Supplementary materials for "High-dimensional varying index coefficient quantile regression model"

Jing Lv[1] and Jialiang Li[2*]

[1]*School of Mathematics and Statistics,*

Southwest University, Chongqing, China

[2]*Department of Statistics and Applied Probability,*

National University of Singapore, Singapore

APPENDIX A: THE PROCEDURE OF GENERATING INITIAL VALUES IN SECTION 2.1

We propose the procedure to generate initial values. Specifically, the algorithm is described in the following.

Step 0. Suppose that functions $m_l(\cdot)$ are linear functions for $l = 1, ..., d$, that is, $Q_\tau(Y|\boldsymbol{X}_i, \boldsymbol{Z}_i) = \sum_{l=1}^{d} m_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l\right) X_{il} = \sum_{l=1}^{d}\left(a_l + b_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l\right)\right) X_{il}$. Let $\boldsymbol{\vartheta} = \left(\boldsymbol{\varsigma}^T, \boldsymbol{\nu}_1^T, ..., \boldsymbol{\nu}_d^T\right)^T$, $\boldsymbol{\varsigma} = (a_1, ..., a_d)^T$, $\boldsymbol{\nu}_l = b_l\boldsymbol{\beta}_l$ for $l = 1, ..., d$ and $\boldsymbol{\Pi}_i = \left(X_{i1}, ..., X_{id}, \boldsymbol{Z}_i^T X_{i1}, ..., \boldsymbol{Z}_i^T X_{id}\right)^T$. We obtain the estimator of $\boldsymbol{\vartheta}$ by minimizing

$$\begin{aligned}
\tilde{\boldsymbol{\vartheta}} &= \arg\min_{\boldsymbol{\vartheta}} \mathcal{Q}_n(\boldsymbol{\vartheta}) \\
&\triangleq \arg\min_{\boldsymbol{\vartheta}} \sum_{i=1}^{n} \rho_\tau\left\{Y_i - \sum_{l=1}^{d}\left(a_l + b_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l\right)\right) X_{il}\right\} \\
&= \arg\min_{\boldsymbol{\vartheta}} \sum_{i=1}^{n} \rho_\tau\left\{Y_i - \boldsymbol{\Pi}_i^T\boldsymbol{\vartheta}\right\},
\end{aligned}$$

and obtain $\tilde{\boldsymbol{\beta}}^0 = (\tilde{\boldsymbol{\beta}}_1^{0T}, ..., \tilde{\boldsymbol{\beta}}_d^{0T})^T$ with $\tilde{\boldsymbol{\beta}}_l^0 = (\tilde{\boldsymbol{\nu}}_l / \|\tilde{\boldsymbol{\nu}}_l\|_2)\, sgn(\tilde{\nu}_{1l}), l = 1, ..., d$ and $\tilde{\nu}_{1l}$ being the first component of $\tilde{\boldsymbol{\nu}}_l$.

Step 1. For a given $\tilde{\boldsymbol{\beta}}^0$, $\check{\boldsymbol{\lambda}}(\tilde{\boldsymbol{\beta}}^0)$ can be attained by $\check{\boldsymbol{\lambda}}\left(\tilde{\boldsymbol{\beta}}^0\right) = \arg\min_{\boldsymbol{\lambda}\in\mathbb{R}^{dJ_n}} \mathcal{L}_{\tau n}\left(\boldsymbol{\lambda}, \tilde{\boldsymbol{\beta}}^0\right)$ with

$\mathcal{L}_{\tau n}(\boldsymbol{\lambda}, \boldsymbol{\beta}) = \sum_{i=1}^{n} \rho_\tau \left\{ Y_i - \sum_{l=1}^{d} \boldsymbol{B}\left(\boldsymbol{Z}_i^T \boldsymbol{\beta}_l\right)^T \boldsymbol{\lambda}_l X_{il} \right\}$. Thus, we obtain the initial estimates of the nonparametric functions $\tilde{m}_l(\cdot, \tilde{\boldsymbol{\beta}}^0) = \boldsymbol{B}(\cdot)^T \check{\boldsymbol{\lambda}}_l(\tilde{\boldsymbol{\beta}}^0)$ for $l = 1, ..., d$.

Step 2. Then we can obtain the updated value of $\boldsymbol{\beta}$ by minimizing the objective function $\mathcal{C}_{\tau n}(\boldsymbol{\beta}) \triangleq \sum_{i=1}^{n} \rho_\tau \left\{ Y_i - \sum_{l=1}^{d} \tilde{m}_l\left(\boldsymbol{Z}_i^T \boldsymbol{\beta}_l\right)^T X_{il} \right\}$ with respect to $\boldsymbol{\beta}$ through nonlinear optimization. For example, the "ucminf" or "optim" function in R software can be applied to find the minimum of $\mathcal{C}_{\tau n}$ by numerical computing approaches. Let $\check{\boldsymbol{\beta}}$ be the minimum of $\mathcal{C}_{\tau n}$ and $\tilde{\boldsymbol{\beta}}_l = \left(\check{\boldsymbol{\beta}}_l / \left\|\check{\boldsymbol{\beta}}_l\right\|_2\right) sgn\left(\check{\beta}_{1l}\right)$ for $l = 1, ..., d$, where $\check{\beta}_{1l}$ is the first component of $\check{\boldsymbol{\beta}}_l$.

Step 3. Repeat Steps 1 and 2 until convergence, and denote the convergent values as $\tilde{\boldsymbol{\beta}}^{ini}$ and $\check{\boldsymbol{\lambda}}^{ini}$.

The above algorithm is used to generate the initial values of $\boldsymbol{\beta}$. For investigating the convergence rate of $\tilde{\boldsymbol{\beta}}^{ini}$ and $\hat{\boldsymbol{\beta}}$, Table 1 reports the 25%, 50%, 75%, quantiles of the number of iterations of $\tilde{\boldsymbol{\beta}}^{ini}$ and $\hat{\boldsymbol{\beta}}$ for Examples 1 and 2, where $\hat{\boldsymbol{\beta}}$ is the proposed estimator given in the subsection 2.1 by using $\tilde{\boldsymbol{\beta}}^{ini}$ as the initial values. Here the maximum number of iterations is taken as 100 in our algorithm. We can see from Table S1 that $\tilde{\boldsymbol{\beta}}^{ini}$ and $\hat{\boldsymbol{\beta}}$ can be converged with a few iterations, indicating our proposed iterative algorithm works well in practice.

Table S1: The 25%, 50%, 75% quantiles of the number of iterations of $\tilde{\boldsymbol{\beta}}^{ini}$ and $\hat{\boldsymbol{\beta}}$ for examples 1-2.

			$\tilde{\boldsymbol{\beta}}^{ini}$			$\hat{\boldsymbol{\beta}}$		
			25%	50%	75%	25%	50%	75%
example 1								
SN	500	$\tau = 0.5$	7	9	12	9	11	18
	1500	$\tau = 0.5$	6	8	11	7	9	14
t_3	500	$\tau = 0.5$	6	9	13	10	14	23
	1500	$\tau = 0.5$	6	8	11	7	9	13
MN	500	$\tau = 0.5$	6	8	13	10	14	34
	1500	$\tau = 0.5$	6	8	12	7	9	13
LA	500	$\tau = 0.5$	6	9	13	9	13	27
	1500	$\tau = 0.5$	6	8	12	7	8	12
example 2								
	500	$\tau = 0.5$	5	7	11	13	21	75
		$\tau = 0.75$	4	6	8	11	16	29
	1500	$\tau = 0.5$	5	7	10	8	10	15
		$\tau = 0.75$	4	6	8	8	10	14

APPENDIX B: ADDITIONAL NUMERICAL RESULTS

The tables and figures of the Appendix B provide numerical results.

Table S2: Simulation results ($\times 10^{-2}$) of Bias, MAD, ESD and ASD for SN for $\tau = 0.5$ in example 1.

	LS								QR							
	$n = 500$				$n = 1500$				$n = 500$				$n = 1500$			
	Bias	MAD	ESD	ASD	Bias	MAD	ESD	ASD	Bias	MAD	ESD	ASD	Bias	MAD	ESD	ASD
β_{11}	-0.287	3.170	4.165	3.122	0.136	1.566	2.009	1.685	-0.229	3.398	4.442	3.642	0.097	1.690	2.144	1.835
β_{12}	0.550	3.871	4.926	4.202	0.076	2.253	2.819	2.294	0.505	4.284	5.392	4.904	0.136	2.514	3.121	2.467
β_{13}	-0.304	2.225	2.867	2.067	-0.202	1.104	1.374	1.158	-0.380	2.345	3.088	2.399	-0.212	1.154	1.418	1.248
β_{21}	-0.295	1.484	1.816	1.683	0.072	0.776	0.986	0.968	-0.255	1.673	2.051	2.133	0.039	0.859	1.102	1.184
β_{22}	0.274	2.443	2.941	2.588	-0.055	1.302	1.636	1.507	0.076	2.792	3.427	3.281	-0.037	1.497	1.860	1.849
β_{23}	-0.062	2.382	3.049	2.954	-0.223	1.278	1.556	1.687	0.099	2.698	3.392	3.670	-0.189	1.440	1.767	2.044
β_{31}	0.042	0.965	1.273	1.076	-0.069	0.499	0.627	0.610	-0.020	1.178	1.475	1.325	-0.068	0.602	0.749	0.711
β_{32}	-0.081	0.750	0.988	0.906	0.033	0.404	0.509	0.514	-0.065	0.904	1.138	1.122	0.032	0.474	0.595	0.601
β_{33}	0.079	1.007	1.286	1.276	0.017	0.583	0.733	0.728	0.124	1.212	1.548	1.571	0.008	0.664	0.830	0.854

Table S3: Simulation results ($\times 10^{-2}$) of Bias, MAD, ESD and ASD for t_3 with $\tau = 0.5$ in example 1.

	LS								QR							
	$n = 500$				$n = 1500$				$n = 500$				$n = 1500$			
	Bias	MAD	ESD	ASD	Bias	MAD	ESD	ASD	Bias	MAD	ESD	ASD	Bias	MAD	ESD	ASD
β_{11}	-0.242	4.778	6.261	4.747	-0.168	2.545	3.241	2.691	-0.109	3.892	5.099	4.056	0.068	2.031	2.582	2.085
β_{12}	0.237	7.357	9.209	6.486	0.167	3.439	4.509	3.666	-0.116	5.870	7.360	5.581	0.332	2.883	3.585	2.853
β_{13}	-0.800	3.276	4.130	3.298	-0.169	1.818	2.298	1.822	-0.473	2.862	3.667	2.771	-0.299	1.407	1.787	1.421
β_{21}	-0.377	2.400	3.016	2.736	0.026	1.343	1.691	1.625	-0.223	1.928	2.420	2.381	0.035	0.952	1.189	1.291
β_{22}	0.189	3.850	4.784	4.208	-0.042	2.062	2.650	2.511	0.008	3.073	3.830	3.720	-0.049	1.562	1.969	2.020
β_{23}	-0.250	3.627	4.631	4.775	-0.294	1.969	2.468	2.839	0.036	2.766	3.544	4.084	-0.178	1.570	1.953	2.207
β_{31}	-0.010	1.583	2.029	1.747	0.025	0.942	1.201	1.034	-0.033	1.234	1.586	1.466	-0.021	0.657	0.816	0.793
β_{32}	-0.077	1.276	1.623	1.491	-0.017	0.760	0.958	0.869	-0.007	1.033	1.330	1.250	0.021	0.510	0.644	0.667
β_{33}	0.028	1.779	2.279	2.109	-0.070	0.899	1.193	1.214	-0.054	1.449	1.826	1.765	-0.055	0.695	0.875	0.946

Table S4: Simulation results ($\times 10^{-2}$) of Bias, MAD, ESD and ASD for MN with $\tau = 0.5$ in example 1.

	LS								QR							
	$n=500$				$n=1500$				$n=500$				$n=1500$			
	Bias	MAD	ESD	ASD	Bias	MAD	ESD	ASD	Bias	MAD	ESD	ASD	Bias	MAD	ESD	ASD
β_{11}	-0.385	5.215	6.554	4.984	-0.257	2.660	3.474	2.832	-0.377	3.699	4.649	3.823	-0.073	1.955	2.433	1.973
β_{12}	0.221	7.526	9.434	7.039	0.220	3.783	4.923	3.909	0.254	4.850	6.193	5.327	0.348	2.646	3.379	2.732
β_{13}	-0.793	3.848	4.948	3.569	-0.168	1.962	2.536	1.963	-0.273	2.585	3.263	2.686	-0.197	1.428	1.805	1.339
β_{21}	-0.470	2.674	3.513	2.991	-0.180	1.440	1.805	1.768	-0.218	1.838	2.327	2.338	-0.069	0.982	1.228	1.295
β_{22}	-0.292	4.200	5.331	4.572	0.245	2.375	2.960	2.736	-0.108	2.983	3.710	3.609	0.137	1.540	1.949	2.008
β_{23}	0.808	3.809	4.686	5.136	-0.326	2.176	2.833	3.048	0.318	2.548	3.198	4.001	-0.231	1.478	1.831	2.195
β_{31}	0.017	1.837	2.327	1.941	0.122	0.978	1.211	1.071	-0.004	1.325	1.700	1.437	0.021	0.704	0.883	0.767
β_{32}	-0.053	1.446	1.890	1.662	-0.099	0.741	0.938	0.912	-0.017	1.047	1.343	1.231	-0.037	0.544	0.689	0.656
β_{33}	-0.154	1.841	2.452	2.336	-0.016	0.931	1.186	1.296	-0.087	1.393	1.764	1.723	0.032	0.701	0.871	0.913

Table S5: Simulation results ($\times 10^{-2}$) of Bias, MAD, ESD and ASD for LA with $\tau = 0.5$ in example 1.

	LS								QR							
	$n=500$				$n=1500$				$n=500$				$n=1500$			
	Bias	MAD	ESD	ASD	Bias	MAD	ESD	ASD	Bias	MAD	ESD	ASD	Bias	MAD	ESD	ASD
β_{11}	-0.393	4.048	5.074	3.965	0.249	1.942	2.487	2.245	-0.350	3.760	4.719	3.747	0.259	1.689	2.106	1.887
β_{12}	-0.345	5.588	7.059	5.527	-0.207	2.677	3.410	3.054	-0.200	4.932	6.222	5.327	-0.298	2.285	2.930	2.544
β_{13}	-0.171	2.703	3.528	2.689	-0.227	1.392	1.735	1.546	-0.144	2.486	3.223	2.474	-0.168	1.139	1.418	1.297
β_{21}	-0.136	2.166	2.671	2.337	0.019	1.093	1.367	1.350	0.033	1.626	2.092	2.221	0.120	0.906	1.109	1.157
β_{22}	-0.338	3.395	4.226	3.591	0.045	1.805	2.302	2.096	-0.448	2.640	3.352	3.453	-0.106	1.485	1.842	1.797
β_{23}	0.377	2.887	3.575	4.085	-0.366	1.724	2.111	2.333	0.341	2.329	2.918	3.856	-0.290	1.333	1.676	1.976
β_{31}	-0.008	1.316	1.705	1.492	0.026	0.769	0.954	0.843	-0.097	1.093	1.424	1.406	0.017	0.608	0.759	0.707
β_{32}	-0.014	1.112	1.412	1.259	-0.042	0.617	0.751	0.720	0.063	0.942	1.221	1.196	-0.024	0.466	0.585	0.601
β_{33}	-0.104	1.506	1.930	1.803	0.028	0.784	0.976	1.013	-0.119	1.412	1.799	1.716	0.009	0.629	0.793	0.848

Table S6: Simulation results of RASE for m_1, m_2, m_3 with $\tau = 0.5$ in example 1.

		LS			QR		
		m_1	m_2	m_3	m_1	m_2	m_3
500	SN	0.098	0.127	0.102	0.110	0.145	0.123
	t_3	0.140	0.176	0.156	0.128	0.163	0.146
	MN	0.148	0.187	0.167	0.119	0.154	0.130
	LA	0.119	0.146	0.141	0.114	0.139	0.134
1500	SN	0.072	0.098	0.061	0.082	0.107	0.069
	t_3	0.094	0.121	0.094	0.092	0.113	0.078
	MN	0.099	0.126	0.099	0.088	0.112	0.083
	LA	0.084	0.113	0.081	0.083	0.107	0.073

Table S7: Simulation results of Bias, MAD, ESD and ASD for $\boldsymbol{\beta}_\tau$ with $\tau = 0.5, 0.75$ and $n = 500, 1500$ in example 2.

	$n = 500$								$n = 1500$							
	$\tau = 0.5$				$\tau = 0.75$				$\tau = 0.5$				$\tau = 0.75$			
	Bias	MAD	ESD	ASD	Bias	MAD	ESD	ASD	Bias	MAD	ESD	ASD	Bias	MAD	ESD	ASD
$\beta_{\tau,11}$	-0.017	0.064	0.078	0.083	0.078	0.085	0.064	0.062	-0.006	0.035	0.043	0.044	0.080	0.080	0.041	0.035
$\beta_{\tau,12}$	-0.001	0.071	0.088	0.088	0.003	0.062	0.078	0.067	-0.004	0.036	0.046	0.047	-0.010	0.035	0.045	0.038
$\beta_{\tau,13}$	0.001	0.052	0.067	0.065	-0.061	0.067	0.056	0.051	0.003	0.024	0.030	0.034	-0.050	0.052	0.037	0.028
$\beta_{\tau,21}$	-0.007	0.181	0.221	0.143	-0.061	0.189	0.218	0.120	-0.036	0.108	0.129	0.097	-0.076	0.118	0.122	0.082
$\beta_{\tau,22}$	-0.052	0.189	0.228	0.129	0.022	0.177	0.225	0.111	-0.036	0.098	0.117	0.092	0.026	0.092	0.114	0.077
$\beta_{\tau,23}$	-0.057	0.153	0.203	0.115	-0.068	0.144	0.174	0.094	0.015	0.079	0.099	0.073	-0.004	0.072	0.094	0.058
$\beta_{\tau,31}$	-0.015	0.086	0.112	0.083	-0.079	0.099	0.097	0.073	0.003	0.043	0.055	0.049	-0.062	0.068	0.056	0.046
$\beta_{\tau,32}$	-0.002	0.115	0.144	0.108	0.000	0.093	0.120	0.095	0.006	0.056	0.072	0.066	-0.001	0.052	0.064	0.059
$\beta_{\tau,33}$	-0.033	0.120	0.159	0.112	0.081	0.120	0.121	0.093	-0.021	0.056	0.069	0.068	0.083	0.090	0.067	0.059

Table S8: Simulation results of RASE for $m_{\tau,1}$, $m_{\tau,2}$, $m_{\tau,3}$ with $\tau = 0.5, 0.75$ and $n = 500, 1500$ in example 2.

	$\tau = 0.5$			$\tau = 0.75$		
	$m_{\tau,1}$	$m_{\tau,2}$	$m_{\tau,3}$	$m_{\tau,1}$	$m_{\tau,2}$	$m_{\tau,3}$
$n = 500$	0.066	0.077	0.104	0.100	0.191	0.216
$n = 1500$	0.042	0.042	0.057	0.066	0.170	0.199

Table S9: Simulation results of variable selection for $\boldsymbol{\beta}$ with $\tau = 0.1, 0.5, 0.75$ and 0.9 in example 3.

		$n=500, p_n=7$						$n=1500, p_n=11$					
		C	IC	CF	O.MSE	P.MSE	U.MSE	C	IC	CF	O.MSE	P.MSE	U.MSE
$\tau=0.1$	SN	11.34	0.000	0.755	2.041	2.188	3.847	19.65	0.000	0.825	0.743	0.769	1.415
	t_3	11.26	0.020	0.690	3.821	5.025	7.794	19.73	0.0000	0.900	1.610	1.688	3.147
	MN	11.43	0.020	0.720	3.587	5.183	7.938	19.79	0.005	0.900	1.450	2.124	3.517
	LA	11.08	0.005	0.615	3.834	5.150	7.801	19.41	0.005	0.825	1.482	2.215	3.505
$\tau=0.5$	SN	11.79	0.000	0.875	1.416	1.459	2.603	19.98	0.000	0.975	0.459	0.459	0.922
	t_3	11.85	0.000	0.925	1.834	1.878	3.542	19.99	0.000	0.990	0.618	0.618	1.285
	MN	11.90	0.000	0.930	1.784	1.795	3.595	19.99	0.000	0.985	0.613	0.613	1.193
	LA	11.85	0.000	0.890	1.852	1.892	3.677	19.98	0.000	0.980	0.574	0.574	1.165
$\tau=0.75$	SN	11.66	0.005	0.850	1.515	1.728	3.013	19.94	0.000	0.955	0.529	0.529	1.048
	t_3	11.78	0.000	0.875	2.133	2.175	4.323	19.97	0.000	0.970	0.796	0.799	1.627
	MN	11.76	0.000	0.840	2.092	2.135	4.051	19.96	0.000	0.965	0.687	0.688	1.404
	LA	11.72	0.000	0.840	2.092	2.192	4.270	19.91	0.000	0.960	0.741	0.747	1.494
$\tau=0.9$	SN	11.15	0.005	0.705	1.831	2.277	4.075	19.72	0.000	0.870	0.676	0.699	1.358
	t_3	11.26	0.010	0.715	3.518	5.306	8.591	19.69	0.000	0.905	1.627	1.790	3.183
	MN	11.13	0.035	0.680	3.525	7.671	10.26	19.79	0.000	0.885	1.356	1.403	2.740
	LA	11.25	0.010	0.705	3.456	4.549	7.139	19.53	0.000	0.835	1.585	2.562	3.902

Notation: the values of last three columns multiplied by 10^{-2} are true simulation results of O.MSE, P.MSE and U.MSE. In addition the number of zero coefficients is 12 for $n = 500$ and 20 for $n = 1500$.

Table S10: Simulations results of linear component identification for $m_l, l = 1, 2, 3, 4$ with $\tau = 0.1, 0.5, 0.75$ and 0.9 in example 3.

		$\tau = 0.1$					$\tau = 0.5$				
		ILC_1	ILC_2	ILC_3	ILC_4	CIL	ILC_1	ILC_2	ILC_3	ILC_4	CIL
$n = 500$	SN	0.000	0.000	0.440	0.525	0.340	0.000	0.000	0.750	0.715	0.600
	t_3	0.000	0.000	0.540	0.540	0.430	0.000	0.000	0.820	0.845	0.740
	MN	0.000	0.005	0.605	0.630	0.470	0.000	0.000	0.780	0.845	0.710
	LA	0.000	0.000	0.500	0.480	0.330	0.000	0.000	0.730	0.815	0.675
$n = 1500$	SN	0.000	0.000	0.775	0.785	0.660	0.000	0.000	0.935	0.945	0.895
	t_3	0.000	0.000	0.805	0.810	0.745	0.000	0.000	0.985	0.990	0.975
	MN	0.000	0.000	0.825	0.900	0.780	0.000	0.000	0.960	0.955	0.925
	LA	0.000	0.000	0.850	0.850	0.770	0.000	0.000	0.940	0.965	0.915
		$\tau = 0.75$					$\tau = 0.9$				
		ILC_1	ILC_2	ILC_3	ILC_4	CIL	ILC_1	ILC_2	ILC_3	ILC_4	CIL
$n = 500$	SN	0.000	0.000	0.620	0.665	0.485	0.000	0.000	0.570	0.545	0.400
	t_3	0.000	0.000	0.750	0.755	0.630	0.000	0.005	0.575	0.610	0.440
	MN	0.000	0.000	0.685	0.685	0.565	0.000	0.005	0.575	0.560	0.410
	LA	0.000	0.000	0.715	0.720	0.585	0.000	0.000	0.575	0.600	0.450
$n = 1500$	SN	0.000	0.000	0.900	0.920	0.870	0.000	0.000	0.725	0.770	0.620
	t_3	0.000	0.000	0.960	0.970	0.955	0.000	0.000	0.895	0.915	0.825
	MN	0.000	0.000	0.925	0.955	0.900	0.000	0.000	0.860	0.880	0.805
	LA	0.000	0.000	0.935	0.960	0.905	0.000	0.000	0.875	0.880	0.775

Table S11: Simulation results of RASE for $m_l, l = 1, 2, 3, 4$ with $\tau = 0.1, 0.5, 0.75$ and 0.9 in example 3.

		n = 500								n = 1500							
		P.RASE				U.RASE				P.RASE				U.RASE			
		m_1	m_2	m_3	m_4	m_1	m_2	m_3	m_4	m_1	m_2	m_3	m_4	m_1	m_2	m_3	m_4
$\tau = 0.1$	SN	0.295	0.686	0.222	0.186	0.282	0.701	0.273	0.237	0.138	0.249	0.055	0.052	0.129	0.264	0.095	0.088
	t_3	0.278	0.698	0.244	0.210	0.270	0.711	0.306	0.270	0.129	0.227	0.068	0.060	0.120	0.244	0.111	0.102
	MN	0.335	0.685	0.212	0.183	0.321	0.713	0.289	0.261	0.159	0.230	0.068	0.052	0.153	0.250	0.119	0.100
	LA	0.281	0.640	0.229	0.217	0.267	0.653	0.286	0.268	0.122	0.231	0.062	0.059	0.114	0.242	0.106	0.097
$\tau = 0.5$	SN	0.163	0.623	0.128	0.133	0.173	0.634	0.210	0.199	0.057	0.230	0.030	0.028	0.058	0.242	0.072	0.071
	t_3	0.157	0.597	0.121	0.102	0.167	0.612	0.221	0.192	0.059	0.217	0.028	0.024	0.063	0.233	0.073	0.068
	MN	0.176	0.584	0.126	0.098	0.176	0.607	0.232	0.196	0.059	0.223	0.030	0.029	0.062	0.237	0.073	0.069
	LA	0.175	0.617	0.161	0.124	0.175	0.634	0.240	0.216	0.053	0.227	0.031	0.026	0.055	0.241	0.072	0.067
$\tau = 0.75$	SN	0.220	0.697	0.166	0.143	0.219	0.710	0.239	0.215	0.082	0.233	0.037	0.035	0.083	0.248	0.078	0.074
	t_3	0.210	0.649	0.141	0.136	0.215	0.665	0.226	0.215	0.079	0.213	0.036	0.030	0.083	0.231	0.088	0.074
	MN	0.226	0.669	0.149	0.144	0.229	0.694	0.238	0.235	0.090	0.215	0.037	0.031	0.089	0.230	0.085	0.076
	LA	0.242	0.647	0.149	0.150	0.247	0.673	0.243	0.239	0.084	0.215	0.037	0.031	0.088	0.230	0.088	0.076
$\tau = 0.9$	SN	0.274	0.667	0.179	0.166	0.263	0.679	0.232	0.220	0.135	0.244	0.061	0.054	0.128	0.257	0.099	0.092
	t_3	0.273	0.618	0.202	0.178	0.261	0.637	0.268	0.248	0.125	0.218	0.056	0.056	0.116	0.235	0.107	0.104
	MN	0.312	0.654	0.232	0.189	0.304	0.676	0.303	0.248	0.156	0.215	0.057	0.054	0.153	0.231	0.113	0.101
	LA	0.277	0.669	0.190	0.175	0.274	0.686	0.248	0.235	0.120	0.214	0.060	0.054	0.116	0.227	0.105	0.095

Notation: P.RASE and U.RASE stand for the RASEs of penalized estimators and unpenalized estimators, respectively.

Table S12: The definitions of all variables in New Zealand workforce study.

Variable	Name	Description
		the intercept
X_1	*intercept*	the intercept
		binary variables
X_2	*sex*	a factor with levels F(1=female) and M (2=male)
X_3	*diabetes*	Do you have diabetes? (0=no,1=yes)
X_4	*nervous*	would you call yourself a nervous person?(0=no,1=yes)
		continuous variables
Z_1	*age*	age in years
Z_2	*cholest*	cholesterol (mmol/L)
Z_3	*dmd*	the largest number of drinks that you have a day
Z_4	*feethour*	the average time that you spend on your feet, either standing or moving
Z_5	*sleep*	the average hours that you usually sleep each night
Z_6	*sbp*	systolic blood pressure (mm Hg)
Z_7	*dbp*	diastolic blood pressure (mm Hg)
		the response variable
Y	*BMI*	the weight (kg) divided by the squared height (m)

Table S13: The estimated loading coefficients $\hat{\boldsymbol{\beta}}_l$ (unEST) and $\bar{\boldsymbol{\beta}}_l$ (penEST) with $l = 1, 2, 3, 4$ in New Zealand workforce study.

		τ=0.1		τ=0.25		τ=0.5		τ=0.75		τ=0.9	
		unEST	penEST	unEST	penEST	unEST	penEST	unEST	penEST	unEST	penEST
X_1=intercept											
$\boldsymbol{\beta}_1$	Z_1	0.273	0.335	0.459	0.852	0.384	0.386	0.292	0.836	0.366	0.288
	Z_2	0.379	0.379	0.515	0.524	0.542	0.542	0.472	0	0.374	0
	Z_3	0.588	0.588	0.354	0	0.460	0.460	0.307	0	0.161	0
	Z_4	0.365	0.364	0.184	0	0.038	0	-0.126	0	0.280	0
	Z_5	-0.132	0	-0.131	0	-0.195	-0.195	-0.190	0	-0.178	0
	Z_6	0.144	0	0.324	0	0.432	0.432	0.445	0	0.162	0
	Z_7	0.515	0.515	0.493	0	0.348	0.348	0.589	0.548	0.751	0.958
X_2=sex											
$\boldsymbol{\beta}_2$	Z_1	0.077	0.211	0.621	0.827	0.402	0.433	0.010	0.079	0.119	0
	Z_2	0.143	0	0.319	0	0.507	0.507	0.269	0	0.036	0.028
	Z_3	0.804	0.802	0.186	0	0.494	0.494	0.248	0	0.118	0.091
	Z_4	0.560	0.559	-0.038	0	-0.197	-0.197	-0.703	-0.754	-0.068	-0.052
	Z_5	0.084	0	0.359	0	-0.127	0	-0.036	0	0.411	0.321
	Z_6	0.074	0	0.174	0	0.523	0.523	0.601	0.652	0.305	0.238
	Z_7	0.030	0	-0.564	-0.562	-0.097	0	-0.098	0	-0.839	-0.910
X_3=diabetes											
$\boldsymbol{\beta}_3$	Z_1	0	0	0.229	0.680	0	0	0.001	0.748	0.001	0.465
	Z_2	-0.201	-0.201	-0.368	0	-0.544	-0.544	0.304	0	-0.182	0
	Z_3	0.238	0.238	0.425	0	-0.410	-0.410	0.743	0.663	0.309	0
	Z_4	0.719	0.719	0.515	0.514	0.405	0.405	-0.010	0	0.157	0
	Z_5	0.488	0.488	0.528	0.523	0.446	0.446	-0.506	0	0.330	0
	Z_6	-0.184	-0.184	-0.294	0	-0.374	-0.374	0.291	0	-0.134	0
	Z_7	0.338	0.338	-0.030	0	-0.182	-0.182	-0.121	0	0.849	0.885
X_4=nervous											
$\boldsymbol{\beta}_4$	Z_1	0.004	0.128	0.002	0.608	0.279	0.324	0.331	0.684	0.009	0.752
	Z_2	-0.375	-0.376	-0.339	0	-0.164	0	-0.473	0	-0.369	0
	Z_3	0.427	0.428	-0.408	0	-0.239	-0.239	-0.201	0	-0.646	-0.659
	Z_4	-0.649	-0.649	-0.070	0	0.401	0.401	0.130	0	0.067	0
	Z_5	0.487	0.487	-0.192	0	0.363	0.363	-0.219	0	-0.496	0
	Z_6	-0.095	0	0.823	0.794	0.695	0.695	0.160	0	-0.284	0
	Z_7	0.101	0	0.004	0	-0.250	-0.250	0.732	0.730	0.339	0

Table S14: The estimated values $\| \bar{\boldsymbol{\lambda}}_l \|_{\boldsymbol{D}}$ ($l = 1, 2, 3, 4$) by using the penalized $\bar{\boldsymbol{\lambda}}_l$ for $\tau = 0.1, 0.25, 0.5, 0.75$ and 0.9 in New Zealand workforce study.

	$\| \bar{\boldsymbol{\lambda}}_1 \|_{\boldsymbol{D}}$	$\| \bar{\boldsymbol{\lambda}}_2 \|_{\boldsymbol{D}}$	$\| \bar{\boldsymbol{\lambda}}_3 \|_{\boldsymbol{D}}$	$\| \bar{\boldsymbol{\lambda}}_4 \|_{\boldsymbol{D}}$
$\tau = 0.1$	$< 10^{-2}$	$< 10^{-2}$	80.02	14.17
$\tau = 0.25$	$< 10^{-2}$	$< 10^{-2}$	163.1	$< 10^{-2}$
$\tau = 0.5$	$< 10^{-2}$	$< 10^{-2}$	43.60	$< 10^{-2}$
$\tau = 0.75$	$< 10^{-2}$	$< 10^{-2}$	93.02	$< 10^{-2}$
$\tau = 0.9$	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$	$< 10^{-2}$

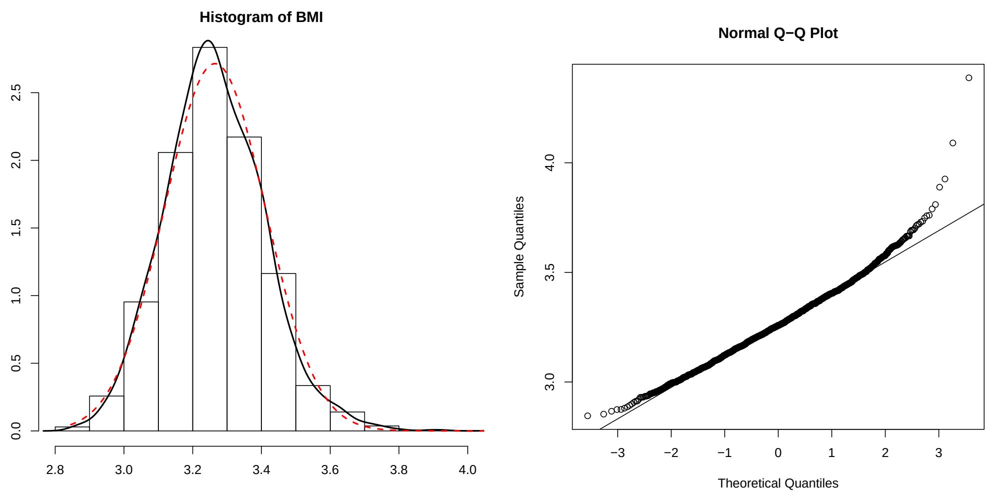

Figure S1: The histogram (left panel) and Q-Q plot (right panel) of BMI.

To assess how well the varying index coefficient model (VICM^1) and its penalized version (VICM^2) fit the data, we adopt the following model assessment tool (Wang et al. (2009)) by comparing the empirical distribution of Y with the simulated distribution from VICM^1 and VICM^2. Specifically, similar to the simulation example 2 of the subsection 5.2, the simulated Y^* and Y^{**} are generated as the uth conditional quantile $\sum_{l=1}^{d} \hat{m}_{u,l}(\boldsymbol{Z}^T\hat{\boldsymbol{\beta}}_{u,l})X_l$ and $\sum_{l=1}^{d} \bar{m}_{u,l}(\boldsymbol{Z}^T\bar{\boldsymbol{\beta}}_{u,l})X_l$ respectively, where $\hat{\boldsymbol{\beta}}_{u,l}$, $\hat{m}_{u,l}$ and $\bar{\boldsymbol{\beta}}_{u,l}$, $\bar{m}_{u,l}$ are uth quantile estimators in VICM^1 and VICM^2. Here we generate u from $U(0,1)$. Repeating this procedure many times, say the sample size n, we can obtain simulated samples Y^* and Y^{**} based on VICM^1 and VICM^2. Obviously, if VICM^1 (or VICM^2) fits data well, the marginal distribution of the simulated Y^* (or Y^{**}) should match that of the observed Y. Figure S2 shows estimated density curves and Q-Q plots of Y, Y^* and Y^{**}, suggest that our proposed models (VICM^1 and VICM^2) fit the data reasonably well. Moreover, the p values for VICM^1 and VICM^2 are obvious more than 0.05 by the well-known two-sample Kolmogorov-Smirnov test (Conover (1971)), indicating there is little difference for marginal distributions of Y and Y^* (or Y^{**}).

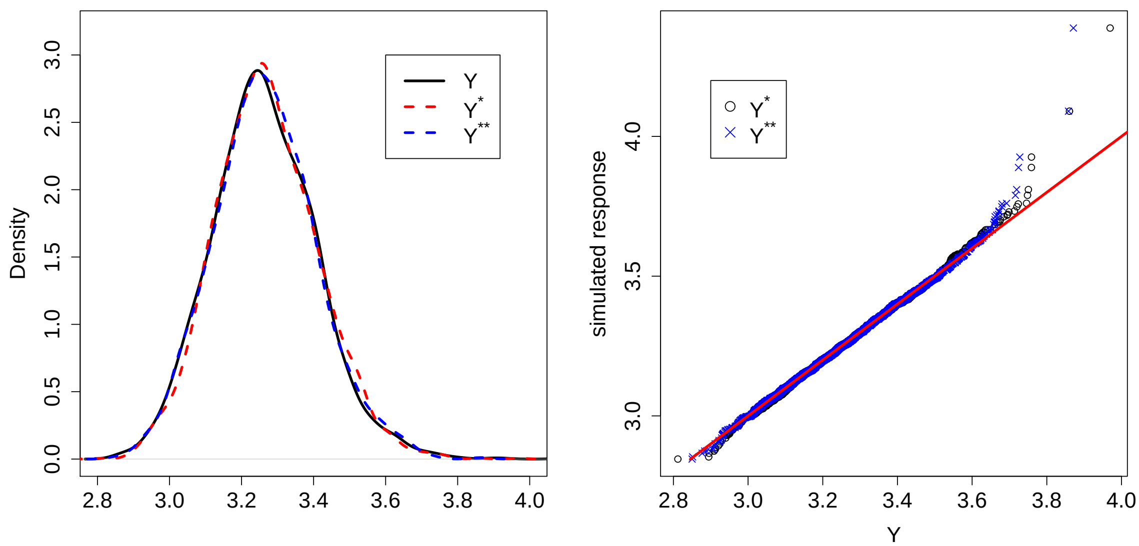

Figure S2: The estimated density curves and Q-Q plots of Y, Y^* and Y^{**} for the New Zealand workforce data. The diagonal line in the Q-Q plot is $y = x$.

APPENDIX C: PROOFS OF TECHNICAL LEMMAS AND THEOREMS

In this document, we present the proofs of technical lemmas and theorems. The technical lemmas are used to prove the theorems of the paper. We first present some notations that will be used in the proofs of lemmas and theorems. For any positive numbers a_n and b_n, let $a_n \asymp b_n$ denote that $\lim_{n\to\infty} a_n/b_n = c$ for a positive constant c. For any vector $\boldsymbol{\xi} = (\xi_1, ..., \xi_s)^T \in \mathbb{R}^s$, denote $\|\boldsymbol{\xi}\|_\infty = \max_{1\le l\le s}|\xi_l|$. For any symmetric matrix $\boldsymbol{A}_{s\times s}$, denote its L_r norm as $\|\boldsymbol{A}\|_r = \max_{\boldsymbol{\xi}\in\mathbb{R}^s, \boldsymbol{\xi}\neq 0}\|\boldsymbol{A}\boldsymbol{\xi}\|_r \|\boldsymbol{\xi}\|_r^{-1}$. For any matrix $\boldsymbol{A} = (A_{ij})_{i=1,j=1}^{s,t}$, denote $|\boldsymbol{A}| = \max_{1\le i\le s, 1\le j\le t}|A_{ij}|$ and $\|\boldsymbol{A}\|_\infty = \max_{1\le i\le s}\sum_{j=1}^{t}|A_{ij}|$. Define $(\boldsymbol{a}_l)_{l=1}^{d} = \left(\boldsymbol{a}_1^T, ..., \boldsymbol{a}_d^T\right)^T$ for any vector $\boldsymbol{a}_l, l = 1, ..., d$. Throughout the paper, C denotes a generic positive constant that can vary from line to line, $\rho_{min}(\cdot)$ and $\rho_{max}(\cdot)$ denote the minimum and maximum eigenvalues of a matrix, respectively and $tr(\cdot)$ is the trace operator of a matrix.

First, we give some properties of B-spline. According to the properties of normalized B-splines, we have that for each $l = 1, ..., d, s = 1, ..., J_n$, (i) $B_s(u_l) \ge 0$ and $\sum_{s=1}^{J_n} B_s(u_l) = 1$ for $u_l \in [a, b]$; (ii) there exist positive constants C_1 and C_2 such that for any $\boldsymbol{a} \in \mathbb{R}^{J_n}$,

$$\frac{C_1}{J_n}\boldsymbol{a}^T\boldsymbol{a} \le \int \boldsymbol{a}^T\boldsymbol{B}(u)\boldsymbol{B}^T(u)\boldsymbol{a}du \le \frac{C_2}{J_n}\boldsymbol{a}^T\boldsymbol{a}.$$

Define $\boldsymbol{D}(\boldsymbol{\beta}) = \left[(\boldsymbol{D}_1(\boldsymbol{\beta}), ..., \boldsymbol{D}_n(\boldsymbol{\beta}))^T\right]_{n\times dJ_n}$, $\boldsymbol{V}(\boldsymbol{\beta}) = E\left(\boldsymbol{D}_i(\boldsymbol{\beta})\boldsymbol{D}_i^T(\boldsymbol{\beta})\right)$ and $\hat{\boldsymbol{V}}(\boldsymbol{\beta}) = n^{-1}\boldsymbol{D}^T(\boldsymbol{\beta})\boldsymbol{D}(\boldsymbol{\beta})$, $\boldsymbol{D}_i(\boldsymbol{\beta}) = (D_{i,sl}(\boldsymbol{\beta}_l), 1 \le s \le J_n, 1 \le l \le d)^T$ with $D_{i,sl}(\boldsymbol{\beta}_l) = B_s(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l)X_{il}$.

Lemma 1. *Under conditions (C1) and (C3), for any vector $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1^T, ..., \boldsymbol{\alpha}_d^T)^T$ with $\boldsymbol{\alpha}_l = (\alpha_{sl}, 1 \le s \le J_n)^T$, there are constants $0 < c_V < C_V < \infty$, such that $\forall\boldsymbol{\beta} \in \Theta$ and for large enough n, we have*

$$(i)\, c_V J_n^{-1}\boldsymbol{\alpha}^T\boldsymbol{\alpha} \le \boldsymbol{\alpha}^T\boldsymbol{V}(\boldsymbol{\beta})\boldsymbol{\alpha} \le C_V J_n^{-1}\boldsymbol{\alpha}^T\boldsymbol{\alpha}, C_V^{-1}J_n\boldsymbol{\alpha}^T\boldsymbol{\alpha} \le \boldsymbol{\alpha}^T\boldsymbol{V}^{-1}(\boldsymbol{\beta})\boldsymbol{\alpha} \le c_V^{-1}J_n\boldsymbol{\alpha}^T\boldsymbol{\alpha},$$

$$(ii) \sup_{1\le s,s'\le J_n, 1\le l\le d}\left|n^{-1}\sum_{i=1}^{n}\left[D_{i,sl}(\boldsymbol{\beta}_l)D_{i,s'l}(\boldsymbol{\beta}_l)\right] - E\left\{D_{i,sl}(\boldsymbol{\beta}_l)D_{i,s'l}(\boldsymbol{\beta}_l)\right\}\right|$$
$$= O_{a.s.}\left(\sqrt{J_n^{-1}n^{-1}\log n}\right),$$

$$(iii) \sup_{1\le s,s'\le J_n, l\neq l'}\left|n^{-1}\sum_{i=1}^{n}\left[D_{i,sl}(\boldsymbol{\beta}_l)D_{i,s'l'}(\boldsymbol{\beta}_l)\right] - E\left\{D_{i,sl}(\boldsymbol{\beta}_l)D_{i,s'l'}(\boldsymbol{\beta}_l)\right\}\right|$$
$$= O_{a.s.}\left(J_n^{-1}\sqrt{n^{-1}\log n}\right).$$

Proof of Lemma 1 Lemma 1 follows from Lemma A.1 of Ma and Song (2015). □

The following lemma states the convergence rate of $\hat{m}_l(u_l, \boldsymbol{\beta}^0)$ and $\hat{\dot{m}}_l(u_l, \boldsymbol{\beta}^0)$ of the nonparametric function $m_l(u_l)$ and its first derivative $\dot{m}_l(u_l)$, for $l = 1, ..., d$.

Lemma 2. *Under conditions (C1)–(C3) and (C5), $J_n \to \infty$ and $n^{-1}J_n^3 = o(1)$, for $1 \leq l \leq d$, we have $|\hat{m}_l(u_l, \boldsymbol{\beta}^0) - m_l(u_l)| = O_p\left(\sqrt{J_n/n} + J_n^{-r}\right)$ uniformly for any $u_l \in [a, b]$, and $\left|\hat{\dot{m}}_l(u_l, \boldsymbol{\beta}^0) - \dot{m}_l(u_l)\right| = O_p\left(\sqrt{J_n^3/n} + J_n^{-r+1}\right)$ uniformly for any $u_l \in [a, b]$.*

Proof of Lemma 2. According to the result on page 149 of de Boor (2001), for $m_l(u_l)$ satisfying condition (C2), there exists a best spline approximation function $m_l^0(u_l) = \boldsymbol{B}(u_l)^T \boldsymbol{\lambda}_l^0$, where $\boldsymbol{\lambda}_l^0$ is the best spline approximation coefficients, such that

$$\sup_{u_l \in [a,b]} \left|m_l(u_l) - m_l^0(u_l)\right| = O(J_n^{-r}), \tag{A.1}$$

Let $\boldsymbol{\lambda}^0 = \left\{(\boldsymbol{\lambda}_l^0)^T : 1 \leq l \leq d\right\}^T$ and $k_n = J_n n^{-1/2} + J_n^{-r+1/2}$. We will show that for sufficient large n, for any given $\epsilon > 0$ there exists a large constant $C > 0$ such that

$$P\left\{\inf_{\|\boldsymbol{v}\|_2 = C} \mathcal{L}_{\tau n}\left(\boldsymbol{\lambda}^0 + k_n \boldsymbol{v}, \boldsymbol{\beta}^0\right) > \mathcal{L}_{\tau n}\left(\boldsymbol{\lambda}^0, \boldsymbol{\beta}^0\right)\right\} \geq 1 - \epsilon. \tag{A.2}$$

This implies that with probability at least $1 - \epsilon$ that there exists a local minimizer for (2.1) in the ball $\{\boldsymbol{\lambda}^0 + k_n \boldsymbol{v} : \|\boldsymbol{v}\|_2 \leq C\}$. Then it follows that there exists a local minimizer $\hat{\boldsymbol{\lambda}}\left(\boldsymbol{\beta}^0\right)$ such that $\left\|\hat{\boldsymbol{\lambda}}\left(\boldsymbol{\beta}^0\right) - \boldsymbol{\lambda}^0\right\|_2 = O_p\left(k_n\right)$. By applying the identity given in Knight (1998) that

$$\rho_\tau\left(r - s\right) - \rho_\tau\left(r\right) = s\left(I\left(r \leq 0\right) - \tau\right) + \int_0^s \left[I\left(r \leq t\right) - I\left(r \leq 0\right)\right] dt, \tag{A.3}$$

we have

$$\begin{aligned}
&\mathcal{L}_{\tau n}\left(\boldsymbol{\lambda}^0 + k_n \boldsymbol{v}, \boldsymbol{\beta}^0\right) - \mathcal{L}_{\tau n}\left(\boldsymbol{\lambda}^0, \boldsymbol{\beta}^0\right) \\
&= k_n \sum_{i=1}^n \left[I\left\{\varepsilon_i + \Delta_i \leq 0\right\} - \tau\right] \left\{\sum_{l=1}^d \boldsymbol{B}(\boldsymbol{Z}_i^T \boldsymbol{\beta}_l^0)^T \boldsymbol{v}_l X_{il}\right\} \\
&\quad + \sum_{i=1}^n \int_0^{k_n \sum_{l=1}^d \boldsymbol{B}(\boldsymbol{Z}_i^T \boldsymbol{\beta}_l^0)^T \boldsymbol{v}_l X_{il}} \left[I\left\{\varepsilon_i + \Delta_i \leq t\right\} - I\left\{\varepsilon_i + \Delta_i \leq 0\right\}\right] dt \\
&\triangleq I + II,
\end{aligned} \tag{A.4}$$

where $\Delta_i = \sum_{l=1}^{d} m_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0\right)X_{il} - \sum_{l=1}^{d}\boldsymbol{B}\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0\right)^T\boldsymbol{\lambda}_l^0 X_{il}$ and $\varepsilon_i = Y_i - \sum_{l=1}^{d} m_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0\right)X_{il}$. Meanwhile, the first term on the right hand side of (A.4) can be divided into the following two parts.

$$\begin{aligned} I &= k_n\sum_{i=1}^{n}\left[I\left\{\varepsilon_i \le 0\right\} - \tau\right]\left\{\sum_{l=1}^{d}\boldsymbol{B}(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0)^T\boldsymbol{v}_l X_{il}\right\} \\ &\quad - k_n\sum_{i=1}^{n}\left[I\left\{\varepsilon_i \le 0\right\} - I\left\{\varepsilon_i + \Delta_i \le 0\right\}\right]\left\{\sum_{l=1}^{d}\boldsymbol{B}(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0)^T\boldsymbol{v}_l X_{il}\right\} \\ &\triangleq I_1 + I_2. \end{aligned} \tag{A.5}$$

Thus $E\left(I_1\right) = 0$. By Lemma 1, for sufficiently large n, there are constants $0 < c_1 < C_1 < \infty$, such that

$$c_1\left\|\boldsymbol{v}\right\|_2^2 J_n^{-1} \le E\left\{\sum_{l=1}^{d}\boldsymbol{B}(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0)^T\boldsymbol{v}_l X_{il}\right\}^2 \le C_1\left\|\boldsymbol{v}\right\|_2^2 J_n^{-1}, \tag{A.6}$$

and by Bernstein's inequality in Bosq (1998), we have with probability 1,

$$c_1\left\|\boldsymbol{v}\right\|_2^2 J_n^{-1} \le n^{-1}\sum_{i=1}^{n}\left\{\sum_{l=1}^{d}\boldsymbol{B}(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0)^T\boldsymbol{v}_l X_{il}\right\}^2 \le C_1\left\|\boldsymbol{v}\right\|_2^2 J_n^{-1}. \tag{A.7}$$

Therefore, by (A.6), there is a constant $0 < C_2 < \infty$ such that for sufficiently large n,

$$Var\left(I_1\right) \le C_2 n k_n^2 E\left\{\sum_{l=1}^{d}\boldsymbol{B}(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0)^T\boldsymbol{v}_l X_{il}\right\}^2 \le n k_n^2 C_1 C_2\left\|\boldsymbol{v}\right\|_2^2 J_n^{-1}.$$

By weak law of large numbers, we have with probability approaching 1, for sufficiently large n, $|I_1| = O_p(n^{1/2}k_n J_n^{-1/2}\left\|\boldsymbol{v}\right\|_2)$. In addition,

$$\begin{aligned} I_2 &= k_n\sum_{i=1}^{n}\left[F_\varepsilon\left(-\Delta_i\left|\boldsymbol{X}_i, \boldsymbol{Z}_i\right.\right) - F_\varepsilon\left(0\left|\boldsymbol{X}_i, \boldsymbol{Z}_i\right.\right)\right]\left\{\sum_{l=1}^{d}\boldsymbol{B}(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0)^T\boldsymbol{v}_l X_{il}\right\} \\ &\quad - k_n\sum_{i=1}^{n}\left\{\left[I\left\{\varepsilon_i \le 0\right\} - I\left\{\varepsilon_i + \Delta_i \le 0\right\}\right]\right. \\ &\quad \left. - \left[F_\varepsilon\left(0\left|\boldsymbol{X}_i, \boldsymbol{Z}_i\right.\right) - F_\varepsilon\left(-\Delta_i\left|\boldsymbol{X}_i, \boldsymbol{Z}_i\right.\right)\right]\right\}\left\{\sum_{l=1}^{d}\boldsymbol{B}(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0)^T\boldsymbol{v}_l X_{il}\right\} \\ &\triangleq I_{21} - I_{22}. \end{aligned} \tag{A.8}$$

Moreover, by (A.1) and (A.6), taking Taylor's explanation for $F_\varepsilon\left(-\Delta_i\left|\boldsymbol{X}_i, \boldsymbol{Z}_i\right.\right)$ at 0 gives

$$\begin{aligned} I_{21} &= -k_n\sum_{i=1}^{n}\left[f_\varepsilon\left(0\left|\boldsymbol{X}_i, \boldsymbol{Z}_i\right.\right)\Delta_i\left(1 + o\left(1\right)\right)\right]\left\{\sum_{l=1}^{d}\boldsymbol{B}(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0)^T\boldsymbol{v}_l X_{il}\right\} \\ &= n k_n E\left\{\sum_{l=1}^{d}\boldsymbol{B}(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0)^T\boldsymbol{v}_l X_{il}\right\} O_p\left(J_n^{-r}\right) \\ &= O_p\left(n k_n J_n^{-r-1/2}\left\|\boldsymbol{v}\right\|_2\right). \end{aligned} \tag{A.9}$$

By direct calculation of the mean and variance, we can show that $I_{22} = O_p\left(n^{1/2}k_nJ_n^{-1/2}\left\|\boldsymbol{v}\right\|_2\right)$. This combined with (A.5), (A.8) and (A.9) lead to

$$\begin{aligned} I &= O_p\left(n^{1/2}k_nJ_n^{-1/2} + nk_nJ_n^{-r-1/2}\right)\left\|\boldsymbol{v}\right\|_2 \\ &= O_p\left(J_n^{1/2} + n^{1/2}J_n^{-r} + n^{1/2}J_n^{-r+1/2} + nJ_n^{-2r}\right)\left\|\boldsymbol{v}\right\|_2. \end{aligned} \tag{A.10}$$

Let $\boldsymbol{\mathcal{X}} = (\boldsymbol{X}_1, ..., \boldsymbol{X}_n)^T$, $\boldsymbol{\mathcal{Z}} = (\boldsymbol{Z}_1, ..., \boldsymbol{Z}_n)^T$ and $\Upsilon_i = k_n \sum_{l=1}^{d} \boldsymbol{B}(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0)^T\boldsymbol{v}_l X_{il}$. By (A.1), (A.4), (A.7), together with the condition (C5), we obtain

$$\begin{aligned} &E\left[II\,|\boldsymbol{\mathcal{X}},\boldsymbol{\mathcal{Z}}\right] \\ &= E\left[\sum_{i=1}^{n}\int_0^{\Upsilon_i}\left[I\left\{\varepsilon_i+\Delta_i\le t\right\} - I\left\{\varepsilon_i+\Delta_i\le 0\right\}\right]dt\,|\boldsymbol{\mathcal{X}},\boldsymbol{\mathcal{Z}}\right] \\ &= \sum_{i=1}^{n}\int_0^{\Upsilon_i}\left[F_\varepsilon\left((t-\Delta_i)\,|\boldsymbol{X}_i,\boldsymbol{Z}_i\right) - F_\varepsilon\left(-\Delta_i\,|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right]dt \\ &= \sum_{i=1}^{n}\int_0^{\Upsilon_i}\left[f_\varepsilon\left(-\Delta_i\,|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)t + o\left(t\right)\right]dt \\ &= \tfrac{1}{2}k_n^2\sum_{i=1}^{n} f_\varepsilon\left(0|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left\{\sum_{l=1}^{d}\boldsymbol{B}(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0)^T\boldsymbol{v}_lX_{il}\right\}^2\{1+o_p(1)\} \\ &= O_p(nk_n^2J_n^{-1})\left\|\boldsymbol{v}\right\|_2^2, \end{aligned} \tag{A.11}$$

and

$$\begin{aligned} &Var\left(II\,|\boldsymbol{\mathcal{X}},\boldsymbol{\mathcal{Z}}\right) \\ &= Var\left[\sum_{i=1}^{n}\int_0^{\Upsilon_i}\left[I\left\{\varepsilon_i+\Delta_i\le t\right\} - I\left\{\varepsilon_i+\Delta_i\le 0\right\}\right]dt\,|\boldsymbol{\mathcal{X}},\boldsymbol{\mathcal{Z}}\right] \\ &\le \sum_{i=1}^{n}E\left[\left(\int_0^{\Upsilon_i}\left[I\left\{\varepsilon_i+\Delta_i\le t\right\} - I\left\{\varepsilon_i+\Delta_i\le 0\right\}\right]dt\right)^2|\boldsymbol{\mathcal{X}},\boldsymbol{\mathcal{Z}}\right] \\ &\le \sum_{i=1}^{n}\int_0^{|\Upsilon_i|}\int_0^{|\Upsilon_i|}\left[F_\varepsilon\left\{(|\Upsilon_i|-\Delta_i)\,|\boldsymbol{X}_i,\boldsymbol{Z}_i\right\} - F_\varepsilon\left(-\Delta_i\,|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right]dt_1dt_2 \\ &\le o\left(\sum_{i=1}^{n}k_n^2\left(\sum_{l=1}^{d}\boldsymbol{B}(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0)^T\boldsymbol{v}_lX_{il}\right)^2\right) \\ &\le o_p\left(nk_n^2J_n^{-1}\right)\left\|\boldsymbol{v}\right\|_2^2. \end{aligned} \tag{A.12}$$

From (A.11) and (A.12), we have

$$II = O_p(\left\|\boldsymbol{v}\right\|_2^2 nk_n^2J_n^{-1}) = O_p(J_n + nJ_n^{-2r})\left\|\boldsymbol{v}\right\|_2^2. \tag{A.13}$$

It follows from (A.10) and (A.13) that, as long as $\left\|\boldsymbol{v}\right\|_2 = C$ is large enough, $\mathcal{L}_{\tau n}\left(\boldsymbol{\lambda}^0 + k_n\boldsymbol{v}, \boldsymbol{\beta}^0\right) - \mathcal{L}_{\tau n}\left(\boldsymbol{\lambda}^0, \boldsymbol{\beta}^0\right)$ in (A.4) is dominated II, which is large and positive for all sufficiently large n. Therefore, (A.2) holds and

$$\left\|\hat{\boldsymbol{\lambda}}(\boldsymbol{\beta}^0) - \boldsymbol{\lambda}^0\right\|_2 = O_p\left(k_n\right) = O_p\left(J_nn^{-1/2} + J_n^{-r+1/2}\right). \tag{A.14}$$

By Lemma 1, (A.1), (A.14) and triangle inequality, for $1 \leq l \leq d$, we have with probability approaching 1,

$$
\begin{aligned}
&\|\hat{m}_l(u_l, \boldsymbol{\beta}^0) - m_l(u_l)\|_2^2 \\
&= \int_a^b \left\{ \boldsymbol{B}^T(u)\hat{\boldsymbol{\lambda}}_l(\boldsymbol{\beta}^0) - \boldsymbol{B}^T(u)\boldsymbol{\lambda}_l^0 + \boldsymbol{B}^T(u)\boldsymbol{\lambda}_l^0 - m_l(u) \right\}^2 du \\
&\leq 2\int_a^b \left\{ \boldsymbol{B}^T(u)\hat{\boldsymbol{\lambda}}_l(\boldsymbol{\beta}^0) - \boldsymbol{B}^T(u)\boldsymbol{\lambda}_l^0 \right\}^2 du + 2\int_a^b \left\{ \boldsymbol{B}^T(u)\boldsymbol{\lambda}_l^0 - m_l(u) \right\}^2 du \\
&= 2\left(\hat{\boldsymbol{\lambda}}_l(\boldsymbol{\beta}^0) - \boldsymbol{\lambda}_l^0\right)^T \int_a^b \boldsymbol{B}(u)\boldsymbol{B}^T(u)\, du \left(\hat{\boldsymbol{\lambda}}_l(\boldsymbol{\beta}^0) - \boldsymbol{\lambda}_l^0\right) \\
&\quad + 2\int_a^b \left\{ \boldsymbol{B}^T(u)\boldsymbol{\lambda}_l^0 - m_l(u) \right\}^2 du \\
&\asymp J_n^{-1}\left\|\hat{\boldsymbol{\lambda}}_l(\boldsymbol{\beta}^0) - \boldsymbol{\lambda}_l^0\right\|_2^2 + O\left(J_n^{-2r}\right) \\
&= O_p\left(J_n n^{-1} + J_n^{-2r}\right).
\end{aligned}
$$

Thus $|\hat{m}_l(u_l, \boldsymbol{\beta}^0) - m_l(u_l)| = O_p\left(\sqrt{J_n/n} + J_n^{-r}\right)$ uniformly for any $u_l \in [a, b]$. According to (de Boor (2001), page 116), $\hat{\dot{m}}_l(u_l, \boldsymbol{\beta}^0) = \boldsymbol{B}_{q-1}(u_l)^T \boldsymbol{\Xi} \hat{\boldsymbol{\lambda}}_l(\boldsymbol{\beta}^0)$, where $\boldsymbol{B}_{q-1}(u_l) = \{B_{q-1,s}(u_l) : 2 \leq s \leq J_n\}^T$ is the $(q-1)$-th order B-spline basis and

$$
\boldsymbol{\Xi} = (q-1)\begin{bmatrix}
\frac{-1}{\xi_{q+1}-\xi_2} & \frac{1}{\xi_{q+1}-\xi_2} & 0 & \cdots & 0 \\
0 & \frac{-1}{\xi_{q+2}-\xi_3} & \frac{1}{\xi_{q+2}-\xi_3} & \cdots & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0 & 0 & \cdots & \frac{-1}{\xi_{N_n+2q-1}-\xi_{N_n+q}} & \frac{1}{\xi_{N_n+2q-1}-\xi_{N_n+q}}
\end{bmatrix}_{(J_n-1)\times J_n}.
$$

Follows from Ma and Song (2015), one has $\|\boldsymbol{\Xi}\|_\infty = O(J_n)$. Similar to the proof of $\hat{m}_l(u_l, \boldsymbol{\beta}^0)$, we have $\left|\hat{\dot{m}}_l(u_l, \boldsymbol{\beta}^0) - \dot{m}_l(u_l)\right| = O_p\left(\sqrt{J_n^3/n} + J_n^{-r+1}\right)$ uniformly for any $u_l \in [a, b]$. This completes the proof. □

For a given $\boldsymbol{\beta}$, let $\tilde{\boldsymbol{\lambda}}(\boldsymbol{\beta}) = \left(\tilde{\boldsymbol{\lambda}}_1^T(\boldsymbol{\beta}), ..., \tilde{\boldsymbol{\lambda}}_d^T(\boldsymbol{\beta})\right)^T$ be the minimizer of $E\{\mathcal{L}_{\tau n}(\boldsymbol{\lambda}, \boldsymbol{\beta}) | \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Z}}\}$, where $\mathcal{L}_{\tau n}(\boldsymbol{\lambda}, \boldsymbol{\beta})$ is the objective function defined in (2.1). Denote $m_i = \sum_{l=1}^d m_l\left(\boldsymbol{Z}_i^T \boldsymbol{\beta}_l^0\right) X_{il}$ for notation simplicity. We next present the following four lemmas to be used in the proof of Theorem 1.

Lemma 3. *Under the same conditions of Theorem 1, we have*

$$
\sup_{1\leq i\leq n}\left|\boldsymbol{D}_i^T(\boldsymbol{\beta}^0)\,\partial\tilde{\boldsymbol{\lambda}}(\boldsymbol{\beta}^0)\Big/\partial\boldsymbol{\beta}_{l,-1}^T - \left\{-\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0, \boldsymbol{\beta}^0\right) X_{il}\mathbb{P}\left(\boldsymbol{Z}_i^T\right)\boldsymbol{J}_l^0\right\}\right| = O_p\left(J_n^{-1}\sqrt{p_n}\right).
$$

Proof of Lemma 3. Let

$$\tilde{\mathcal{L}}_{\tau n}(\boldsymbol{\lambda}, \boldsymbol{\beta}) = n^{-1} \sum_{i=1}^{n} \rho_\tau \left\{ Y_i - \sum_{l=1}^{d} \boldsymbol{B}\big(\boldsymbol{Z}_i^T \boldsymbol{\beta}_l\big)^T \boldsymbol{\lambda}_l X_{il} \right\} - n^{-1} \sum_{i=1}^{n} \rho_\tau \{Y_i\}.$$

By applying the identity given in (A.3), we have

$$\begin{aligned} &\tilde{\mathcal{L}}_{\tau n}(\boldsymbol{\lambda}, \boldsymbol{\beta}) \\ &= n^{-1} \sum_{i=1}^{n} \left[\sum_{l=1}^{d} \boldsymbol{B}\big(\boldsymbol{Z}_i^T \boldsymbol{\beta}_l\big)^T \boldsymbol{\lambda}_l X_{il} \left\{ I(Y_i \le 0) - \tau \right\} \right] \\ &\quad + n^{-1} \sum_{i=1}^{n} \int_0^{\sum_{l=1}^{d} \boldsymbol{B}(\boldsymbol{Z}_i^T \boldsymbol{\beta}_l)^T \boldsymbol{\lambda}_l X_{il}} \left\{ I(Y_i \le t) - I(Y_i \le 0) \right\} dt \\ &= n^{-1} \sum_{i=1}^{n} \left[\int_0^{\sum_{l=1}^{d} \boldsymbol{B}(\boldsymbol{Z}_i^T \boldsymbol{\beta}_l)^T \boldsymbol{\lambda}_l X_{il}} I(Y_i \le t)\, dt - \tau \sum_{l=1}^{d} \boldsymbol{B}\big(\boldsymbol{Z}_i^T \boldsymbol{\beta}_l\big)^T \boldsymbol{\lambda}_l X_{il} \right]. \end{aligned}$$

Denote $F_Y(t \,|\boldsymbol{X}, \boldsymbol{Z}) = E\{I(Y \le t) \,|\boldsymbol{X}, \boldsymbol{Z}\}$. Then we have

$$\begin{aligned} &E\left\{ \tilde{\mathcal{L}}_{\tau n}(\boldsymbol{\lambda}, \boldsymbol{\beta}) \,|\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Z}} \right\} \\ &= n^{-1} \sum_{i=1}^{n} \left[\int_0^{\sum_{l=1}^{d} \boldsymbol{B}(\boldsymbol{Z}_i^T \boldsymbol{\beta}_l)^T \boldsymbol{\lambda}_l X_{il}} F_Y(t \,|\boldsymbol{X}_i, \boldsymbol{Z}_i)\, dt - \tau \sum_{l=1}^{d} \boldsymbol{B}\big(\boldsymbol{Z}_i^T \boldsymbol{\beta}_l\big)^T \boldsymbol{\lambda}_l X_{il} \right]. \end{aligned}$$

By the definition of $\tilde{\boldsymbol{\lambda}}(\boldsymbol{\beta})$, we have that $\tilde{\boldsymbol{\lambda}}(\boldsymbol{\beta})$ is the minimizer of $E\left\{ \tilde{\mathcal{L}}_{\tau n}(\boldsymbol{\lambda}, \boldsymbol{\beta}) \,|\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Z}} \right\}$ for a given $\boldsymbol{\beta}$. Hence

$$\begin{aligned} &\partial E\left\{ \tilde{\mathcal{L}}_{\tau n}\left(\tilde{\boldsymbol{\lambda}}(\boldsymbol{\beta}), \boldsymbol{\beta}\right) |\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Z}} \right\} / \partial \boldsymbol{\lambda} \\ &= \left[\partial E\left\{ \tilde{\mathcal{L}}_{\tau n}(\boldsymbol{\lambda}, \boldsymbol{\beta}) \,|\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Z}} \right\} / \partial \boldsymbol{\lambda} \right] |_{\boldsymbol{\lambda} = \tilde{\boldsymbol{\lambda}}(\boldsymbol{\beta})} \\ &= n^{-1} \sum_{i=1}^{n} \boldsymbol{D}_i(\boldsymbol{\beta}) \left\{ F_Y\left(\boldsymbol{D}_i^T(\boldsymbol{\beta})\, \tilde{\boldsymbol{\lambda}}(\boldsymbol{\beta}) \,|\boldsymbol{X}_i, \boldsymbol{Z}_i \right) - \tau \right\} = \mathbf{0}. \end{aligned}$$

Since $\partial E\left\{ \tilde{\mathcal{L}}_{\tau n}(\boldsymbol{\lambda}(\boldsymbol{\beta}), \boldsymbol{\beta}) \,|\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Z}} \right\} / \partial \boldsymbol{\lambda}$ is differentiable with respect to $\boldsymbol{\lambda}$ and $\boldsymbol{\beta}$ and

$$\begin{aligned} &\partial^2 E\left\{ \tilde{\mathcal{L}}_{\tau n}\left(\tilde{\boldsymbol{\lambda}}(\boldsymbol{\beta}), \boldsymbol{\beta}\right) |\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Z}} \right\} / \partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}^T \\ &= n^{-1} \sum_{i=1}^{n} \boldsymbol{D}_i(\boldsymbol{\beta})\, f_Y\left(\boldsymbol{D}_i^T(\boldsymbol{\beta})\, \tilde{\boldsymbol{\lambda}}(\boldsymbol{\beta}) \,|\boldsymbol{X}_i, \boldsymbol{Z}_i \right) \boldsymbol{D}_i^T(\boldsymbol{\beta}) \ne \mathbf{0}, \end{aligned}$$

by Implicit Function Theorem, we then can find $\partial \tilde{\boldsymbol{\lambda}}(\boldsymbol{\beta}^0) / \partial \boldsymbol{\beta}_{l,-1}^T$ through

$$\partial^2 E\left\{ \tilde{\mathcal{L}}_{\tau n}\left(\tilde{\boldsymbol{\lambda}}(\boldsymbol{\beta}^0), \boldsymbol{\beta}^0\right) |\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Z}} \right\} / \partial \boldsymbol{\lambda} \partial \boldsymbol{\beta}_{l,-1}^T = \mathbf{0}.$$

Define

$$\begin{pmatrix} \dot{\boldsymbol{B}}\left(\boldsymbol{Z}_i^T \boldsymbol{\beta}_1^0\right) X_{i1} \boldsymbol{Z}_i^T \boldsymbol{J}_1^0 & \cdots & \mathbf{0}_{J_n \times (p_n - 1)} \\ \vdots & \ddots & \vdots \\ \mathbf{0}_{J_n \times (p_n - 1)} & \cdots & \dot{\boldsymbol{B}}\left(\boldsymbol{Z}_i^T \boldsymbol{\beta}_d^0\right) X_{id} \boldsymbol{Z}_i^T \boldsymbol{J}_d^0 \end{pmatrix} \triangleq (\boldsymbol{\Lambda}_{i1}, ..., \boldsymbol{\Lambda}_{id})$$

where $\mathbf{0}_{s\times k}$ is the $s \times k$ dimensional matrix with 0's as its elements, we have

$$\begin{aligned}
&\partial^2 E\left\{\tilde{\mathcal{L}}_{\tau n}\left(\tilde{\boldsymbol{\lambda}}(\boldsymbol{\beta}^0),\boldsymbol{\beta}^0\right)|\boldsymbol{\mathcal{X}},\boldsymbol{\mathcal{Z}}\right\}/\partial\boldsymbol{\lambda}\partial\boldsymbol{\beta}_{l,-1}^T \\
&=\partial^2 E\left\{\tilde{\mathcal{L}}_{\tau n}\left(\tilde{\boldsymbol{\lambda}}(\boldsymbol{\beta}),\boldsymbol{\beta}\right)|\boldsymbol{\mathcal{X}},\boldsymbol{\mathcal{Z}}\right\}/\partial\boldsymbol{\lambda}\partial\boldsymbol{\beta}_{l,-1}^T|_{\boldsymbol{\beta}=\boldsymbol{\beta}^0} \\
&= n^{-1}\sum_{i=1}^n \boldsymbol{\Lambda}_{il}\left\{F_Y\left(\boldsymbol{D}_i^T\left(\boldsymbol{\beta}^0\right)\tilde{\boldsymbol{\lambda}}\left(\boldsymbol{\beta}^0\right)|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\tau\right\} \\
&\quad + n^{-1}\sum_{i=1}^n\left[\boldsymbol{D}_i\left(\boldsymbol{\beta}^0\right)f_Y\left(\boldsymbol{D}_i^T\left(\boldsymbol{\beta}^0\right)\tilde{\boldsymbol{\lambda}}\left(\boldsymbol{\beta}^0\right)|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right. \\
&\quad \left.\times\left(\dot{\boldsymbol{B}}^T\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0\right)\tilde{\boldsymbol{\lambda}}_l\left(\boldsymbol{\beta}^0\right)X_{il}\boldsymbol{Z}_i^T\boldsymbol{J}_l^0+\boldsymbol{D}_i^T\left(\boldsymbol{\beta}^0\right)\left(\partial\tilde{\boldsymbol{\lambda}}\left(\boldsymbol{\beta}^0\right)/\partial\boldsymbol{\beta}_{l,-1}^T\right)\right)\right] \\
&= \mathbf{0}.
\end{aligned} \tag{A.15}$$

Denote

$$\begin{aligned}
\boldsymbol{\Xi}_n &= n^{-1}\sum_{i=1}^n f_Y\left(\boldsymbol{D}_i^T\left(\boldsymbol{\beta}^0\right)\tilde{\boldsymbol{\lambda}}\left(\boldsymbol{\beta}^0\right)|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\boldsymbol{D}_i\left(\boldsymbol{\beta}^0\right)\boldsymbol{D}_i^T\left(\boldsymbol{\beta}^0\right), \\
\boldsymbol{\Delta}_{n1l} &= n^{-1}\sum_{i=1}^n\boldsymbol{\Lambda}_{il}\left\{F_Y\left(\boldsymbol{D}_i^T\left(\boldsymbol{\beta}^0\right)\tilde{\boldsymbol{\lambda}}\left(\boldsymbol{\beta}^0\right)|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\tau\right\}, \\
\boldsymbol{\Delta}_{n2l} &= n^{-1}\sum_{i=1}^n\boldsymbol{D}_i\left(\boldsymbol{\beta}^0\right)f_Y\left(\boldsymbol{D}_i^T\left(\boldsymbol{\beta}^0\right)\tilde{\boldsymbol{\lambda}}\left(\boldsymbol{\beta}^0\right)|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\dot{\boldsymbol{B}}^T\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0\right)\tilde{\boldsymbol{\lambda}}_l\left(\boldsymbol{\beta}^0\right)X_{il}\boldsymbol{Z}_i^T\boldsymbol{J}_l^0.
\end{aligned}$$

Then, by (A.15), we have

$$\partial\tilde{\boldsymbol{\lambda}}\left(\boldsymbol{\beta}^0\right)\Big/\partial\boldsymbol{\beta}_{l,-1}^T = -\boldsymbol{\Xi}_n^{-1}\left(\boldsymbol{\Delta}_{n1l}+\boldsymbol{\Delta}_{n2l}\right). \tag{A.16}$$

By Lemma 4 and (A.1), we have

$$\sup_{u_l\in[a,b]}\left|\boldsymbol{B}^T\left(u_l\right)\tilde{\boldsymbol{\lambda}}\left(\boldsymbol{\beta}^0\right)-m_l\left(u_l\right)\right| = O_p\left(J_n^{-r}\right), \tag{A.17}$$

$$\sup_{u_l\in[a,b]}\left|\dot{\boldsymbol{B}}^T\left(u_l\right)\tilde{\boldsymbol{\lambda}}\left(\boldsymbol{\beta}^0\right)-\dot{m}_l\left(u_l\right)\right| = O_p\left(J_n^{-r+1}\right). \tag{A.18}$$

Result (A.17) implies that

$$\sup_{1\le i\le n}\left|F_Y\left(\boldsymbol{D}_i^T\left(\boldsymbol{\beta}^0\right)\tilde{\boldsymbol{\lambda}}\left(\boldsymbol{\beta}^0\right)|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-\tau\right| = O_p\left(J_n^{-r}\right), \tag{A.19}$$

and

$$\sup_{1\le i\le n}\left|f_Y\left(\boldsymbol{D}_i^T\left(\boldsymbol{\beta}^0\right)\tilde{\boldsymbol{\lambda}}\left(\boldsymbol{\beta}^0\right)|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-f_Y\left(m_i\,|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right| = O_p\left(J_n^{-r}\right). \tag{A.20}$$

Moreover, it can be proved by Bernstein's inequality that

$$\sup_{s,l} n^{-1}\sum_{i=1}^n\left|\dot{B}_s\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0\right)X_{il}\right| = O_p\left(1\right), \tag{A.21}$$

and

$$\sup_{s,l} n^{-1} \sum_{i=1}^{n} \left| B_s \left(\boldsymbol{Z}_i^T \boldsymbol{\beta}_l^0 \right) X_{il} \right| = O_p \left(J_n^{-1} \right). \tag{A.22}$$

(A.19) and (A.21) imply that

$$\begin{aligned} |\boldsymbol{\Delta}_{n1l}| &\leq \sup_{s,l} n^{-1} \sum_{i=1}^{n} |\dot{B}_s \left(\boldsymbol{Z}_i^T \boldsymbol{\beta}_l^0 \right) X_{il}| \sup_{1 \leq i \leq n} \left\| \boldsymbol{Z}_i^T \boldsymbol{J}_l^0 \right\|_2 \\ &\quad \times \sup_{1 \leq i \leq n} \left| F_Y \left(\boldsymbol{D}_i^T (\boldsymbol{\beta}^0) \tilde{\boldsymbol{\lambda}} (\boldsymbol{\beta}^0) | \boldsymbol{X}_i, \boldsymbol{Z}_i \right) - \tau \right| \\ &= O_p \left(\sqrt{p_n} J_n^{-r} \right), \end{aligned} \tag{A.23}$$

and by (A.18), (A.20)–(A.22), we have

$$\begin{aligned} &\left| \boldsymbol{\Delta}_{n2l} - n^{-1} \sum_{i=1}^{n} \boldsymbol{D}_i (\boldsymbol{\beta}^0) f_Y (m_i | \boldsymbol{X}_i, \boldsymbol{Z}_i) \dot{m}_l \left(\boldsymbol{Z}_i^T \boldsymbol{\beta}_l^0, \boldsymbol{\beta}^0 \right) X_{il} \boldsymbol{Z}_i^T \boldsymbol{J}_l^0 \right| \\ &\leq \sup_{1 \leq i \leq n} \left| f_Y \left(\boldsymbol{D}_i^T (\boldsymbol{\beta}^0) \tilde{\boldsymbol{\lambda}} (\boldsymbol{\beta}^0) | \boldsymbol{X}_i, \boldsymbol{Z}_i \right) - f_Y (m_i | \boldsymbol{X}_i, \boldsymbol{Z}_i) \right| \\ &\quad \times \sup_{s,l} n^{-1} \sum_{i=1}^{n} \left| B_s \left(\boldsymbol{Z}_i^T \boldsymbol{\beta}_l^0 \right) X_{il} \right| \sup_{1 \leq i \leq n} \left\| \dot{\boldsymbol{B}}^T \left(\boldsymbol{Z}_i^T \boldsymbol{\beta}_l^0 \right) \tilde{\boldsymbol{\lambda}}_l (\boldsymbol{\beta}^0) X_{il} \boldsymbol{Z}_i^T \boldsymbol{J}_l^0 \right\|_2 \\ &\quad + \sup_{1 \leq i \leq n} f_Y (m_i | \boldsymbol{X}_i, \boldsymbol{Z}_i) \left| \dot{\boldsymbol{B}}^T \left(\boldsymbol{Z}_i^T \boldsymbol{\beta}_l^0 \right) \tilde{\boldsymbol{\lambda}}_l (\boldsymbol{\beta}^0) - \dot{m}_l \left(\boldsymbol{Z}_i^T \boldsymbol{\beta}_l^0, \boldsymbol{\beta}^0 \right) \right| \\ &\quad \times \sup_{1 \leq i \leq n} \left\| X_{il} \boldsymbol{Z}_i^T \boldsymbol{J}_l^0 \right\|_2 \sup_{s,l} n^{-1} \sum_{i=1}^{n} \left| B_s \left(\boldsymbol{Z}_i^T \boldsymbol{\beta}_l^0 \right) X_{il} \right| \\ &= O_p \left(J_n^{-r} \sqrt{p_n} \right). \end{aligned} \tag{A.24}$$

Hence

$$\begin{aligned} &\left| \boldsymbol{\Delta}_{n1l} + \boldsymbol{\Delta}_{n2l} - n^{-1} \sum_{i=1}^{n} \boldsymbol{D}_i (\boldsymbol{\beta}^0) f_Y (m_i | \boldsymbol{X}_i, \boldsymbol{Z}_i) \dot{m}_l \left(\boldsymbol{Z}_i^T \boldsymbol{\beta}_l^0, \boldsymbol{\beta}^0 \right) (\boldsymbol{\beta}^0) X_{il} \boldsymbol{Z}_i^T \boldsymbol{J}_l^0 \right| \\ &= O_p \left(J_n^{-r} \sqrt{p_n} \right). \end{aligned} \tag{A.25}$$

By Lemma 1, one has with probability approaching 1, for large enough n, $\forall \boldsymbol{\beta} \in \Theta$,

$$\begin{aligned} c_V J_n^{-1} \boldsymbol{\alpha}^T \boldsymbol{\alpha} &\leq \boldsymbol{\alpha}^T \hat{\boldsymbol{V}} (\boldsymbol{\beta}) \boldsymbol{\alpha} \leq C_V J_n^{-1} \boldsymbol{\alpha}^T \boldsymbol{\alpha}, \\ C_V^{-1} J_n \boldsymbol{\alpha}^T \boldsymbol{\alpha} &\leq \boldsymbol{\alpha}^T \hat{\boldsymbol{V}}^{-1} (\boldsymbol{\beta}) \boldsymbol{\alpha} \leq c_V^{-1} J_n \boldsymbol{\alpha}^T \boldsymbol{\alpha} \end{aligned}$$

for any vector $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1^T, ..., \boldsymbol{\alpha}_d^T)^T$ with $\boldsymbol{\alpha}_l = (\alpha_{sl}, 1 \leq s \leq J_n)^T$. By Lemma 1 (i) and Demko (1986), it can be proven that $\forall \boldsymbol{\beta} \in \Theta$ and for large enough n, there is a constant $0 < C_V^* < \infty$ such that $\|\boldsymbol{V}^{-1} (\boldsymbol{\beta})\|_\infty \leq C_V^* J_n$. Following this result, Lemma 1 (ii) and Lemma 1 (iii), it can be proved that $\forall \boldsymbol{\beta} \in \Theta$, $\left\| \hat{\boldsymbol{V}}^{-1} (\boldsymbol{\beta}) \right\|_\infty = O_p (J_n)$. By this result, (A.20) and condition (C5), we have

$$\left\| \boldsymbol{\Xi}_n^{-1} \right\|_\infty = O_p (J_n),$$

$$\left\|\left\{n^{-1}\sum_{i=1}^{n} f_Y\left(m_i\left|\boldsymbol{X}_i,\boldsymbol{Z}_i\right.\right)\boldsymbol{D}_i\left(\boldsymbol{\beta}^0\right)\boldsymbol{D}_i^T\left(\boldsymbol{\beta}^0\right)\right\}^{-1}\right\|_\infty = O_p\left(J_n\right).$$

The above results and (A.20), we have

$$\begin{aligned}
&\left\|\boldsymbol{\Xi}_n^{-1}-\left\{n^{-1}\sum_{i=1}^{n} f_Y\left(m_i\left|\boldsymbol{X}_i,\boldsymbol{Z}_i\right.\right)\boldsymbol{D}_i\left(\boldsymbol{\beta}^0\right)\boldsymbol{D}_i^T\left(\boldsymbol{\beta}^0\right)\right\}^{-1}\right\|_\infty\\
&\leq \left\|\boldsymbol{\Xi}_n^{-1}\right\|_\infty\left\|\left\{n^{-1}\sum_{i=1}^{n} f_Y\left(m_i\left|\boldsymbol{X}_i,\boldsymbol{Z}_i\right.\right)\boldsymbol{D}_i\left(\boldsymbol{\beta}^0\right)\boldsymbol{D}_i^T\left(\boldsymbol{\beta}^0\right)\right\}^{-1}\right\|_\infty\\
&\quad\times\left\|\boldsymbol{\Xi}_n-n^{-1}\sum_{i=1}^{n} f_Y\left(m_i\left|\boldsymbol{X}_i,\boldsymbol{Z}_i\right.\right)\boldsymbol{D}_i\left(\boldsymbol{\beta}^0\right)\boldsymbol{D}_i^T\left(\boldsymbol{\beta}^0\right)\right\|_\infty\\
&= O_p\left(J_n^2\right)O_p\left(J_n^{-r-1}\right)=O_p\left(J_n^{-r+1}\right).
\end{aligned}\tag{A.26}$$

By Bernstein's inequality, (A.22) and conditions (C4) and (C5), we have

$$\left|n^{-1}\sum_{i=1}^{n}\boldsymbol{D}_i\left(\boldsymbol{\beta}^0\right)f_Y\left(m_i\left|\boldsymbol{X}_i,\boldsymbol{Z}_i\right.\right)\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\boldsymbol{Z}_i^T\boldsymbol{J}_l^0\right| = O_p\left(J_n^{-1}\sqrt{p_n}\right).\tag{A.27}$$

Therefore, by (A.16), (A.25)– (A.27) and $f_Y\left(m_i\left|\boldsymbol{X}_i,\boldsymbol{Z}_i\right.\right)=f_\varepsilon\left(0\left|\boldsymbol{X}_i,\boldsymbol{Z}_i\right.\right)$, we have

$$\begin{aligned}
&\left|\partial\tilde{\boldsymbol{\lambda}}\left(\boldsymbol{\beta}^0\right)\Big/\partial\boldsymbol{\beta}_{l,-1}^T-\left[-\left\{n^{-1}\sum_{i=1}^{n} f_\varepsilon\left(0\left|\boldsymbol{X}_i,\boldsymbol{Z}_i\right.\right)\boldsymbol{D}_i\left(\boldsymbol{\beta}^0\right)\boldsymbol{D}_i^T\left(\boldsymbol{\beta}^0\right)\right\}^{-1}\right.\right.\\
&\quad\left.\left.\times n^{-1}\sum_{i=1}^{n}\boldsymbol{D}_i\left(\boldsymbol{\beta}^0\right)f_\varepsilon\left(0\left|\boldsymbol{X}_i,\boldsymbol{Z}_i\right.\right)\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\boldsymbol{Z}_i^T\boldsymbol{J}_l^0\right]\right|\\
&\leq \left\|\boldsymbol{\Xi}_n^{-1}\right\|_\infty\left|\boldsymbol{\Delta}_{n1l}+\boldsymbol{\Delta}_{n2l}-n^{-1}\sum_{i=1}^{n}\boldsymbol{D}_i\left(\boldsymbol{\beta}^0\right)f_\varepsilon\left(0\left|\boldsymbol{X}_i,\boldsymbol{Z}_i\right.\right)\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\boldsymbol{Z}_i^T\boldsymbol{J}_l^0\right|\\
&\quad+\left\|\boldsymbol{\Xi}_n^{-1}-\left\{n^{-1}\sum_{i=1}^{n} f_\varepsilon\left(0\left|\boldsymbol{X}_i,\boldsymbol{Z}_i\right.\right)\boldsymbol{D}_i\left(\boldsymbol{\beta}^0\right)\boldsymbol{D}_i^T\left(\boldsymbol{\beta}^0\right)\right\}^{-1}\right\|_\infty\\
&\quad\times\left|n^{-1}\sum_{i=1}^{n}\boldsymbol{D}_i\left(\boldsymbol{\beta}^0\right)f_\varepsilon\left(0\left|\boldsymbol{X}_i,\boldsymbol{Z}_i\right.\right)\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\boldsymbol{Z}_i^T\boldsymbol{J}_l^0\right|\\
&= O_p\left(J_n\right)O_p\left(J_n^{-r}\sqrt{p_n}\right)+O_p\left(J_n^{1-r}\right)O_p\left(J_n^{-1}\sqrt{p_n}\right)\\
&= O_p\left(J_n^{-r+1}\sqrt{p_n}\right).
\end{aligned}\tag{A.28}$$

Define $\hat{\boldsymbol{\delta}}_k=\underset{\boldsymbol{\delta}_k\in\mathbb{R}^{dJ_n}}{\arg\min}\sum_{i=1}^{n} f_\varepsilon(0\left|\boldsymbol{X}_i,\boldsymbol{Z}_i\right.)\left\{Z_{ik}-\boldsymbol{D}_i^T\left(\boldsymbol{\beta}^0\right)\boldsymbol{\delta}_k\right\}^2$, we have

$$\begin{aligned}
&\mathbb{P}_n\left(Z_{ik}\right)\\
&=\boldsymbol{D}_i\left(\boldsymbol{\beta}^0\right)^T\hat{\boldsymbol{\delta}}_k\\
&=\boldsymbol{D}_i(\boldsymbol{\beta}^0)^T\left(\sum_{i=1}^{n} f_\varepsilon(0\left|\boldsymbol{X}_i,\boldsymbol{Z}_i\right.)\boldsymbol{D}_i(\boldsymbol{\beta}^0)\boldsymbol{D}_i^T(\boldsymbol{\beta}^0)\right)^{-1}\sum_{i=1}^{n} f_\varepsilon(0\left|\boldsymbol{X}_i,\boldsymbol{Z}_i\right.)\boldsymbol{D}_i\left(\boldsymbol{\beta}^0\right)Z_{ik},
\end{aligned}\tag{A.29}$$

which is the B-spline estimator of $\mathbb{P}\left(Z_{ik}\right)$. By this result, we have

$$\begin{aligned}
&\boldsymbol{D}_i^T\left(\boldsymbol{\beta}^0\right)\left\{n^{-1}\sum_{i=1}^n f_{\varepsilon}\left(0\left|\boldsymbol{X}_i,\boldsymbol{Z}_i\right.\right)\boldsymbol{D}_i\left(\boldsymbol{\beta}^0\right)\boldsymbol{D}_i^T\left(\boldsymbol{\beta}^0\right)\right\}^{-1} \\
&\quad\times n^{-1}\sum_{i=1}^n f_{\varepsilon}\left(0\left|\boldsymbol{X}_i,\boldsymbol{Z}_i\right.\right)\boldsymbol{D}_i\left(\boldsymbol{\beta}^0\right)\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\boldsymbol{Z}_i^T\boldsymbol{J}_l^0 \\
&=\mathbb{P}_n\left(\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\boldsymbol{Z}_i^T\boldsymbol{J}_l^0\right) \\
&=\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\mathbb{P}_n\left(\boldsymbol{Z}_i^T\right)\boldsymbol{J}_l^0.
\end{aligned} \tag{A.30}$$

Moreover, by properties B-splines, condition (C2), together with the definition of $\mathbb{P}(\cdot)$ given in subsection 2.2, it can be shown that

$$\begin{aligned}
&\left\|\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\mathbb{P}_n\left(\boldsymbol{Z}_i^T\right)\boldsymbol{J}_l^0-\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\mathbb{P}\left(\boldsymbol{Z}_i^T\right)\boldsymbol{J}_l^0\right\|_{\infty} \\
&=O_p\left(J_n^{1/2}\sqrt{p_n/n}+\sqrt{p_n}J_n^{-1}\right).
\end{aligned} \tag{A.31}$$

Hence, by (A.28), (A.30)– (A.31) and condition $n^{1/(2r+2)} \ll J_n \ll n^{1/4}$, we have

$$\sup_{1\le i\le n}\left|\boldsymbol{D}_i^T\left(\boldsymbol{\beta}^0\right)\tilde{\boldsymbol{\lambda}}\left(\boldsymbol{\beta}^0\right)/\boldsymbol{\beta}_{l,-1}^T-\left\{-\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\mathbb{P}\left(\boldsymbol{Z}_i^T\right)\boldsymbol{J}_l^0\right\}\right|=O_p\left(J_n^{-1}\sqrt{p_n}\right)$$

This completes the Lemma 3. □

Lemma 4. *Under the same conditions of Theorem 1, we have*

$$\left\|\tilde{\boldsymbol{\lambda}}\left(\boldsymbol{\beta}^0\right)-\boldsymbol{\lambda}^0\right\|_2=O_p\left(J_n^{-r+1/2}\right).$$

Proof of Lemma 4. Since $\tilde{\boldsymbol{\lambda}}\left(\boldsymbol{\beta}^0\right)$ minimizes $E\left\{\mathcal{L}_{\tau n}\left(\boldsymbol{\lambda},\boldsymbol{\beta}^0\right)|\boldsymbol{\mathcal{X}},\boldsymbol{\mathcal{Z}}\right\}$, then we have

$$E\left\{\mathcal{L}_{\tau n}\left(\tilde{\boldsymbol{\lambda}}\left(\boldsymbol{\beta}^0\right),\boldsymbol{\beta}^0\right)|\boldsymbol{\mathcal{X}},\boldsymbol{\mathcal{Z}}\right\}\le E\left\{\mathcal{L}_{\tau n}\left(\boldsymbol{\lambda},\boldsymbol{\beta}^0\right)|\boldsymbol{\mathcal{X}},\boldsymbol{\mathcal{Z}}\right\}$$

for any $\boldsymbol{\lambda}\in\mathbb{R}^{dJ_n}$. Hence

$$\begin{aligned}
&n^{-1}\sum_{i=1}^n E\left[\rho_\tau\left\{Y_i-\sum_{l=1}^d \boldsymbol{B}\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0\right)^T\tilde{\boldsymbol{\lambda}}_l\left(\boldsymbol{\beta}^0\right)X_{il}\right\}\right] \\
&\le n^{-1}\sum_{i=1}^n E\left[\rho_\tau\left\{Y_i-\sum_{l=1}^d \boldsymbol{B}\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0\right)^T\boldsymbol{\lambda}_l^0X_{il}\right\}\right]
\end{aligned}$$

Moreover, by the definition of $Q_\tau\left(Y_i|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)=\sum_{l=1}^d m_l(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0)X_{il}$, we have

$$\begin{aligned}
&n^{-1}\sum_{i=1}^n E\left[\rho_\tau\left\{Y_i-\sum_{l=1}^d m_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0\right)X_{il}\right\}\right] \\
&\le n^{-1}\sum_{i=1}^n E\left[\rho_\tau\left\{Y_i-\sum_{l=1}^d \boldsymbol{B}\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0\right)^T\tilde{\boldsymbol{\lambda}}_l\left(\boldsymbol{\beta}^0\right)X_{il}\right\}\right] \\
&\le n^{-1}\sum_{i=1}^n E\left[\rho_\tau\left\{Y_i-\sum_{l=1}^d \boldsymbol{B}\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0\right)^T\boldsymbol{\lambda}_l^0X_{il}\right\}\right]
\end{aligned}$$

By the above result and

$$n^{-1}\sum_{i=1}^{n} E\left[\sum_{l=1}^{d}\left(m_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0\right)X_{il}-\boldsymbol{B}\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0\right)^T\boldsymbol{\lambda}_l^0X_{il}\right)\right]^2\leq CJ_n^{-2r},$$

we have

$$n^{-1}\sum_{i=1}^{n} E\left[\sum_{l=1}^{d}\left(m_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0\right)X_{il}-\boldsymbol{B}\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0\right)^T\tilde{\boldsymbol{\lambda}}_l\left(\boldsymbol{\beta}^0\right)X_{il}\right)\right]^2\leq CJ_n^{-2r}$$

for some constant $0<C<\infty$. Thus

$$\begin{aligned}
&n^{-1}\sum_{i=1}^{n}\left[\sum_{l=1}^{d}\left(\boldsymbol{B}\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0\right)^T\tilde{\boldsymbol{\lambda}}_l\left(\boldsymbol{\beta}^0\right)X_{il}-\boldsymbol{B}\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0\right)^T\boldsymbol{\lambda}_l^0X_{il}\right)\right]^2\\
&=n^{-1}\sum_{i=1}^{n}\left[\boldsymbol{D}_i^T\left(\boldsymbol{\beta}^0\right)\tilde{\boldsymbol{\lambda}}\left(\boldsymbol{\beta}^0\right)-\boldsymbol{D}_i^T\left(\boldsymbol{\beta}^0\right)\boldsymbol{\lambda}^0\right]^2\\
&=O_p\left(J_n^{-2r}\right).
\end{aligned}$$

Hence by the above result and with probability 1 that

$$c_1J_n^{-1}\leq\rho_{\min}\left(n^{-1}\boldsymbol{D}^T\left(\boldsymbol{\beta}^0\right)\boldsymbol{D}\left(\boldsymbol{\beta}^0\right)\right)\leq\rho_{\max}\left(n^{-1}\boldsymbol{D}^T\left(\boldsymbol{\beta}^0\right)\boldsymbol{D}\left(\boldsymbol{\beta}^0\right)\right)\leq C_1J_n^{-1}$$

we have

$$\left\|\tilde{\boldsymbol{\lambda}}\left(\boldsymbol{\beta}^0\right)-\boldsymbol{\lambda}^0\right\|_2=O_p\left(J_n^{-r+1/2}\right).$$

Lemma 5. *Let*

$$\boldsymbol{\Pi}_i\left(\boldsymbol{\beta}^0\right)=-\psi_{\tau h}\left\{Y_i-m_i\right\}\left[\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\tilde{\boldsymbol{Z}}_i\right]_{l=1}^{d},$$

$m_i=\sum_{l=1}^{d}m_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0\right)X_{il}$ *and* $f^{(i)}(v)=d^if(v)/dv^i$*. Under the same conditions of Theorem 1, we have*

$$\begin{aligned}
&(i)\ E\boldsymbol{\Pi}_i\left(\boldsymbol{\beta}^0\right)=(-h)^{\nu}(\nu!)^{-1}C_kE\left\{\left[\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\tilde{\boldsymbol{Z}}_i\right]_{l=1}^{d}f_{\varepsilon}^{(\nu-1)}\left(0\,|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right\}\\
&\quad+o\left(h^{\nu}\sqrt{p_n}\right);\\
&(ii)\ E\frac{\partial\boldsymbol{\Pi}_i\left(\boldsymbol{\beta}^0\right)}{\partial\boldsymbol{\beta}_{-1}^T}=E\left\{f_{\varepsilon}\left(0\,|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(\left[\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\tilde{\boldsymbol{Z}}_i\right]_{l=1}^{d}\right)^{\otimes2}\right\}+o(1).
\end{aligned}$$

Proof of Lemma 5. By a change of variables, we have

$$
\begin{aligned}
&E\boldsymbol{\Pi}_i\left(\boldsymbol{\beta}^0\right)\\
&=E\left\{E\left[\left(G\left(\tfrac{m_i-Y_i}{h}\right)-\tau\right)\left[\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\tilde{\boldsymbol{Z}}_i\right]_{l=1}^d|\boldsymbol{\mathcal{X}},\boldsymbol{\mathcal{Z}}\right]\right\}\\
&=E\left\{\left[\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\tilde{\boldsymbol{Z}}_i\right]_{l=1}^d\left(\int_{-\infty}^{\infty}G\left(\tfrac{-\varepsilon}{h}\right)f_\varepsilon\left(\varepsilon\,|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)d\varepsilon-\tau\right)\right\}\\
&=E\left\{\left[\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\tilde{\boldsymbol{Z}}_i\right]_{l=1}^d\left(\int_{\infty}^{-\infty}G\left(v\right)f_\varepsilon\left(-hv\,|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)d(-hv)-\tau\right)\right\}\\
&=E\left\{\left[\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\tilde{\boldsymbol{Z}}_i\right]_{l=1}^d\left[\int_{\infty}^{-\infty}\left(\int_{x<v}K\left(x\right)dx\right)f_\varepsilon\left(-hv\,|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)d\left(-hv\right)-\tau\right]\right\}\\
&=E\left\{\left[\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\tilde{\boldsymbol{Z}}_i\right]_{l=1}^d\left[-\int_{-1}^{1}\int_{x}^{\infty}f_\varepsilon\left(-hv\,|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)d\left(-hv\right)K\left(x\right)dx-\tau\right]\right\}\\
&=E\left\{\left[\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\tilde{\boldsymbol{Z}}_i\right]_{l=1}^d\right.\\
&\quad\left.\times\left[\int_{-1}^{1}F_\varepsilon\left(-hu\,|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)K\left(u\right)du-\int_{-1}^{1}F_\varepsilon\left(0\,|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)K\left(u\right)du\right]\right\}\\
&=E\left\{\left[\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\tilde{\boldsymbol{Z}}_i\right]_{l=1}^d\int_{-1}^{1}\left[F_\varepsilon\left(-hu\,|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-F_\varepsilon\left(0\,|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right]K\left(u\right)du\right\}
\end{aligned}
$$

Then, by a Taylor expansion, we have

$$
\begin{aligned}
&F_\varepsilon\left(-hu\,|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-F_\varepsilon\left(0\,|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\\
&=f_\varepsilon\left(0\,|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(-hu\right)+\tfrac{1}{2!}f_\varepsilon^{(1)}\left(0\,|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(-hu\right)^2+\cdots\\
&\quad+\tfrac{1}{\nu!}f_\varepsilon^{(\nu-1)}\left(0\,|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left(-hu\right)^\nu\left\{1+o\left(1\right)\right\}.
\end{aligned}
$$

These results with condition (C6), we can obtain

$$
\begin{aligned}
&E\boldsymbol{\Pi}_i\left(\boldsymbol{\beta}^0\right)\\
&=(-h)^\nu(\nu!)^{-1}C_kE\left\{\left[\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)\boldsymbol{X}_{il}\boldsymbol{J}_l^{0T}\tilde{\boldsymbol{Z}}_i\right]_{l=1}^d f_\varepsilon^{(\nu-1)}\left(0\,|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right\}\\
&\quad+o\left(h^\nu\sqrt{p_n}\right).
\end{aligned}
$$

Similarly, part (ii) holds by noting that

$$
\frac{\partial\boldsymbol{\Pi}_i\left(\boldsymbol{\beta}^0\right)}{\partial\boldsymbol{\beta}_{-1}^T}=\frac{1}{h}K\left(\frac{m_i-Y_i}{h}\right)\left(\left[\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\tilde{\boldsymbol{Z}}_i\right]_{l=1}^d\right)^{\otimes 2}
$$

and

$$
\begin{aligned}
&E\frac{\partial\boldsymbol{\Pi}_i\left(\boldsymbol{\beta}^0\right)}{\partial\boldsymbol{\beta}_{-1}^T}\\
&=E\left\{E\left[\tfrac{1}{h}K\left(\tfrac{m_i-Y_i}{h}\right)\left(\left[\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\tilde{\boldsymbol{Z}}_i\right]_{l=1}^d\right)^{\otimes 2}|\boldsymbol{\mathcal{X}},\boldsymbol{\mathcal{Z}}\right]\right\}\\
&=E\left\{\left(\left[\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\tilde{\boldsymbol{Z}}_i\right]_{l=1}^d\right)^{\otimes 2}f_\varepsilon\left(0\,|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right\}\\
&\quad+E\left\{\left(\left[\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\tilde{\boldsymbol{Z}}_i\right]_{l=1}^d\right)^{\otimes 2}\int_{-1}^{1}\left[f_\varepsilon\left(-hu\,|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)-f_\varepsilon\left(0\,|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\right]K\left(u\right)du\right\}.
\end{aligned}
$$

Lemma 6. *Under the same conditions of Theorem 1, we have*

$$\mathcal{R}_{\tau nh}\left(\boldsymbol{\beta}_{-1}^{0}\right)=-\sum_{i=1}^{n}\psi_{\tau}\left\{\varepsilon_i\right\}\left[\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\tilde{\boldsymbol{Z}}_i\right]_{l=1}^{d}+o_p\left(\sqrt{np_n}\right).$$

Proof of Lemma 6. Based on Lemma 2, we can obtain

$$\hat{\dot{m}}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\boldsymbol{Z}_i=\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\boldsymbol{Z}_i+O_p\left(J_n^{3/2}\sqrt{p_n/n}+\sqrt{p_n}J_n^{-r+1}\right). \tag{A.32}$$

By Lemmas 2 and 4, we have

$$\left\|\tilde{\boldsymbol{\lambda}}\left(\boldsymbol{\beta}^0\right)-\boldsymbol{\lambda}^0\right\|_2=O_p\left(J_n^{-r+1/2}\right),\ \left\|\hat{\boldsymbol{\lambda}}\left(\boldsymbol{\beta}^0\right)-\boldsymbol{\lambda}^0\right\|_2=O_p\left(J_n^{-r+1/2}+J_nn^{-1/2}\right). \tag{A.33}$$

By Lemma 3, (A.32), (A.33) and the following decomposition

$$\hat{\boldsymbol{\lambda}}\left(\boldsymbol{\beta}^0\right)=\hat{\boldsymbol{\lambda}}\left(\boldsymbol{\beta}^0\right)-\tilde{\boldsymbol{\lambda}}\left(\boldsymbol{\beta}^0\right)+\tilde{\boldsymbol{\lambda}}\left(\boldsymbol{\beta}^0\right),$$

we have

$$\begin{aligned}&\hat{\dot{m}}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\boldsymbol{Z}_i+\left(\partial\hat{\boldsymbol{\lambda}}(\boldsymbol{\beta}^0)^T\Big/\partial\boldsymbol{\beta}_{l,-1}\right)\boldsymbol{D}_i\left(\boldsymbol{\beta}^0\right)\\&=\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\boldsymbol{Z}_i-\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\mathbb{P}\left(\boldsymbol{Z}_i\right)\\&\quad+O_p\left(J_n^{-1}\sqrt{p_n}+\sqrt{p_nJ_n^3}n^{-1/2}\right)\times\mathbf{1}_{(p_n-1)}\\&=\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\tilde{\boldsymbol{Z}}_i+O_p\left(J_n^{-1}\sqrt{p_n}+\sqrt{p_nJ_n^3}n^{-1/2}\right)\times\mathbf{1}_{(p_n-1)},\end{aligned}$$

where $\mathbf{1}_s$ is an $s-$vector of 1's. Hence, by (2.3) and the above result, one has

$$\begin{aligned}&\mathcal{R}_{\tau nh}\left(\boldsymbol{\beta}_{-1}^0\right)\\&=-\sum_{i=1}^{n}\psi_{\tau h}\left\{Y_i-\boldsymbol{D}_i^T\left(\boldsymbol{\beta}^0\right)\hat{\boldsymbol{\lambda}}\left(\boldsymbol{\beta}^0\right)\right\}\\&\quad\times\left[\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\tilde{\boldsymbol{Z}}_i+O_p\left(J_n^{-1}\sqrt{p_n}+\sqrt{p_nJ_n^3}n^{-1/2}\right)\times\mathbf{1}_{(p_n-1)}\right]_{l=1}^{d}\\&=-\sum_{i=1}^{n}\psi_{\tau h}\left\{Y_i-m_i+m_i-\boldsymbol{D}_i^T\left(\boldsymbol{\beta}^0\right)\hat{\boldsymbol{\lambda}}\left(\boldsymbol{\beta}^0\right)\right\}\\&\quad\times\left[\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\tilde{\boldsymbol{Z}}_i+O_p\left(J_n^{-1}\sqrt{p_n}+\sqrt{p_nJ_n^3}n^{-1/2}\right)\times\mathbf{1}_{(p_n-1)}\right]_{l=1}^{d}\\&\triangleq-\sum_{i=1}^{n}\psi_{\tau h}\left\{Y_i-m_i\right\}\left[\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\tilde{\boldsymbol{Z}}_i\right]_{l=1}^{d}-\boldsymbol{I}_1-\boldsymbol{I}_2-\boldsymbol{I}_3,\end{aligned} \tag{A.34}$$

where

$$\begin{aligned}\boldsymbol{I}_1&=\sum_{i=1}^{n}\frac{1}{h}K\left(\frac{Y_i-m_i^*}{h}\right)\left(m_i-\boldsymbol{D}_i^T\left(\boldsymbol{\beta}^0\right)\hat{\boldsymbol{\lambda}}\left(\boldsymbol{\beta}^0\right)\right)\left[\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\tilde{\boldsymbol{Z}}_i\right]_{l=1}^{d},\\\boldsymbol{I}_2&=\sum_{i=1}^{n}\psi_{\tau h}\left\{Y_i-m_i\right\}O_p\left(J_n^{-1}\sqrt{p_n}+\sqrt{p_nJ_n^3}n^{-1/2}\right)\times\mathbf{1}_{d(p_n-1)},\\\boldsymbol{I}_3&=\sum_{i=1}^{n}\frac{1}{h}K\left(\frac{Y_i-m_i^*}{h}\right)\left(m_i-\boldsymbol{D}_i^T\left(\boldsymbol{\beta}^0\right)\hat{\boldsymbol{\lambda}}\left(\boldsymbol{\beta}^0\right)\right)O_p\left(J_n^{-1}\sqrt{p_n}+\sqrt{p_nJ_n^3}n^{-1/2}\right)\times\mathbf{1}_{d(p_n-1)},\end{aligned}$$

where m_i^* lies between m_i and $\boldsymbol{D}_i^T(\boldsymbol{\beta}^0)\hat{\boldsymbol{\lambda}}(\boldsymbol{\beta}^0)$. We will prove that $\|\boldsymbol{I}_j\|_2 = o_p(\sqrt{np_n})$ for each $j = 1, 2, 3$. By Lemma 2 and (A.1), we have

$$\begin{aligned}
&\left|m_i - \boldsymbol{D}_i^T(\boldsymbol{\beta}^0)\hat{\boldsymbol{\lambda}}(\boldsymbol{\beta}^0)\right| \\
&\leq \left|\sum_{l=1}^d m_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0\right)X_{il} - \boldsymbol{D}_i^T(\boldsymbol{\beta}^0)\boldsymbol{\lambda}^0\right| + \left|\boldsymbol{D}_i^T(\boldsymbol{\beta}^0)\left(\hat{\boldsymbol{\lambda}}(\boldsymbol{\beta}^0) - \boldsymbol{\lambda}^0\right)\right| \\
&= O_p\left(J_n^{-r} + J_n^{1/2}n^{-1/2}\right).
\end{aligned}$$

This combines with the weak law of large numbers and condition $n^{1/(2r+2)} \ll J_n \ll n^{1/4}$, we have $\|\boldsymbol{I}_1\|_2 = o_p(\sqrt{np_n})$. Moreover,

$$\begin{aligned}
&\|\boldsymbol{I}_2\|_2 = O_p(\sqrt{n})\,O_p\left(J_n^{-1}\sqrt{p_n} + \sqrt{p_nJ_n^3}n^{-1/2}\right) = o_p(\sqrt{np_n});\\
&\|\boldsymbol{I}_3\|_2 = O_p\left(nJ_n^{-r} + J_n^{1/2}n^{1/2}\right)O_p\left(J_n^{-1}\sqrt{p_n} + \sqrt{p_nJ_n^3}n^{-1/2}\right) = o_p(\sqrt{np_n}).
\end{aligned}$$

Therefore, by (A.34) and the above results, we have

$$\mathcal{R}_{\tau nh}\left(\boldsymbol{\beta}_{-1}^0\right) = -\sum_{i=1}^n \psi_{\tau h}\{Y_i - m_i\}\left[\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0, \boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\tilde{\boldsymbol{Z}}_i\right]_{l=1}^d + o_p(\sqrt{np_n}). \tag{A.35}$$

Writing $G_{ni} \triangleq G(-\varepsilon_i/h) - I(\varepsilon_i \leq 0)$ and rearranging terms, we have

$$\begin{aligned}
&-\sum_{i=1}^n \psi_{\tau h}\{Y_i - m_i\}\left[\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0, \boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\tilde{\boldsymbol{Z}}_i\right]_{l=1}^d \\
&= \sum_{i=1}^n (G(-\varepsilon_i/h) - \tau)\left[\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0, \boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\tilde{\boldsymbol{Z}}_i\right]_{l=1}^d \\
&= \sum_{i=1}^n (I(\varepsilon_i \leq 0) - \tau)\left[\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0, \boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\tilde{\boldsymbol{Z}}_i\right]_{l=1}^d \\
&\quad + \left\{\sum_{i=1}^n G_{ni}\left[\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0, \boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\tilde{\boldsymbol{Z}}_i\right]_{l=1}^d - nE\left(G_{ni}\left[\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0, \boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\tilde{\boldsymbol{Z}}_i\right]_{l=1}^d\right)\right\} \\
&\quad + nE\left(G_{ni}\left[\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0, \boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\tilde{\boldsymbol{Z}}_i\right]_{l=1}^d\right).
\end{aligned} \tag{A.36}$$

Since, for each $\epsilon > 0$

$$\begin{aligned}
&P\left\{\left\|\sum_{i=1}^n G_{ni}\left[\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0, \boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\tilde{\boldsymbol{Z}}_i\right]_{l=1}^d \right.\right.\\
&\qquad \left.\left. - \ nE\left(G_{ni}\left[\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0, \boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\tilde{\boldsymbol{Z}}_i\right]_{l=1}^d\right)\right\|_2 > \epsilon\right\} \\
&\leq \epsilon^{-2}nE\left\{[G(-\varepsilon_i/h) - I(\varepsilon_i \leq 0)]^2\left\|\left[\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0, \boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\tilde{\boldsymbol{Z}}_i\right]_{l=1}^d\right\|_2^2\right\} \\
&\leq Cntr\left\{E\left(\left[\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0, \boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\tilde{\boldsymbol{Z}}_i\right]_{l=1}^d\right)^{\otimes 2}\right\}P\{-h \leq \varepsilon \leq h\} \\
&= O(nhp_n),
\end{aligned}$$

where the last inequality uses the fact that $G\left(-\varepsilon_i/h\right) - I\left(\varepsilon_i \leq 0\right)$ differs from zero only if $|\varepsilon_i| \leq h$ by condition (C6). Therefore,

$$\begin{aligned}
&\left\| \sum_{i=1}^{n} G_{ni} \left[\dot{m}_l \left(\boldsymbol{Z}_i^T \boldsymbol{\beta}_l^0, \boldsymbol{\beta}^0 \right) X_{il} \boldsymbol{J}_l^{0T} \tilde{\boldsymbol{Z}}_i \right]_{l=1}^{d} - nE \left(G_{ni} \left[\dot{m}_l \left(\boldsymbol{Z}_i^T \boldsymbol{\beta}_l^0, \boldsymbol{\beta}^0 \right) X_{il} \boldsymbol{J}_l^{0T} \tilde{\boldsymbol{Z}}_i \right]_{l=1}^{d} \right) \right\|_2 \\
&= O_p\left(\sqrt{nhp_n}\right) \\
&= o_p\left(\sqrt{np_n}\right).
\end{aligned} \tag{A.37}$$

Also, the last term on the right hand side of (A.36) is also $o_p\left(\sqrt{np_n}\right)$ using condition (C7) $nh^{2\nu} \to 0$ and Lemma 5 since

$$nE\left(G_{ni} \left[\dot{m}_l \left(\boldsymbol{Z}_i^T \boldsymbol{\beta}_l^0, \boldsymbol{\beta}^0 \right) X_{il} \boldsymbol{J}_l^{0T} \tilde{\boldsymbol{Z}}_i \right]_{l=1}^{d} \right) = nE\left(\boldsymbol{\Pi}_i\left(\boldsymbol{\beta}^0\right)\right) = O\left(nh^{\nu}\sqrt{p_n}\right) = o\left(\sqrt{np_n}\right).$$

This result coupled with (A.35)–(A.37) leads to

$$\mathcal{R}_{\tau nh}\left(\boldsymbol{\beta}_{-1}^0\right) = -\sum_{i=1}^{n} \psi_\tau\left\{\varepsilon_i\right\} \left[\dot{m}_l \left(\boldsymbol{Z}_i^T \boldsymbol{\beta}_l^0, \boldsymbol{\beta}^0 \right) X_{il} \boldsymbol{J}_l^{0T} \tilde{\boldsymbol{Z}}_i \right]_{l=1}^{d} + o_p\left(\sqrt{np_n}\right).$$

Proof of Theorem 1 To prove (i), it suffices to show that $\forall \delta > 0$, there exists a sufficiently large constant $\triangle > 0$ such that for sufficiently large n,

$$P\left\{ \inf_{\left\|\boldsymbol{\beta}_{-1} - \boldsymbol{\beta}_{-1}^0\right\|_2 = \Delta\sqrt{p_n/n}} \left(\boldsymbol{\beta}_{-1} - \boldsymbol{\beta}_{-1}^0\right)^T \mathcal{R}_{\tau nh}\left(\boldsymbol{\beta}_{-1}\right) > 0 \right\} \geq 1 - \delta. \tag{A.38}$$

(A.38) is sufficient to ensure the existence of the root $\hat{\boldsymbol{\beta}}_{-1}$ of the equation $\mathcal{R}_{\tau nh}\left(\boldsymbol{\beta}_{-1}\right) = \mathbf{0}$ such that $\left\|\boldsymbol{\beta}_{-1} - \boldsymbol{\beta}_{-1}^0\right\|_2 = O_p(\sqrt{p_n/n})$. This implies that we need to calculate the sign of $\left(\boldsymbol{\beta}_{-1} - \boldsymbol{\beta}_{-1}^0\right)^T \mathcal{R}_{\tau nh}\left(\boldsymbol{\beta}_{-1}\right)$ on $\left\{\boldsymbol{\beta}_{-1} : \left\|\boldsymbol{\beta}_{-1} - \boldsymbol{\beta}_{-1}^0\right\|_2 = \Delta\sqrt{p_n/n}\right\}$. Note that

$$\begin{aligned}
&\left(\boldsymbol{\beta}_{-1} - \boldsymbol{\beta}_{-1}^0\right)^T \mathcal{R}_{\tau nh}\left(\boldsymbol{\beta}_{-1}\right) \\
&= \left(\boldsymbol{\beta}_{-1} - \boldsymbol{\beta}_{-1}^0\right)^T \mathcal{R}_{\tau nh}\left(\boldsymbol{\beta}_{-1}^0\right) + \left(\boldsymbol{\beta}_{-1} - \boldsymbol{\beta}_{-1}^0\right)^T \frac{\partial}{\partial \boldsymbol{\beta}_{-1}^T} \mathcal{R}_{\tau nh}\left(\boldsymbol{\beta}_{-1}^0\right) \left(\boldsymbol{\beta}_{-1} - \boldsymbol{\beta}_{-1}^0\right) \{1 + o_p(1)\} \\
&= I_{n1} + I_{n2}.
\end{aligned}$$

We first consider I_{n1}. By Lemma 6, we have $\mathcal{R}_{\tau nh}\left(\boldsymbol{\beta}_{-1}^0\right) = \boldsymbol{S}_n\left(\boldsymbol{\beta}_{-1}^0\right) + o_p\left(\sqrt{np_n}\right)$, where

$$\boldsymbol{S}_n\left(\boldsymbol{\beta}_{-1}^0\right) = -\sum_{i=1}^{n} \psi_\tau\left\{\varepsilon_i\right\} \left[\dot{m}_l \left(\boldsymbol{Z}_i^T \boldsymbol{\beta}_l^0, \boldsymbol{\beta}^0 \right) X_{il} \boldsymbol{J}_l^{0T} \tilde{\boldsymbol{Z}}_i \right]_{l=1}^{d}.$$

For any $\boldsymbol{\beta}_{-1}$ such that $\left\|\boldsymbol{\beta}_{-1}-\boldsymbol{\beta}_{-1}^0\right\|=\Delta\sqrt{p_n/n}$, we have

$$|I_{n1}|\leq \Delta\sqrt{p_n/n}\left\|\boldsymbol{S}_n\left(\boldsymbol{\beta}_{-1}^0\right)\right\|_2+o_p(p_n).$$

Note that $E\{\boldsymbol{S}_n\left(\boldsymbol{\beta}_{-1}^0\right)\}=\mathbf{0}$, and

$$\begin{aligned}E\left[\left\|\boldsymbol{S}_n\left(\boldsymbol{\beta}_{-1}^0\right)\right\|_2^2\right]&=E\left[\sum_{i=1}^n\psi_\tau^2\left\{\varepsilon_i\right\}tr\left(\left\{\left[\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\tilde{\boldsymbol{Z}}_i\right]_{l=1}^d\right\}^{\otimes 2}\right)\right]\\&=O\left(np_n\right).\end{aligned}$$

Thus, $\left\|\boldsymbol{S}_n\left(\boldsymbol{\beta}_{-1}^0\right)\right\|_2=O_p\left(\sqrt{np_n}\right)$ and $|I_{n1}|\leq\Delta O_p\left(p_n\right)$. Next, by Taylor expansion, (2.3) and Lemmas 5 and 6, we have

$$\begin{aligned}I_{n2}&=\left(\boldsymbol{\beta}_{-1}-\boldsymbol{\beta}_{-1}^0\right)^T\left[\sum_{i=1}^n\frac{1}{h}K\left(-\frac{\varepsilon_i}{h}\right)\left\{\left[\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\tilde{\boldsymbol{Z}}_i\right]_{l=1}^d\right\}^{\otimes 2}\right]\\&\quad\times\left(\boldsymbol{\beta}_{-1}-\boldsymbol{\beta}_{-1}^0\right)\left\{1+o_p\left(1\right)\right\}\\&\to n\left(\boldsymbol{\beta}_{-1}-\boldsymbol{\beta}_{-1}^0\right)^TE\left[f_\varepsilon\left(0\,|\boldsymbol{X}_i,\boldsymbol{Z}_i\right)\left\{\left[\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\tilde{\boldsymbol{Z}}_i\right]_{l=1}^d\right\}^{\otimes 2}\right]\\&\quad\times\left(\boldsymbol{\beta}_{-1}-\boldsymbol{\beta}_{-1}^0\right)\left\{1+o_p\left(1\right)\right\}\\&\leq Cn\rho_{max}\left(E\left\{\left[\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\tilde{\boldsymbol{Z}}_i\right]_{l=1}^d\right\}^{\otimes 2}\right)\left\|\boldsymbol{\beta}_{-1}-\boldsymbol{\beta}_{-1}^0\right\|_2^2\\&=O_p\left(\Delta^2p_n\right)\end{aligned}$$

For sufficiently large Δ, $\left(\boldsymbol{\beta}_{-1}-\boldsymbol{\beta}_{-1}^0\right)^T\mathcal{R}_{\tau nh}\left(\boldsymbol{\beta}_{-1}\right)$ on $\left\{\boldsymbol{\beta}_{-1}:\left\|\boldsymbol{\beta}_{-1}-\boldsymbol{\beta}_{-1}^0\right\|_2=\Delta\sqrt{p_n/n}\right\}$ is dominated by I_{n2}, which is large and positive for all sufficiently large n. Thus, (A.38) holds.

Next, we need prove (ii). Let

$$\Gamma_{ni}=-n^{-1/2}\boldsymbol{e}_n^T\boldsymbol{M}^{-1/2}\left(\boldsymbol{\beta}_{-1}^0\right)\psi_\tau\left\{\varepsilon_i\right\}\left[\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\tilde{\boldsymbol{Z}}_i\right]_{l=1}^d$$

and

$$n^{-1/2}\boldsymbol{e}_n^T\boldsymbol{M}^{-1/2}\left(\boldsymbol{\beta}_{-1}^0\right)\boldsymbol{S}_n\left(\boldsymbol{\beta}_{-1}^0\right)=\sum_{i=1}^n\ \Gamma_{ni}.$$

Since $E(\psi_\tau\left\{\varepsilon_i\right\})=0$ and $Cov\left(n^{-1/2}\boldsymbol{S}_n\left(\boldsymbol{\beta}_{-1}^0\right)\right)=\tau(1-\tau)\boldsymbol{M}\left(\boldsymbol{\beta}_{-1}^0\right)$, we have

$$E\left(n^{-1/2}\boldsymbol{e}_n^T\boldsymbol{M}^{-1/2}\left(\boldsymbol{\beta}_{-1}^0\right)\boldsymbol{S}_n\left(\boldsymbol{\beta}_{-1}^0\right)\right)=0$$

and

$$Var\left(n^{-1/2}\boldsymbol{e}_n^T\boldsymbol{M}^{-1/2}\left(\boldsymbol{\beta}_{-1}^0\right)\boldsymbol{S}_n\left(\boldsymbol{\beta}_{-1}^0\right)\right)=\tau(1-\tau).$$

To establish the asymptotic normality, it suffices to check the Lindberg condition, that is, $\forall\delta>0$, $\sum_{i=1}^{n}E\left[\Gamma_{ni}^2I\left(\left|\Gamma_{ni}\right|>\delta\right)\right]\to 0$. For any $\delta>0$,

$$\sum_{i=1}^{n}E\left[\Gamma_{ni}^2I\left(\left|\Gamma_{ni}\right|>\delta\right)\right]\leq n\left\{E\left(\Gamma_{ni}^4\right)\right\}^{1/2}\left\{P\left(\left|\Gamma_{ni}\right|>\delta\right)\right\}^{1/2}, \tag{A.39}$$

where

$$\begin{aligned}&E\Gamma_{ni}^2\\&=E\left[n^{-1}\boldsymbol{e}_n^T\boldsymbol{M}^{-1/2}\left(\boldsymbol{\beta}_{-1}^0\right)\left\{\left[\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\tilde{\boldsymbol{Z}}_i\right]_{l=1}^d\right\}^{\otimes 2}\boldsymbol{M}^{-1/2}\left(\boldsymbol{\beta}_{-1}^0\right)\boldsymbol{e}_n\left(\psi_\tau\left\{\varepsilon_i\right\}\right)^2\right]\\&\triangleq\tau(1-\tau)\gamma_{ni},\end{aligned}$$

where

$$\gamma_{ni}=E\left\{n^{-1}\boldsymbol{e}_n^T\boldsymbol{M}^{-1/2}\left(\boldsymbol{\beta}_{-1}^0\right)\left\{\left[\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\tilde{\boldsymbol{Z}}_i\right]_{l=1}^d\right\}^{\otimes 2}\boldsymbol{M}^{-1/2}\left(\boldsymbol{\beta}_{-1}^0\right)\boldsymbol{e}_n\right\}.$$

Next, we will prove that $\max_{1\leq i\leq n}\gamma_{ni}\to 0$ as $n\to\infty$. Note that

$$\gamma_{ni}\leq n^{-1}\rho_{\max}\left(E\left\{\left[\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\tilde{\boldsymbol{Z}}_i\right]_{l=1}^d\right\}^{\otimes 2}\right)\rho_{\min}^{-1}\left(\boldsymbol{M}\left(\boldsymbol{\beta}_{-1}^0\right)\right).$$

Therefore, we have

$$\gamma_{ni}\leq n^{-1}\frac{\rho_{\max}\left(E\left\{\left[\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\tilde{\boldsymbol{Z}}_i\right]_{l=1}^d\right\}^{\otimes 2}\right)}{\rho_{\min}\left(E\left\{\left[\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\tilde{\boldsymbol{Z}}_i\right]_{l=1}^d\right\}^{\otimes 2}\right)}\leq n^{-1}\frac{tr\left(E\left\{\left[\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\tilde{\boldsymbol{Z}}_i\right]_{l=1}^d\right\}^{\otimes 2}\right)}{\rho_{\min}\left(E\left\{\left[\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\tilde{\boldsymbol{Z}}_i\right]_{l=1}^d\right\}^{\otimes 2}\right)}.$$

It follows that $\max_{1\leq i\leq n}\gamma_{ni}\leq O\left(p_n/n\right)$. By Cauchy inequality, we have

$$P\left(\left|\Gamma_{ni}\right|>\delta\right)\leq\frac{E\left(\Gamma_{ni}^2\right)}{\delta^2}=O\left(\frac{p_n}{n}\right), \tag{A.40}$$

and

$$\begin{aligned}&E\left(\Gamma_{ni}^4\right)\\&\leq Cn^{-2}\rho_{\max}^2\left(E\left\{\left[\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\tilde{\boldsymbol{Z}}_i\right]_{l=1}^d\right\}^{\otimes 2}\right)\rho_{\min}^{-2}\left(E\boldsymbol{M}\left(\boldsymbol{\beta}_{-1}^0\right)\right)\\&\leq Cn^{-2}\left\{tr\left(\left\{\left[\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0,\boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\tilde{\boldsymbol{Z}}_i\right]_{l=1}^d\right\}^{\otimes 2}\right)\right\}^2\rho_{\min}^{-2}\left(\boldsymbol{M}\left(\boldsymbol{\beta}_{-1}^0\right)\right)\\&=O\left(\frac{p_n^2}{n^2}\right).\end{aligned} \tag{A.41}$$

Combining (A.39)–(A.41) and condition $n^{-1}p_n^3 = o(1)$, we have

$$\sum_{i=1}^{n} E\left[\Gamma_{ni}^2 I\left(|\Gamma_{ni}| > \delta\right)\right] = O\left(\sqrt{\frac{p_n^3}{n}}\right) = o\left(1\right).$$

Thus, Γ_{ni} satisfies the conditions of the Lindeberg-Feller central limit theorem (see van der Vaart (1998)). This also means that $\sum_{i=1}^{n}\Gamma_{ni} = n^{-1/2}\boldsymbol{e}_n^T\boldsymbol{M}^{-1/2}\left(\boldsymbol{\beta}_{-1}^0\right)\boldsymbol{S}_n\left(\boldsymbol{\beta}_{-1}^0\right)$ has an asymptotic multivariate normal distribution

$$n^{-1/2}\boldsymbol{e}_n^T\boldsymbol{M}^{-1/2}\left(\boldsymbol{\beta}_{-1}^0\right)\boldsymbol{S}_n\left(\boldsymbol{\beta}_{-1}^0\right)\xrightarrow{d} N\left(0, \tau(1-\tau)\right). \tag{A.42}$$

By (2.3) and Lemmas 5 and 6, it is straightforward to prove that

$$\begin{aligned}
&\partial\mathcal{R}_{\tau nh}\left(\boldsymbol{\beta}_{-1}^0\right)/\partial\boldsymbol{\beta}_{-1}^T\\
&= \sum_{i=1}^{n} f_\varepsilon\left(0\,|\boldsymbol{X}_i, \boldsymbol{Z}_i\right)\left\{\left[\dot{m}_l\left(\boldsymbol{Z}_i^T\boldsymbol{\beta}_l^0, \boldsymbol{\beta}^0\right)X_{il}\boldsymbol{J}_l^{0T}\tilde{\boldsymbol{Z}}_i\right]_{l=1}^{d}\right\}^{\otimes 2} + o_p\left(n\right)\\
&\triangleq n\boldsymbol{H}\left(\boldsymbol{\beta}_{-1}^0\right) + o_p\left(n\right).
\end{aligned} \tag{A.43}$$

Thus by Taylor series expansion, Lemma 6 and (A.43), we have

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{-1} - \boldsymbol{\beta}_{-1}^0 &= -\left\{\partial\mathcal{R}_{\tau nh}\left(\boldsymbol{\beta}_{-1}^0\right)/\partial\boldsymbol{\beta}_{-1}^T\right\}^{-1}\mathcal{R}_{\tau nh}\left(\boldsymbol{\beta}_{-1}^0\right)\left\{1 + o_p\left(1\right)\right\}\\
&= -n^{-1}\boldsymbol{H}^{-1}\left(\boldsymbol{\beta}_{-1}^0\right)\boldsymbol{S}_n\left(\boldsymbol{\beta}_{-1}^0\right)\left\{1 + o_p\left(1\right)\right\}
\end{aligned}$$

and

$$\begin{aligned}
&n^{1/2}\boldsymbol{e}_n^T\boldsymbol{M}^{-1/2}\left(\boldsymbol{\beta}_{-1}^0\right)\boldsymbol{H}\left(\boldsymbol{\beta}_{-1}^0\right)\left(\hat{\boldsymbol{\beta}}_{-1} - \boldsymbol{\beta}_{-1}^0\right)\\
&= -n^{-1/2}\boldsymbol{e}_n^T\boldsymbol{M}^{-1/2}\left(\boldsymbol{\beta}_{-1}^0\right)\boldsymbol{S}_n\left(\boldsymbol{\beta}_{-1}^0\right)\left\{1 + o_p\left(1\right)\right\}\\
&\xrightarrow{d} N\left(0, \tau(1-\tau)\right).
\end{aligned}$$

Therefore, we complete the proof. □

Proof of Theorem 2 Since $\left\|\hat{\boldsymbol{\beta}}_{-1} - \boldsymbol{\beta}_{-1}^0\right\|_2 = O_p\left(\sqrt{p_n/n}\right)$, Theorem 2 follows from this result and Lemma 2. □

Proof of Theorem 3 Let $\alpha_n = O\left(\sqrt{p_n}\left(n^{-1/2} + a_n\right)\right)$. It suffices to show that $\forall\delta > 0$, there exists a sufficiently large constant $\Delta > 0$ such that for sufficiently large n,

$$P\left\{\inf_{\left\|\boldsymbol{\beta}_{-1} - \boldsymbol{\beta}_{-1}^0\right\|_2 = \Delta\alpha_n}\left(\boldsymbol{\beta}_{-1} - \boldsymbol{\beta}_{-1}^0\right)^T\left[\mathcal{R}_{\tau nh}\left(\boldsymbol{\beta}_{-1}\right) + n\boldsymbol{b}_{\alpha_1}\left(\boldsymbol{\beta}_{-1}\right)\right] > 0\right\} \geq 1 - \delta. \tag{A.44}$$

From the proof of Theorem 1 (i), we can obtain $\left(\boldsymbol{\beta}_{-1}-\boldsymbol{\beta}_{-1}^0\right)^T\mathcal{R}_{\tau nh}\left(\boldsymbol{\beta}_{-1}\right) = O_p\left(n\alpha_n^2\Delta^2\right)$.

Define

$$\boldsymbol{b}_{\boldsymbol{\alpha_1}}(\boldsymbol{\beta}_{-1}^0) = \left[\dot{p}_{\alpha_1}\left(|\beta_{12}^0|\right)\operatorname{sgn}\left(\beta_{12}^0\right), ..., \dot{p}_{\alpha_1}\left(\left|\beta_{1p_n}^0\right|\right)\operatorname{sgn}\left(\beta_{1p_n}^0\right)\right.$$
$$\left. ..., \dot{p}_{\alpha_1}\left(|\beta_{d2}^0|\right)\operatorname{sgn}\left(\beta_{d2}^0\right), ..., \dot{p}_{\alpha_1}\left(\left|\beta_{dp_n}^0\right|\right)\operatorname{sgn}\left(\beta_{dp_n}^0\right)\right]^T$$

and

$$\boldsymbol{\Sigma}_{\alpha_1}(\boldsymbol{\beta}_{-1}^0) = diag\left\{\ddot{p}_{\alpha_1}\left(\left|\beta_{12}^0\right|\right), ..., \ddot{p}_{\alpha_1}\left(\left|\beta_{1p_n}^0\right|\right), ..., \ddot{p}_{\alpha_1}\left(\left|\beta_{dp_n}^0\right|\right)\right\}.$$

By the definition of $\boldsymbol{b}_{\alpha_1}\left(\boldsymbol{\beta}_{-1}\right)$ given in section 3, taking Taylor's explanation for $\boldsymbol{b}_{\alpha_1}\left(\boldsymbol{\beta}_{-1}\right)$ at $\boldsymbol{\beta}_{-1}^0$ gives

$$\boldsymbol{b}_{\alpha_1}\left(\boldsymbol{\beta}_{-1}\right) = \boldsymbol{b}_{\alpha_1}\left(\boldsymbol{\beta}_{-1}^0\right) + \boldsymbol{\Sigma}_{\alpha_1}\left(\boldsymbol{\beta}_{-1}^0\right)\{1+o(1)\}\left(\boldsymbol{\beta}_{-1}-\boldsymbol{\beta}_{-1}^0\right).$$

Moreover, by conditions (C9) and (C10), we have

$$\begin{aligned}
&\left|\left(\boldsymbol{\beta}_{-1}-\boldsymbol{\beta}_{-1}^0\right)^T n\boldsymbol{b}_{\alpha_1}\left(\boldsymbol{\beta}_{-1}\right)\right| \\
&\leq \left|\left(\boldsymbol{\beta}_{-1}-\boldsymbol{\beta}_{-1}^0\right)^T n\boldsymbol{b}_{\alpha_1}\left(\boldsymbol{\beta}_{-1}^0\right)\right| \\
&\quad + \left|\left(\boldsymbol{\beta}_{-1}-\boldsymbol{\beta}_{-1}^0\right)^T n\boldsymbol{\Sigma}_{\alpha_1}\left(\boldsymbol{\beta}_{-1}^0\right)\{1+o(1)\}\left(\boldsymbol{\beta}_{-1}-\boldsymbol{\beta}_{-1}^0\right)\right| \\
&\leq n\sqrt{d(p_n-1)}\max_{l=1,...,d,j=2,...,p_n}\dot{p}_{\alpha_1}\left(\left|\beta_{lj}^0\right|\right)\left\|\boldsymbol{\beta}_{-1}-\boldsymbol{\beta}_{-1}^0\right\|_2 \\
&\quad + n\max_{l=1,...,d,j=2,...,p_n}\left|\ddot{p}_{\alpha_1}\left(\left|\beta_{lj}^0\right|\right)\right|\left\|\boldsymbol{\beta}_{-1}-\boldsymbol{\beta}_{-1}^0\right\|_2^2 \\
&\leq na_n\sqrt{d(p_n-1)}\left\|\boldsymbol{\beta}_{-1}-\boldsymbol{\beta}_{-1}^0\right\|_2 + nb_n\left\|\boldsymbol{\beta}_{-1}-\boldsymbol{\beta}_{-1}^0\right\|_2^2 \\
&= O_p\left(n\alpha_n^2\Delta\right) + o_p\left(n\alpha_n^2\Delta^2\right).
\end{aligned}$$

Thus $\left(\boldsymbol{\beta}_{-1}-\boldsymbol{\beta}_{-1}^0\right)^T[\mathcal{R}_{\tau nh}\left(\boldsymbol{\beta}_{-1}\right) + n\boldsymbol{b}_{\alpha_1}\left(\boldsymbol{\beta}_{-1}\right)]$ is dominated by $\left(\boldsymbol{\beta}_{-1}-\boldsymbol{\beta}_{-1}^0\right)^T\mathcal{R}_{\tau nh}\left(\boldsymbol{\beta}_{-1}\right)$ for sufficiently large Δ, which is large and positive for all sufficiently large n. Thus, (A.44) holds.

Proof of Theorem 4. From Theorem 3 for a sufficiently large $C>0$, $\bar{\boldsymbol{\beta}}_{\alpha_1,-1}$ lies in the ball $\left\{\boldsymbol{\beta}_{-1}^0 + \alpha_n\boldsymbol{u} : \|\boldsymbol{u}\|_2 \leq C\right\}$ with probability converging to 1, where $\alpha_n = \sqrt{p_n}\left(n^{-1/2}+a_n\right)$. From the proof of Theorem 1, we have $\left\|\mathcal{R}_{\tau nh}\left(\boldsymbol{\beta}_{-1}^0\right)\right\|_2 = O_p\left(\sqrt{np_n}\right)$. By the assumption that $\boldsymbol{\beta}_{-1}-\boldsymbol{\beta}_{-1}^0 = O_p(\sqrt{p_n/n})$, and for some small $\epsilon_n = C\alpha_n$ and $l=1,...,d$, $j=s_l+1,...,p_n$, we have

$$\mathcal{R}_{\tau nh}\left(\beta_{lj}\right) + nb_{\alpha_1}\left(\beta_{lj}\right) = n\alpha_1\left\{O_p\left(\frac{\sqrt{p_n/n}}{\alpha_1}\right) + \frac{\dot{p}_{\alpha_1}\left(|\beta_{lj}|\right)}{\alpha_1}\operatorname{sgn}\left(\beta_{lj}\right)\right\}. \tag{A.45}$$

Using $\sqrt{p_n/n}/\alpha_1 \to 0$ and condition (C8), the sign of β_{lj} determines the sign of $\mathcal{R}_{\tau nh}(\boldsymbol{\beta}_{-1}) + nb_{\alpha_1}(\boldsymbol{\beta}_{-1})$. Hence, (A.45) implies that

$$\mathcal{R}_{\tau nh}(\beta_{lj}) + nb_{\alpha_1}(\beta_{lj}) = \begin{cases} > 0, 0 < \beta_{lj} < \epsilon_n \\ < 0, -\epsilon_n < \beta_{lj} < 0, \end{cases}$$

for $l = 1, ..., d$, $j = s_l + 1, ..., p_n$. This implies $\bar{\beta}_{\alpha_1, lj} = 0$ with probability converging to 1 for $l = 1, ..., d$, $j = s_l + 1, ..., p_n$. This completes the proof of (i). Next we need to prove (ii). As shown in Theorem 3, $\bar{\boldsymbol{\beta}}_{\alpha_1,-1}$ is $\sqrt{n/p_n}$ consistent. By the proof of (i), each component of $\bar{\boldsymbol{\beta}}^{(1)}_{\alpha_1,-1}$ stays away from zero for a sufficiently large n. At the same time $\bar{\boldsymbol{\beta}}^{(2)}_{\alpha_1,-1} = 0$ with probability tending to 1. Thus, $\bar{\boldsymbol{\beta}}^{(1)}_{\alpha_1,-1}$ satisfies

$$n^{-1}\mathcal{R}_{\tau nh}\left(\boldsymbol{\beta}^{(1)}_{-1}\right) + \boldsymbol{b}_{\alpha_1}\left(\boldsymbol{\beta}^{(1)}_{-1}\right) = 0,$$

where $\boldsymbol{b}_{\alpha_1}\left(\boldsymbol{\beta}^{(1)}_{-1}\right) = \left(\dot{p}_{\alpha_1}(|\beta_{lj}|)\,\mathrm{sgn}(\beta_{lj})\right)_{l=1,j=2}^{d,s_l}$ is a $(s_1 + ... + s_d - d) \times 1$ vector. Applying the Taylor expansion to $\boldsymbol{b}_{\alpha_1}\left(\boldsymbol{\beta}^{(1)}_{-1}\right)$, we get that

$$\dot{p}_{\alpha_1}(|\beta_{lj}|)\,\mathrm{sgn}(\beta_{lj}) = \dot{p}_{\alpha_1}\left(\left|\beta^0_{lj}\right|\right)\mathrm{sgn}\left(\beta^0_{lj}\right) + \ddot{p}_{\alpha_1}\left(\left|\beta^0_{lj}\right|\right)\{1 + o(1)\}\left(\beta_{lj} - \beta^0_{lj}\right)$$

for $l = 1, ..., d, j = 2, ..., s_l$. By the condition (C12), for $l = 1, ..., d, j = 2, ..., s_l$, we can obtain $\dot{p}_{\alpha_1}\left(\left|\beta^0_{lj}\right|\right) = 0$ as $\alpha_1 \to 0$, together with condition (C10), some simple calculations yields

$$n^{-1}\mathcal{R}_{\tau nh}\left(\boldsymbol{\beta}^{(1)}_{-1}\right) + o_p\left(\boldsymbol{\beta}^{(1)}_{-1} - \boldsymbol{\beta}^{0(1)}_{-1}\right) = 0.$$

By Taylor expansion, Lemma 6, and the previous result,

$$\begin{aligned} &\sqrt{n}\left(\boldsymbol{\beta}^{(1)}_{-1} - \boldsymbol{\beta}^{0(1)}_{-1}\right) \\ &= -\left[n^{-1}\partial\mathcal{R}_{\tau nh}\left(\boldsymbol{\beta}^{0(1)}_{-1}\right)\Big/\partial\boldsymbol{\beta}^{(1)T}_{-1}\right]^{-1} n^{-1/2}\mathcal{R}_{\tau nh}\left(\boldsymbol{\beta}^{0(1)}_{-1}\right)\{1 + o_p(1)\} \\ &= -(\mathbb{H}^{(1)})^{-1} n^{-1/2}\boldsymbol{S}_n\left(\boldsymbol{\beta}^{0(1)}_{-1}\right) + o_p(1), \end{aligned}$$

where $\boldsymbol{S}_n\left(\boldsymbol{\beta}^{0(1)}_{-1}\right)$ is $\sum_{l=1}^{d}(s_l - 1) \times 1$ sub-vector of $\boldsymbol{S}_n\left(\boldsymbol{\beta}^0_{-1}\right)$ corresponding to $\boldsymbol{\beta}^{0(1)}_{-1}$. Thus, by the central limit theorem and Slutsky's theorem, we can complete the proof of Theorem 4 (ii). □

Proof of Theorem 5. Let $\boldsymbol{\lambda}^0 = \left\{(\boldsymbol{\lambda}_l^0)^T : 1 \leq l \leq d\right\}^T$, $k_n = J_n n^{-1/2} + J_n^{-r+1/2}$. Our aim is to show that for any given $\epsilon > 0$ there is a large constant $C > 0$ such that, for a sufficiently large n, we have

$$P\left\{\inf_{\|\boldsymbol{v}\|_2 = C} \mathcal{L}_{\tau n}^{*}\left(\boldsymbol{\lambda}^0 + k_n \boldsymbol{v}, \boldsymbol{\beta}^0\right) > \mathcal{L}_{\tau n}^{*}\left(\boldsymbol{\lambda}^0, \boldsymbol{\beta}^0\right)\right\} \geq 1 - \epsilon. \tag{A.46}$$

This implies with probability of at least $1-\epsilon$ that there exists a local minimizer in the ball $\{\boldsymbol{\lambda}^0 + k_n \boldsymbol{v} : \|\boldsymbol{v}\|_2 \leq C\}$. This in turn implies that there exists a local minimizer $\bar{\boldsymbol{\lambda}}(\boldsymbol{\beta}^0)$ such that $\left\|\bar{\boldsymbol{\lambda}}(\boldsymbol{\beta}^0) - \boldsymbol{\lambda}^0\right\|_2 = O_p(k_n)$. We can find a constant $c > 0$ such that $\|\boldsymbol{\lambda}_l^0\|_{\boldsymbol{D}} = \sqrt{\boldsymbol{\lambda}_l^{0T} \boldsymbol{D} \boldsymbol{\lambda}_l^0} > c$ for $1 \leq l \leq d_1$. Because $\|(\boldsymbol{\lambda}_l^0 + k_n \boldsymbol{v}_l) - \boldsymbol{\lambda}_l^0\|_2^2 = o_p(J_n)$ and that $\alpha_2 = o(1)$, we have

$$P\left\{p_{\alpha_2}\left(\left\|\boldsymbol{\lambda}_l^0 + k_n \boldsymbol{v}_l\right\|_{\boldsymbol{D}}\right) = p_{\alpha_2}\left(\left\|\boldsymbol{\lambda}_l^0\right\|_{\boldsymbol{D}}\right)\right\} \to 1 \text{ if } 1 \leq l \leq d_1.$$

This fact implies that

$$n \sum_{l=1}^{d} p_{\alpha_2}\left(\left\|\boldsymbol{\lambda}_l^0 + k_n \boldsymbol{v}_l\right\|_{\boldsymbol{D}}\right) - n \sum_{l=1}^{d} p_{\alpha_2}\left(\left\|\boldsymbol{\lambda}_l^0\right\|_{\boldsymbol{D}}\right) \geq 0$$

with probability tending to 1 since $\|\boldsymbol{\lambda}_l^0\|_{\boldsymbol{D}} = \sqrt{\boldsymbol{\lambda}_l^0 \boldsymbol{D} \boldsymbol{\lambda}_l^0} = 0$ for $d_1 + 1 \leq l \leq d$. By the proof of Lemma 2 and the previous result, we have

$$\begin{aligned}
&\mathcal{L}_{\tau n}^{*}(\boldsymbol{\lambda}^0 + k_n \boldsymbol{v}, \boldsymbol{\beta}^0) - \mathcal{L}_{\tau n}^{*}(\boldsymbol{\lambda}^0, \boldsymbol{\beta}^0) \\
&= \mathcal{L}_{\tau n}(\boldsymbol{\lambda}^0 + k_n \boldsymbol{v}, \boldsymbol{\beta}^0) - \mathcal{L}_{\tau n}(\boldsymbol{\lambda}^0, \boldsymbol{\beta}^0) + n \sum_{l=1}^{d} p_{\alpha_2}(\|\boldsymbol{\lambda}_l^0 + k_n \boldsymbol{v}_l\|_{\boldsymbol{D}}) - n \sum_{l=1}^{d} p_{\alpha_2}(\|\boldsymbol{\lambda}_l^0\|_{\boldsymbol{D}}) \\
&\geq \mathcal{L}_{\tau n}(\boldsymbol{\lambda}^0 + k_n \boldsymbol{v}, \boldsymbol{\beta}^0) - \mathcal{L}_{\tau n}(\boldsymbol{\lambda}^0, \boldsymbol{\beta}^0). \\
&> 0.
\end{aligned}$$

Hence, (A.46) holds. This implies that $\left\|\bar{\boldsymbol{\lambda}}(\boldsymbol{\beta}^0) - \boldsymbol{\lambda}^0\right\|_2 = O_p(k_n)$. This result combines with (A.1) and triangle inequality, for $1 \leq l \leq d$, we have with probability approaching 1,

$$\begin{aligned}
&\|\bar{m}_l(u_l, \boldsymbol{\beta}^0) - m_l(u_l)\|^2 \\
&= \int_a^b \left\{\boldsymbol{B}^T(u)\bar{\boldsymbol{\lambda}}_l(\boldsymbol{\beta}^0) - \boldsymbol{B}^T(u)\boldsymbol{\lambda}_l^0 + \boldsymbol{B}^T(u)\boldsymbol{\lambda}_l^0 - m_l(u)\right\}^2 du \\
&\leq 2\int_a^b \left\{\boldsymbol{B}^T(u)\bar{\boldsymbol{\lambda}}_l(\boldsymbol{\beta}^0) - \boldsymbol{B}^T(u)\boldsymbol{\lambda}_l^0\right\}^2 du + 2\int_a^b \left\{\boldsymbol{B}^T(u)\boldsymbol{\lambda}_l^0 - m_l(u)\right\}^2 du \\
&= 2\left(\bar{\boldsymbol{\lambda}}_l(\boldsymbol{\beta}^0) - \boldsymbol{\lambda}_l^0\right)^T \int_a^b \boldsymbol{B}(u)\boldsymbol{B}^T(u)\,du\left(\bar{\boldsymbol{\lambda}}_l(\boldsymbol{\beta}^0) - \boldsymbol{\lambda}_l^0\right) + 2\int_a^b \left\{\boldsymbol{B}^T(u)\boldsymbol{\lambda}_l^0 - m_l(u)\right\}^2 du \\
&\asymp J_n^{-1}\left\|\bar{\boldsymbol{\lambda}}_l(\boldsymbol{\beta}^0) - \boldsymbol{\lambda}_l^0\right\|_2^2 + O(J_n^{-2r}) \\
&= O_p(J_n n^{-1} + J_n^{-2r}).
\end{aligned}$$

Thus, for $1 \le l \le d$, we have $|\bar{m}_l(u_l, \boldsymbol{\beta}^0) - m_l(u_l)| = O_p\left(\sqrt{J_n/n} + J_n^{-r}\right)$ uniformly for any $u_l \in [a,b]$. Theorem 5 follows from this result and Theorem 3. □

Proof of Theorem 6. Suppose for some $l = 1 + d_1, ..., d$, $\boldsymbol{B}\left(\boldsymbol{Z}_i^T \bar{\boldsymbol{\beta}}_{\alpha_1 l}\right)^T \bar{\boldsymbol{\lambda}}_l$ is not a linear function. Define $\bar{\boldsymbol{\lambda}}^*$ to be the same as $\bar{\boldsymbol{\lambda}}$ except that $\bar{\boldsymbol{\lambda}}_l$ is replaced by its projection onto the subspace $\left\{\boldsymbol{\lambda}_l : \boldsymbol{B}\left(\boldsymbol{Z}_i^T \bar{\boldsymbol{\beta}}_{\alpha_1 l}\right)^T \boldsymbol{\lambda}_l \text{ represents a linear function}\right\}$. By definition of $\bar{\boldsymbol{\lambda}}$ and $\bar{\boldsymbol{\lambda}}^*$, and $\bar{\boldsymbol{\lambda}}_l^{*T} \boldsymbol{D} \bar{\boldsymbol{\lambda}}_l^* = 0$. we have

$$\begin{aligned}
0 &\ge \mathcal{L}_{\tau n}^*\left(\bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\beta}}_{\alpha_1}\right) - \mathcal{L}_{\tau n}^*\left(\bar{\boldsymbol{\lambda}}^*, \bar{\boldsymbol{\beta}}_{\alpha_1}\right) \\
&= \sum_{i=1}^n \rho_\tau\left\{Y_i - \sum_{l=1}^d \boldsymbol{B}\left(\boldsymbol{Z}_i^T \bar{\boldsymbol{\beta}}_{\alpha_1 l}\right)^T \bar{\boldsymbol{\lambda}}_l X_{il}\right\} \\
&\quad - \sum_{i=1}^n \rho_\tau\left\{Y_i - \sum_{l=1}^d \boldsymbol{B}\left(\boldsymbol{Z}_i^T \bar{\boldsymbol{\beta}}_{\alpha_1 l}\right)^T \bar{\boldsymbol{\lambda}}_l^* X_{il}\right\} + n p_{\alpha_2}\left(\sqrt{\bar{\boldsymbol{\lambda}}_l^T \boldsymbol{D} \bar{\boldsymbol{\lambda}}_l}\right) \\
&= I - II + III.
\end{aligned} \tag{A.47}$$

Let $\Psi_i = \sum_{l=1}^d \boldsymbol{B}\left(\boldsymbol{Z}_i^T \bar{\boldsymbol{\beta}}_{\alpha_1 l}\right)^T\left(\bar{\boldsymbol{\lambda}}_l^* - \boldsymbol{\lambda}_l^0\right) X_{il} + \sum_{l=1}^d \left(\boldsymbol{B}\left(\boldsymbol{Z}_i^T \bar{\boldsymbol{\beta}}_{\alpha_1 l}\right)^T \boldsymbol{\lambda}_l^0 - m_l\left(\boldsymbol{Z}_i^T \boldsymbol{\beta}_l^0\right)\right) X_{il}$. Since $\rho_\tau(u) - \rho_\tau(v) \ge 2(\tau - I(v \le 0))(u - v)$ for any $u, v \in \mathbb{R}$, we have

$$\begin{aligned}
I - II &\ge -\sum_{i=1}^n (2\tau - 2I(\varepsilon_i \le 0)) \sum_{l=1}^d \boldsymbol{B}\left(\boldsymbol{Z}_i^T \bar{\boldsymbol{\beta}}_{\alpha_1 l}\right)^T\left(\bar{\boldsymbol{\lambda}}_l - \bar{\boldsymbol{\lambda}}_l^*\right) X_{il} \\
&\quad - \sum_{i=1}^n (2I(\varepsilon_i \le 0) - 2I(\varepsilon_i \le \Psi_i)) \sum_{l=1}^d \boldsymbol{B}\left(\boldsymbol{Z}_i^T \bar{\boldsymbol{\beta}}_{\alpha_1 l}\right)^T\left(\bar{\boldsymbol{\lambda}}_l - \bar{\boldsymbol{\lambda}}_l^*\right) X_{il} \\
&\ge \left(-\left\|\sum_{i=1}^n (2\tau - 2I(\varepsilon_i \le 0)) \sum_{l=1}^d \boldsymbol{B}\left(\boldsymbol{Z}_i^T \bar{\boldsymbol{\beta}}_{\alpha_1 l}\right) X_{il}\right\|_2 \right. \\
&\quad \left. -\left\|\sum_{i=1}^n (2I(\varepsilon_i \le 0) - 2I(\varepsilon_i \le \Psi_i)) \sum_{l=1}^d \boldsymbol{B}\left(\boldsymbol{Z}_i^T \bar{\boldsymbol{\beta}}_{\alpha_1 l}\right) X_{il}\right\|_2\right) \left\|\bar{\boldsymbol{\lambda}}_l - \bar{\boldsymbol{\lambda}}_l^*\right\|_2 \\
&= O_p\left\{-\left[\sqrt{n/J_n} + nJ_n^{-1/2}\left(\sqrt{J_n/n} + J_n^{-r}\right)\right]\right\} \left\|\bar{\boldsymbol{\lambda}} - \bar{\boldsymbol{\lambda}}^*\right\|_2 \\
&= O_p\left\{-\left[\sqrt{n} + nJ_n^{-r-1/2}\right]\right\} \left\|\bar{\boldsymbol{\lambda}} - \bar{\boldsymbol{\lambda}}^*\right\|_2
\end{aligned} \tag{A.48}$$

since $\max_{1 \le i \le n} \Psi_i = O_p\left(\sqrt{J_n/n} + J_n^{-r}\right)$. On the other hand,

$$\sqrt{\bar{\boldsymbol{\lambda}}_l^T \boldsymbol{D} \bar{\boldsymbol{\lambda}}_l} = \sqrt{\left(\bar{\boldsymbol{\lambda}}_l - \boldsymbol{\lambda}_l^0\right)^T \boldsymbol{D}\left(\bar{\boldsymbol{\lambda}}_l - \boldsymbol{\lambda}_l^0\right)} = O_p\left((J_n/n)^{1/2} + J_n^{-r}\right) = o(\alpha_2)$$

by the proof of Theorem 5 and assumption $\left((J_n/n)^{1/2} + J_n^{-r}\right)^{-1} \alpha_2 \to \infty$, we have

$$p_{\alpha_2}\left(\sqrt{\bar{\boldsymbol{\lambda}}_l^T \boldsymbol{D} \bar{\boldsymbol{\lambda}}_l}\right) = \alpha_2 \sqrt{\bar{\boldsymbol{\lambda}}_l^T \boldsymbol{D} \bar{\boldsymbol{\lambda}}_l} \tag{A.49}$$

with probability tending to 1, by the definition of the SCAD penalty function. It remains to show that

$$\left\|\bar{\boldsymbol{\lambda}}-\bar{\boldsymbol{\lambda}}^{*}\right\|_{2}=\left\|\bar{\boldsymbol{\lambda}}_{l}-\bar{\boldsymbol{\lambda}}_{l}^{*}\right\|_{2}=O_{p}\left(\sqrt{J_{n}\bar{\boldsymbol{\lambda}}_{l}^{T}\boldsymbol{D}\bar{\boldsymbol{\lambda}}_{l}}\right). \tag{A.50}$$

In fact, if this is true, by (A.47), (A.49) and (A.50), we have $III=n\alpha_{2}J_{n}^{-1/2}\left\|\bar{\boldsymbol{\lambda}}-\bar{\boldsymbol{\lambda}}^{*}\right\|_{2}$. Then, by the condition $\left((J_{n}/n)^{1/2}+J_{n}^{-r}\right)^{-1}\alpha_{2}\rightarrow\infty$, it is easy to show that III is of higher order than $O_{p}\left\{\left[\sqrt{n}+nJ_{n}^{-r-1/2}\right]\right\}\left\|\bar{\boldsymbol{\lambda}}-\bar{\boldsymbol{\lambda}}^{*}\right\|_{2}$. Thus, it follows that $I-II$ is dominated by III, which is large and positive for all sufficiently large n. Hence, we have a contradiction if $\bar{\boldsymbol{\lambda}}_{l}^{T}\boldsymbol{D}\bar{\boldsymbol{\lambda}}_{l}>0$. That is, with probability approaching 1, $\|\bar{\boldsymbol{\lambda}}_{l}\|_{\boldsymbol{D}}=\sqrt{\bar{\boldsymbol{\lambda}}_{l}^{T}\boldsymbol{D}\bar{\boldsymbol{\lambda}}_{l}}=0$ for $1+d_{1}\leq l\leq d$.

Next, we prove (A.50). Note that $\bar{\boldsymbol{\lambda}}_{l}^{*T}\boldsymbol{D}\bar{\boldsymbol{\lambda}}_{l}^{*}=0$. Furthermore, since $\bar{\boldsymbol{\lambda}}_{l}^{*}$ is the projection of $\bar{\boldsymbol{\lambda}}_{l}$ onto $\left\{\boldsymbol{\lambda}_{l}:\boldsymbol{\lambda}_{l}^{T}\boldsymbol{D}\boldsymbol{\lambda}_{l}=0\right\}$, $\bar{\boldsymbol{\lambda}}_{l}-\bar{\boldsymbol{\lambda}}_{l}^{*}$ is orthogonal to this projection space. Thus $\bar{\boldsymbol{\lambda}}_{l}^{T}\boldsymbol{D}\bar{\boldsymbol{\lambda}}_{l}=\left(\bar{\boldsymbol{\lambda}}_{l}-\bar{\boldsymbol{\lambda}}_{l}^{*}\right)^{T}\boldsymbol{D}\left(\bar{\boldsymbol{\lambda}}_{l}-\bar{\boldsymbol{\lambda}}_{l}^{*}\right)$ and $\left(\bar{\boldsymbol{\lambda}}_{l}-\bar{\boldsymbol{\lambda}}_{l}^{*}\right)^{T}\boldsymbol{D}\left(\bar{\boldsymbol{\lambda}}_{l}-\bar{\boldsymbol{\lambda}}_{l}^{*}\right)\Big/\left\|\bar{\boldsymbol{\lambda}}_{l}-\bar{\boldsymbol{\lambda}}_{l}^{*}\right\|_{2}^{2}$ lies between the minimum and the maximum positive eigenvalues of $\boldsymbol{D}$, which is of order $1/J_{n}$. □

References

Bosq, D. (1998). Nonparametric Statistics for Stochastic Processes. NewYork: Springer-Verlag.

Conover, W. J. (1971). Practical Nonparametric Statistics. New York: John Wiley & Sons.

de Boor, C. (2001). A practical guide to splines. Springer, New York.

Demko, S. (1986). Spectral bounds for $|a^{-1}|_{\infty}$. *J. Approx. Theory* **48**, 207–212.

DeVore, R. A. and Lorentz, G. G. (1993). Constructive approximation. Springer-Verlag, Berlin.

Fan, J. and Peng, H. (2004). Nonconcave penalized likelihood with a diverging number of parameters. *Ann. Statist.* **32**, 928-961.

Horowitz, J. L. (1998). Bootstrap methods for median regression models. *Econometrica* **66**, 1327–1351.

Knight, K. (1998). Limiting distributions for L_1 regression estimators under general conditions. *Ann. Statist.* **26**, 755–770.

Ma, S. and Song, P. X.-K. (2015). Varying index coefficient models. *J. Amer. Statist. Assoc.* **110**, 341–356.

van der Vaart, A. W. (1998). Asymptotic statistics. Cambridge Univ. Press.

Wang, H.J. and Zhu, Z. (2011). Empirical likelihood for quantile regression models with longitudinal data. *J. Stat. Plan. Infer.* **141**, 1603–1615.

Wang, H. J., Zhu, Z. and Zhou, J. (2009). Quantile regression in partially linear varying coefficient models. *Ann. Statist.* **37**, 3841–3866.

Whang, Y. J. (2006). Smoothed empirical likelihood methods for quantile regression models. *Economet. Theor.* **22**, 173–205.

Zou, H. and Zhang, H. H. (2009). On the adaptive Elastic-Net with a diverging number of parameters. *Ann. Statist.* **37**, 1733–1751.

School of Mathematics and Statistics, Southwest University, Chongqing 400715, China

E-mail: (lvjing@swu.edu.cn)

Department of Statistics and Applied Probability, National University of Singapore, Singapore 119077

E-mail: (stalj@nus.edu.sg)