ /2+2} n^{-1/2} + q_n s_n^{\ /2+1} n^{-1/2})$. Under conditions (A3) and (A5), it is obvious that $|\lambda_{\min}(\check{N}_1^{\mathrm{T}}\check{N}_1/n) - \lambda_{\min}(U_1)| = o_p(1)$. Similarly $|\lambda_{\max}(\check{N}_1^{\mathrm{T}}\check{N}_1/n) - \lambda_{\max}(U_1)| \le ||\check{N}_1^{\mathrm{T}}\check{N}_1/n - U_1||_1 = o_p(1)$, and we skip the details here.

Proof of Lemma 4. By Lemma 3 and (B5), we know that $\check{N}_1^{\mathrm{T}}\check{N}_1$ is invertible, hence $P_{\check{N}_1}$ exists. For Λ_1, we have

$$\Lambda_1 = P_{\check{N}_1}(Y_M - \check{N}_1\tilde{\eta}_{10}) = P_{\check{N}_1}\{Y_M - \tilde{N}_1\tilde{\eta}_{10} - \nu + \nu + (\tilde{N}_1 - \check{N}_1)\tilde{\eta}_{10}\}$$
$$= P_{\check{N}_1}\{\epsilon + \nu + (\tilde{N}_1 - \check{N}_1)\tilde{\eta}_{10}\},$$

where $\epsilon = Y_M - \tilde{N}_1\tilde{\eta}_{10} - \nu = (\epsilon_1, ..., \epsilon_n)^{\mathrm{T}}$, $\nu = (\nu_1, ..., \nu_n)^{\mathrm{T}}$ with $\nu = \sum_{=1} \sum_{= \ n+1}^{\infty} \xi \quad b \quad {}_0$.

For $P_{\check{N}_1}\epsilon$, we have

$$E\|P_{\check{N}_1}\epsilon\|^2 = E(\epsilon^{\mathrm{T}} P_{\check{N}_1}\epsilon) = E\{E(\epsilon^{\mathrm{T}} P_{\check{N}_1}\epsilon \mid \epsilon)\} = E[\mathrm{tr}\{P_{\check{N}_1} E(\epsilon\epsilon^{\mathrm{T}})\}]$$
$$= \sigma^2 \mathrm{tr}(P_{\check{N}_1}) = \sigma^2(q_n + g s_n) = O(q_n + s_n),$$

hence $\|P_{\check{N}_1}\epsilon\|^2 = O_p(q_n + s_n)$.

For $P_{\check{N}_1}(\tilde{N}_1 - \check{N}_1)\tilde{\eta}_{10}$, we have

$$\|P_{\check{N}_1}(\tilde{N}_1 - \check{N}_1)\tilde{\eta}_{10}\|^2 \le \|(\tilde{N}_1 - \check{N}_1)\tilde{\eta}_{10}\|^2 \le O\big[\sum_{=1}^{n}\{\sum_{=1}\sum_{=1}^{n}(\hat{\xi} \quad - \xi \quad)b \quad _0\}^2\big]$$

$$\le O\big[2g\sum_{=1}^{n}\sum_{=1}\{\sum_{=1}^{n}(\tilde{\xi} \quad - \xi \quad)b \quad _0\}^2 + 2g\sum_{=1}^{n}\sum_{=1}\{\sum_{=1}^{n}(\hat{\xi} \quad - \tilde{\xi} \quad)b \quad _0\}^2\big]$$

$$\le O\big[\sum_{=1}\sum_{=1}^{n}\|x \quad \|^2 O_p\big((\sum_{=1}^{n} k^{-} \; kn^{-1/2})^2\big]\big) + O_p\big[\sum_{=1}\sum_{=1}^{n}\{\|\hat{x} \quad - x \quad \|^2(\sum_{=1}^{n} k^{-} \;)^2\}\big]$$

$$= O_p(1),$$

since $b > 2$ implies that $\sum_{=1}\sum_{=1}^{n}\|x \quad \|^2 O_p\{(\sum_{=1}^{n} k^{-} \; kn^{-1/2})^2\} = O_p(1)$, and Lemma 1 entails that $\sum_{=1}\sum_{=1}^{n}\{\|\hat{x} \quad - x \quad \|^2(\sum_{=1}^{n} k^{-} \;)^2\} = O_p(1)$. It then follows that $\|P_{\check{N}_1}(\tilde{N}_1 - \check{N}_1)\tilde{\eta}_{10}\|^2 = O_p(1)$.

For $P_{\check{N}_1}\nu$, it is obvious that $\|P_{\check{N}_1}\nu\|^2 \le \|\nu\|^2$. For $i = 1, \dots, n$, we have

$$E(\nu^2) = O\Big\{\sum_{=1}\text{var}\big(\sum_{= \; n+1}^{\infty}\xi \quad b \quad _0\big)\Big\} = O\Big(\sum_{=1}\sum_{= \; n+1}^{\infty} b^2 \; _0 w \quad \Big)$$

$$= O\Big(\sum_{=1}\sum_{= \; n+1}^{\infty} k^{-2} \; k^{-1}\Big) = O(s_n^{-2} \;).$$

It follows that $\|P_{\check{N}_1}\nu\|^2 = O_p(ns_n^{-2} \;)$. In summary,

$$\|\Lambda_1\|_2^2 \le O(\|P_{\check{N}_1}\epsilon\|^2 + \|P_{\check{N}_1}(\tilde{N}_1 - \check{N}_1)\tilde{\eta}_{10}\|^2 + \|P_{\check{N}_1}\nu\|^2)$$
$$= O_p(q_n + s_n + 1 + ns_n^{-2} \;) = O_p(q_n + s_n) = O_p(r_n^2),$$

where $r_n^2 = q_n + s_n$. This completes the proof.

5. PROOFS OF MAIN THEOREMS

Proof of Theorem 1. First, we constrain $Q_n(\tilde{\eta})$ on the subspace, where the true zero parameters are set as 0, that is $\{\tilde{\eta} \in R^{d \; n+p_n} : \tilde{b}^{(1)} = 0, k = g+1, \dots, d, \gamma^{(2)} = 0\}$, and prove consistency in the $(gs_n + q_n)$-dimensional space. Define the constrained penalized function

$$\bar{Q}_n(\tilde{\eta}_1) = \sum_{=1}^{n}\{y \quad - \sum_{=1}\sum_{=1}^{n}(\hat{\xi} \quad w^{-1/2})\tilde{\psi} \; \dots$$

$u^{(1)} = (u_1^{(1)\mathrm{T}}, \ldots, u^{(1)\mathrm{T}})^{\mathrm{T}}$ and $u \ \ = (u_1, \ldots, u_{q_n})^{\mathrm{T}}$. We have

$$
\begin{aligned}
&\bar{Q}_n(\tilde{\eta}_{10} + \alpha_n u) - \bar{Q}_n(\tilde{\eta}_{10}) \\
\geq\ & \|\check{N}_1 \alpha_n u\|^2 - 2\Lambda_1^{\mathrm{T}} \check{N}_1 \alpha_n u + 2n[\sum_{=1}^{q_n} \{J_{\lambda_n}(|\gamma_{\ 0} + \alpha_n u \ \ |) - J_{\lambda_n}(|\gamma_{\ 0}|)\} \\
&+ \sum_{=1} \{J_{\lambda_{jn}}(\|b_{\ 0}^{(1)} + \alpha_n A^{-1} u^{(1)}\|) - J_{\lambda_{jn}}(\|b_{\ 0}^{(1)}\|)\}] \\
\geq\ & n\lambda_{\min}(\check{N}_1^{\mathrm{T}} \check{N}_1/n)\alpha_n^2 \|u\|^2 - 2n^{1/2}\|\Lambda_1\|_2 \lambda_{\max}^{1/2}(\check{N}_1^{\mathrm{T}} \check{N}_1/n)\alpha_n \|u\| \\
&+2n[\sum_{=1}^{q_n} \{J'_{\lambda_n}(|\gamma_{\ 0}|)\,\mathrm{sgn}(\gamma_{\ 0})\alpha_n u \ \ + J''_{\lambda_n}(|\gamma_{\ 0}|)\alpha_n^2 (u \ \ )^2 (1 + o(1))\} \\
&+ \sum_{=1} \{J'_{\lambda_{jn}}(\|b_{\ 0}^{(1)}\|)\alpha_n \|A^{-1} u^{(1)}\| + J''_{\lambda_{jn}}(\|b_{\ 0}^{(1)}\|)\alpha_n^2 \|A^{-1} u^{(1)}\|^2 (1 + o(1))\}].
\end{aligned}
$$

The Taylor expansion in the second inequality holds because $\alpha_n w_{\ n}^{-1/2} \leq r_n s_n^{\ /2} n^{-1/2} = o(1)$ by condition (A5). As for the smoothly clipped absolute deviation penalty, we have $J'_{\lambda_n}(|\gamma_{\ 0}|) = J''_{\lambda_n}(|\gamma_{\ 0}|) = 0$ for all $l = 1, \ldots, q_n$ under condition (A7), and $J'_{\lambda_{jn}}(\|b_{\ 0}^{(1)}\|) = J''_{\lambda_{jn}}(\|b_{\ 0}^{(1)}\|) = 0$ for all $j = 1, \ldots, g$ since $\|b_{\ 0}^{(1)}\| \geq C_1$ 22

When $j = 1, \dots, g$, since $\|\hat{b}^{(1)}\| \geq \|b_0^{(1)}\| - \|\hat{b}^{(1)} - b_0^{(1)}\|$, $\|\hat{b}^{(1)} - b_0^{(1)}\| = o_p(s_n^{\ /2}\alpha_n) = o_p(1)$ and $\min_{\ =1} \quad \|b_0^{(1)}\| > C_1$ for some positive constant C_1, hence, we have $\min_{\ =1} \quad \|\hat{b}^{(1)}\|/\lambda_n \to \infty$ in probability. It follows that

$$pr(\|\hat{b}^{(1)}\| \geq a\lambda_{\ n} \text{ for } j = 1, \dots, g) \to 1. \quad (3)$$

When $l = 1, \dots, q_n$ and $j = 1, \dots, g$, $S\ (\check{\eta}) = 0$ and $S^*(\check{\eta}) = 0$ hold trivially since (2), (3) hold and $\check{\eta}_1$ is a local minimizer of $\bar{Q}_n(\tilde{\eta}_1)$.

Next, we show that $\check{\eta}$ satisfy

$$|S\ (\tilde{\eta})| \leq \lambda_n, \text{ and } , |\gamma\ | < \lambda_n \text{ for } l = q_n + 1, \dots, p_n$$

and

$$||S^*(\tilde{\eta})|| \leq \lambda_{\ n}, \text{ and } , ||b^{(1)}|| < \lambda_{\ n} \text{ for } j = g + 1, \dots, d.$$

Since $\hat{\gamma}\ = 0$ for $l = q_n + 1, \dots, p_n$ and $\hat{b}^{(1)}\ = 0$ for $j = g + 1, \dots, d$, it suffices to show that

$$pr(|S\ (\check{\eta})| > \lambda_n \text{ for some } l = q_n + 1, \dots, p_n) \to 0, \quad (4)$$

and

$$pr(||S^*(\check{\eta})|| > \lambda_{\ n} \text{ for some } j = g + 1, \dots, d) \to 0. \quad (5)$$

Denote $d_n = (S_{q_n+1}(\check{\eta}), ..., S_{p_n}(\check{\eta}))^{\mathrm{T}} = -n^{-1}\{Z_M^{(2)}\}^{\mathrm{T}}(Y_M - \check{N}_1\check{\eta}_1)$, where

$$Y_M - \check{N}_1\check{\eta}_1 = \epsilon + \nu + (\tilde{N}_1 - \check{N}_1)\tilde{\eta}_{10} + \check{N}_1(\tilde{\eta}_{10} - \check{\eta}_1),$$

and $\nu = (\nu_1, ..., \nu_n)^{\mathrm{T}}$ with $\nu\ = \sum_{\ =1} \sum_{\ =\ n+1}^{\infty} \xi \quad b_{\ 0}$. From the proof of Lemma 4 and previous derivations, we have

$$\|\nu + (\tilde{N}_1 - \check{N}_1)\tilde{\eta}_{10} + \check{N}_1(\tilde{\eta}_{10} - \check{\eta}_1)\|^2 = O_p(r_n^2) = o_p(n\lambda_n^2).$$

Moreover, we have $\max_{\ =q_n+1 \quad p_n} \sum_{\ =1}^{n} z^2\ = O_p(n)$ by (B5). Thus

$$\begin{aligned} &\|n^{-1}\{Z_M^{(2)}\}^{\mathrm{T}}(\nu + (\tilde{N}_1 - \check{N}_1)\tilde{\eta}_{10} + \check{N}_1(\tilde{\eta}_{10} - \check{\eta}_1))\|_\infty \\ \leq\ & n^{-1}\Big(\max_{\ =q_n+1 \quad p_n} \sum_{\ =1}^{n} z^2\Big)^{1/2}\|\nu + (\tilde{N}_1 - \check{N}_1)\tilde{\eta}_{10} + \check{N}_1(\tilde{\eta}_{10} - \check{\eta}_1)\| = o_p(\lambda_n). \end{aligned}$$

Next, we denote $n^{-1/2}\{Z_M^{(2)}\}^{\mathrm{T}}\epsilon = (\delta_{q_n+1}, ..., \delta_{p_n})^{\mathrm{T}}$ with $\delta\ = n^{-1/2}\sum_{\ =1}^{n} z\ \ \epsilon$ for $l = q_n + 1, \dots, p_n$. As $\{\epsilon\ , i = 1, \dots, n\}$ and $\{z\ , i = 1, \dots, n\}$ are subGaussian random variables and independent of each other, combined with (B5), $\{\delta\ , l = 1, \dots, p_n\}$ are subexponential random variables satisfying that, for any constant $C > 0$, there exists positive constants C_1 and C_2 that do not depend on l,

$$pr(|\delta\ | > Cn^{1/2}\lambda_n) \leq C_2 \exp(-2^{-1}C_1 n^{1/2}\lambda_n),$$

and it follows that

$$\begin{aligned} pr(\|n^{-1/2}\{Z_M^{(2)}\}^{\mathrm{T}}\epsilon\|_\infty > Cn^{1/2}\lambda_n) &\leq \sum_{\ =q_n+1}^{p_n} pr(|\delta\ | > Cn^{1/2}\lambda_n) \\ &\leq O\{p_n \exp(-2^{-1}C_1 n^{1/2}\lambda_n)\} = O\{\exp(n\ \ \ - 2^{-1}C_1 n^{1/2}\lambda_n)\} = o(1), \end{aligned}$$

provided that (A6) and (A7) hold. Hence, it is obvious that $\|n^{-1}\{Z_M^{(2)}\}^{\mathrm{T}}\epsilon\|_\infty = o_p(\lambda_n)$. In conclusion, we have $\|d_n\|_\infty \le \|n^{-1}\{Z_M^{(2)}\}^{\mathrm{T}}\epsilon\|_\infty + \|n^{-1}\{Z_M^{(2)}\}^{\mathrm{T}}\{\nu + (\tilde{N}_1 - \check{N}_1)\tilde{\eta}_{10} + \check{N}_1(\tilde{\eta}_{10} - \check{\eta}_1)\}\|_\infty = o_p(\lambda_n)$ and this implies that (4) holds.

Next, we start to show (5). For $j = g+1, \dots, d$, we have $S^*(\check{\eta}) = -n^{-1}\hat{M}^{\mathrm{T}}(Y_M - \check{N}_1\check{\eta}_1) = (S_{\ 1}, ..., S_{\ n})^{\mathrm{T}}$ such that $\sum_{\ =1}^{n} S^2$ is bounded by

$$2\sum_{=1}^{n}\left(n^{-1}\sum_{=1}^{n}\hat{\xi}\quad\epsilon\ \right)^2 + 2n^{-2}\sum_{=1}^{n}\sum_{=1}^{n}\hat{\xi}^2\ \ \|\nu + (\tilde{N}_1 - \check{N}_1)\tilde{\eta}_{10} + \check{N}_1(\tilde{\eta}_{10} - \check{\eta}_1)\|^2.$$

For $\sum_{\ =1}^{n}\left(n^{-1}\sum_{=1}^{n}\hat{\xi}\quad\epsilon\ \right)^2$, we have

$$E\{\sum_{=1}^{n}\left(n^{-1}\sum_{=1}^{n}\hat{\xi}\quad\epsilon\ \right)^2\} = n^{-2}\sum_{=1}^{n}\sum_{=1}^{n}\sigma^2 E(\hat{\xi}^2\ \ ) \le n^{-2}\sum_{=1}^{n}\sum_{=1}^{n}\sigma^2 E(\|\hat{x}\ \ \|^2) = O(s_n n^{-1}).$$

It is easy to see that $\sum_{\ =1}^{n}\sum_{=1}^{n}\hat{\xi}^2\ \ = O_p(n\sum_{\ =1}^{n} w\ \ ) = O_p(n)$. Thus

$$\sum_{=1}^{n} S^2\ \ = O_p(s_n n^{-1}) + n^{-2}O_p(n)O_p(r_n^2) = O_p\{(q_n + s_n)n^{-1}\},$$

and it follows that

$$\|S^*(\check{\eta})\|^2 = \sum_{=1}^{n} S^2\ \ = O_p\{n^{-1}(q_n + s_n)\} = o_p(\lambda_n^2),$$

under (A7), which entails (5) immediately. Hence, we conclude that $\check{\eta}$ is a local minimizer of $Q_n(\tilde{\eta})$ such that $||\hat{\gamma} - \gamma_0|| = O_p(r_n n^{-1/2})$, $||\check{b}^{(1)} - \tilde{b}_0^{(1)}|| = O_p(r_n n^{-1/2})$, $||\hat{b}^{(1)} - b_0^{(1)}|| = O_p(r_n s_n^{\ /2} n^{-1/2})$, and $pr(\hat{b}^{(1)} = 0, j = g+1, \dots, d, \hat{\gamma}_2 = 0) \to 1$. This completes the proof.

where the last equality is by Lemma 2 (b), and 1_{q_n} is the vector of ones with dimension q_n. For $\nabla T_{2n}(\check{\eta})$,

$$\begin{aligned}\nabla T_{2n}(\check{\eta}) &= 2\sum_{=1}^{d}(\vartheta^{\ (2)} + \vartheta^{\ (3)} + \vartheta^{\ (4)}) \\ &= 2\sum_{=1}^{d}(\vartheta^{\ (2)} + \vartheta^{\ (3)} + \vartheta^{\ (5)} + \vartheta^{\ (6)} + \vartheta^{\ (7)} + \vartheta^{\ (8)}) \\ &= o_p(n^{1/2})1_{q_n} + 2\sum_{=1}^{d}\vartheta^{\ (8)},\end{aligned}$$

where the last two equalities are by Lemma 2 (b). Notice that $\vartheta^{\ (8)} = 0$ for $j = g+1, \ldots, d$,

$$\begin{aligned}\nabla Q_n(\check{\eta}) &= \nabla L_n(\check{\eta}) + \nabla T_{1n}(\check{\eta}) + \nabla T_{2n}(\check{\eta}) + \nabla T_{3n}(\check{\eta}) + \nabla T_{4n}(\check{\eta}) \\ &= o_p(n^{1/2})1_{q_n} + 2\sum_{=1}\vartheta^{\ (8)} + (-2\sum_{=1}^{n}\epsilon\ z^{(1)}) \\ &\quad +2\sum_{=1}^{n} z^{(1)} z^{(1)^{\mathrm{T}}}(\hat{\gamma}^{(1)} - \gamma_0^{(1)}) = 0,\end{aligned}$$

and it follows that

$$\begin{aligned}n^{-1/2}\sum_{=1}^{n} z^{(1)} z^{(1)^{\mathrm{T}}}(\hat{\gamma}^{(1)} - \gamma_0^{(1)}) &= n^{-1/2}\sum_{=1}^{n}\epsilon\ z^{(1)} - n^{-1/2}\sum_{=1}\vartheta^{\ (8)} + o_p(1)1_{q_n} \\ &= \sum_{=1}^{n} W_{\ n} + o_p(1)1_{q_n},\end{aligned}$$

with $W_{\ n} = n^{-1/2}\epsilon\ z^{(1)} - n^{-1/2}\sum_{=1}\sum_{=1}^{n}\sum_{\neq} b_{\ \ 0}(w\ \ - w\ \ )^{-1}\langle\Xi\ ,\phi\ \ \rangle\int(x\ \ \otimes x\ \ - K\ )\phi\ \ \phi\ \ $. It is obvious that $\{W_{\ n}, i = 1, \ldots, n\}$ are independently and identically distributed. with zero mean and $\mathrm{cov}(W_{\ n}) = n^{-1}(\sigma^2\Sigma_1 + B_n)$. It satisfies to show that $V_n^{-1/2}A_n\sum_{=1}^{n} W_{\ n}$ converges to $N(0, I)$ in distribution, where $V_n = A_n(\sigma^2\Sigma_1 + B_n)A_n^{\mathrm{T}} \to V^* = \sigma^2 H^* + B^*$, and V^* is positive definite. We start to show this by Linderberg Feller theorem.

First, we denote $\Pi_{\ n} = V_n^{-1/2}A_n W_{\ n}$, and it is trivial that $\{\Pi_{\ n}, i = 1, \ldots, n\}$ are independently and identically distributed centered random vectors with $\mathrm{cov}(\sum_{=1}^{n}\Pi_{\ n}) = I$. Second, for any $\zeta > 0$, we have

$$\sum_{=1}^{n} E\big[\|\Pi_{\ n}\|_2^2 1_{q_n}\{\|\Pi_{\ n}\| > \zeta\}\big] \le nE(\|\Pi_{1n}\|^4)^{1/2}\{pr(\|\Pi_{1n}\| > \zeta)\}^{1/2},$$

where

$$\begin{aligned}\|\Pi_{1n}\|^2 = \Pi_{1n}^{\mathrm{T}}\Pi_{1n} = W_{1n}^{\mathrm{T}}A_n^{\mathrm{T}}V_n^{-1}A_n W_{1n} &\le \lambda_{\max}(A_n^{\mathrm{T}}V_n^{-1}A_n)\|W_{1n}\|^2 \\ \le \lambda_{\max}(V_n^{-1})\lambda_{\max}(A_n^{\mathrm{T}}A_n)\|W_{1n}\|^2 &= O(1)\|W_{1n}\|^2.\end{aligned}$$

It follows that $E\big(\|\Pi_{1n}\|^4\big) = O(1)E\big(\|W_{1n}\|^4\big) = O(1)q_n^2/n^2$. Moreover,

$$\begin{aligned} pr(\|\Pi_{1n}\| > \zeta) &= pr(\|V_n^{-1/2}A_nW_{1n}\| > \zeta) \le E(\|V_n^{-1/2}A_nW_{1n}\|^2)/\zeta^2 \\ &\le \lambda_{\max}(A_n^{\mathrm{T}}V_n^{-1}A_n)E(\|W_{1n}\|^2)/\zeta^2 \\ &= O(1)E(\|W_{1n}\|^2) = O(1)q_n/n. \end{aligned}$$

By previous derivations and assumptions, we know

$$\begin{aligned} \sum_{=1}^{n} E\big[\|\Pi_{\ n}\|^2 1_{q_n}\{\|\Pi_{\ n}\| > \zeta\}\big] &= nO(q_n/n)O(q_nn^{-1/2}) \\ &= O(q_n^{3/2}n^{-1/2}) = o(1), \end{aligned}$$

since $q_n = o(n^{1/3})$. By Linderberg Feller theorem, we conclude that $V_n^{-1/2}A_n\sum_{=1}^{n} W_{\ n}$ converges to $N(0, I)$ in distribution, which completes the proof.

REFERENCES