Supplementary file for "Estimation of conditional prevalence from group testing data with missing covariates"

Aurore Delaigle[1], Wei Huang[1] and Shaoke Lei[2]
[1]School of Mathematics and Statistics and Australian Research Council Centre of Excellence for Mathematical and Statistical Frontiers (ACEMS), University of Melbourne, Australia.
[2]Health Services, Murdoch Children's Research Institute and Health Services Research Unit, The Royal Children's Hospital, Australia.

A Side results

A.1 Estimator of q_0

Delaigle and Meister (2011) have proposed to estimate q_0 by likelihood estimator $\widehat{q}_0$, which is defined as the solution of $\Phi(Z_1^*, \ldots, Z_J^*; q) = 0$, with

$$\Phi(Z_1^*, \ldots, Z_J^*; q) = \sum_{j=1}^{J} n_j(Z_j^* - q^{n_j})/(\sum_{k=0}^{n_j-1} q^k)\,.$$

Since their estimator is based only on the Z_j^*'s, which are all available in our case, we can compute it from our data.

A.2 Proof of (3.5), (3.6) and (3.8)

To prove (3.5), using repeatedly the the assumption at (2.2) and the fact that the $(X_{i,j}, \Delta_{i,j}, Y_{i,j})$'s are i.i.d., we have

$$\begin{aligned}\mathbb{P}(\Delta_{i,j} = 1|X_{i,j} = x, Z_j^* = 1) &= \mathbb{P}(\Delta_{i,j} = 1|X_{i,j} = x, Y_{1,j} = 0, Y_{2,j} = 0, \ldots, Y_{n_j,j} = 0)\\ &=\mathbb{P}(\Delta_{i,j} = 1|Y_{i,j} = 0, X_{i,j} = x) = \mathbb{P}(\Delta_{i,j} = 1|Y_{i,j} = 0)\end{aligned}$$

$$
\begin{aligned}
&=\frac{\mathbb{P}(\Delta_{i,j}=1,Y_{i,j}=0)}{\mathbb{P}(Y_{i,j}=0)}=\frac{\mathbb{P}(\Delta_{i,j}=1,Y_{1,j}=0,Y_{2,j}=0,\ldots,Y_{n_j,j}=0)}{\mathbb{P}(Y_{1,j}=0,Y_{2,j}=0,\ldots,Y_{n_j,j}=0)}\\
&=\mathbb{P}(\Delta_{i,j}=1|Z_j^*=1)\equiv p_0\,.
\end{aligned}
$$

To prove (3.6), using the result $\mathbb{P}(A,B|C)=\mathbb{P}(A|B,C)\mathbb{P}(B|C)$, (2.2), and using $p(x)=P(Y_{i,j}=1|X_{i,j}=x)$, we can write

$$
\begin{aligned}
\pi(x)&=\mathbb{P}(\Delta_{i,j}=1|X_{i,j}=x)=\sum_{k=0}^{1}\mathbb{P}(\Delta_{i,j}=1|Y_{i,j}=k,X_{i,j}=x)\mathbb{P}(Y_{i,j}=k|X_{i,j}=x)\\
&=\sum_{k=0}^{1}\mathbb{P}(\Delta_{i,j}=1|Y_{i,j}=k)\mathbb{P}(Y_{i,j}=k|X_{i,j}=x)=p_0\{1-p(x)\}+p_1p(x)\,, \qquad \text{(A.1)}
\end{aligned}
$$

where $p_1=\mathbb{P}(\Delta_{i,j}=1|Y_{i,j}=1)$ and since we have proved above that $\mathbb{P}(\Delta_{i,j}=1|Y_{i,j}=0)=p_0$.

To prove (3.8), we write

$$
\begin{aligned}
p_1 &= \mathbb{P}(\Delta_{i,j}=1,Y_{i,j}=1)/\mathbb{P}(Y_{i,j}=1)\\
&= \big\{\mathbb{P}(\Delta_{i,j}=1)-\mathbb{P}(\Delta_{i,j}=1|Y_{i,j}=0)\mathbb{P}(Y_{i,j}=0)\big\}/(1-q_0)\\
&= (\mu_\Delta-p_0q_0)/(1-q_0)\,,
\end{aligned}
$$

where $\mu_\Delta=\mathbb{E}(\Delta_{i,j})$.

A.3 Discussion about the assumptions from section 4.1

Conditions (A1) to (A3) and (A6) to (A8) are standard in penalised spline estimation. Intuitively, we need p to be smooth enough to be able use a piecewise polynomial of degree d locally around x. (A3) is a natural assumption since it is unreasonable to impose a penalty on a $d+1$ or higher order derivative of a spline with degree only d. This assumption is also easy to fulfill as d and ℓ are chosen by the user, and usually in practice we take $d=3$ and $\ell=2$ (see Ruppert et al., 2003). (A4) requires that $\mathbb{P}(\Delta=1|Y=1)$ and $\mathbb{P}(\Delta=1|Y=0)$

be strictly larger than 0, which is reasonable as in practice, X can usually be observed for at least some individuals for which $Y = 0$ or $Y = 1$. (A5) is needed to derive the asymptotic properties of our estimator. It means that $\mathbb{P}(\Delta = 1|X = x) \neq 0$ or 1 for any $x \in [a, b]$, which is often satisfied in practice. Indeed, usually, all values of X can potentially be observed or missing. (A6) and (A7) are also easy to satisfy since we can choose K and the location of the knots. (A7) ensures that the knots are placed almost uniformly. It was used by Barrow and Smith (1978) and is commonly used in the spline literature (see e.g. Zhou et al., 1998 and Claeskens et al., 2009). (A8) guarantees that we have enough data on $[a, b]$ to perform the estimation. (A9) and (A10) are natural conditions in group testing studies (see e.g. Delaigle and Meister, 2011 and Delaigle et al., 2014).

Note that our work is in the random design setting (see Assumption (A8)). In this setting, under Assumptions (A1) and (A7), Zhou et al. (1998) have derived the asymptotic bias and variance expressions for the unpenalised least squares spline estimator for complete ungrouped i.i.d. data. Their proof requires that $K = o(n^r)$ for some $r \in (0, 1/2]$, which is stronger than our Assumption (A6). Also in the complete ungrouped case and for the unpenalised least squares spline estimator, Huang (2003) used the condition $K \log K/n \to 0$ as $n \to \infty$ to derive their asymptotic bias and variance conditional on the $X_{i,j}$'s. Our assumption $K(\log K)^2/n \to 0$ as $n \to \infty$ is slightly stronger than theirs; it enables us to derive unconditional bias and variance expressions for the penalised spline estimator.

Assumption (A11) is needed for our results and is also easy to fulfill since we can choose λ. In order to investigate the general theoretical behavior of their penalised spline estimator, Claeskens et al. (2009) derived the asymptotic bias and variance for the penalised spline estimator calculated from complete ungrouped i.i.d. data in the fixed design setting, without any condition on λ. However, it is not clear that, without such an assumption, their Lemma A1 is valid; see our discussion below Corollary D.15. Our assumption $\lambda = o(K)$ is not required to establish consistency but enables us to derive the dominating part of the bias expression of our estimator.

Assumption (A12) is also made by Delaigle et al. (2014). It guarantees that the weight function φ_j, which may depend on n, is uniformly bounded and smooth.

A.4 Discussion about Theorem 1

In the non-grouped i.i.d. case without missing data considered by Zhou et al. (1998) and Claeskens et al. (2009), the asymptotic bias and variance of our spline estimator $\widehat{p}_s(x)$ reduce to B_δ and $n^{-1}\boldsymbol{N}^\top(x)\boldsymbol{H}_\lambda^{-1}\boldsymbol{G}_g\boldsymbol{H}_\lambda^{-1}\boldsymbol{N}(x) + o_p\{(n\delta)^{-1}\}$, respectively, where $\boldsymbol{G}_p = \int_a^b [p(y)\{1-p(y)\}]\boldsymbol{N}(y)\boldsymbol{N}^\top(y)f_X(y)\,dy$ and $\boldsymbol{H}_\lambda^{-1} = \lambda n^{-1}\boldsymbol{D}_\ell + \boldsymbol{G}_p$. These expressions cannot be directly compared with those of Zhou et al. (1998), who use a non penalised type of splines, nor with those of Claeskens et al. (2009), who consider only the fixed design case. However, our bias term coincides with theirs and our variance term has the same order as theirs.

The bias term $B_\delta(x)\xi(x)$ in Theorem 1 can vanish at some x. Indeed, $B_\delta(x)$ is a multiple of one of the $B_{d+1}\{(x-t_j)/(t_{j+1}-t_j)\}$'s, with B_d as in (4.4), whereas $(x-t_j)/(t_{j+1}-t_j)$ diverges and oscillates in $[0,1)$ as n increases (this is because the number of knots $K \uparrow$ as $n \uparrow$, and their location changes with n). Since B_{d+1} has some zeros on $[0,1]$, then B_δ does on $[a,b]$.

A.5 Calculations for the results in section 4.2

Using (4.3) to (4.6) and recalling the definition of ξ from Theorem 1, we write $\int_a^b \xi^2(x)V_{n,\delta}(x)\,dx = \left(p_0p_1\sum_{j=1}^J n_j\varphi_j\right)^{-2}\sum_{j=1}^J n_j\varphi_j^2 V_j$, where $V_j = \int_a^b \pi^4(x)\boldsymbol{N}^\top(x)\tilde{\boldsymbol{H}}_{n,\lambda}^{-1}\tilde{\boldsymbol{G}}_{g,j}\tilde{\boldsymbol{H}}_{n,\lambda}^{-1}\boldsymbol{N}(x)\,dx$, $\tilde{\boldsymbol{H}}_{n,\lambda} = \lambda\boldsymbol{D}_\ell/n + \int_a^b \boldsymbol{N}(y)\boldsymbol{N}^\top(y)\pi(y)f_X(y)\,dy$ and $\tilde{\boldsymbol{G}}_{g,j} = \int_a^b g(y)\{q_0^{1-n_j} - g(y)\}\boldsymbol{N}(y)\boldsymbol{N}^\top(y)\pi(y)f_X(y)\,dy$. To minimise $\int_a^b \xi^2(x)V_{n,\delta}(x)\,dx$

w.r.t. φ_j, we solve, for each j, $\partial \int_a^b \xi^2(x) V_{n,\delta}(x)\, dx / \partial \varphi_j = 0$, which gives

$$\varphi_j V_j = \sum_{k=1}^{J} n_k \varphi_k^2 V_k \Big/ \Big(\sum_{k=1}^{J} n_k \varphi_k \Big) .$$

The right hand side of the above equation is a constant of j. Since we rescale the φ_j's by $\sum_j n_k \varphi_k$ (see (3.13)), they are invariant to scale and we take that constant equal to 1. Thus we can take our φ_j's equal to $\varphi_j^*(q_0) = 1/V_j$.

A.6 Calculations for section 5.2

Recall that $\mathcal{X}_j = \{X_{i,j}, i = 1, \dots, n_j\}$ and $\mathit{\Delta}_j = \{\Delta_{i,j}, i = 1, \dots, n_j\}$. For $j = 1, \dots, J$, since $Y_j^* = 0 \iff Y_{1,j} = \dots = Y_{n_0,j} = 0$, using (2.2) and p_0 defined at page 8 in Delaigle et al. (2017), we have

$$\mathbb{P}(Y_j^* = 0, \mathit{\Delta}_j = \delta_j | \mathcal{X}_j; \boldsymbol{\theta}) = \prod_{i=1}^{n_j} \mathbb{P}(Y_{i,j} = 0, \Delta_{i,j} = \delta_{i,j} | X_{i,j}; \boldsymbol{\theta}) = \prod_{i=1}^{n_j} A_{ij}(\delta_{i,j}, X_{i,j}) \, ,$$

where $A_{ij}(\delta_{i,j}, X_{i,j}) = \{1 - p_{\boldsymbol{\theta}}(X_{i,j})\}\{p_0 \delta_{i,j} + (1-p_0)(1-\delta_{i,j})\}$ and $\delta_j = \{\delta_{i,j}, i = 1, \dots, n_j\}$. Using (3.6) and p_1 defined at page 8 in Delaigle et al. (2017), we deduce that

$$f(\mathcal{Y}, \mathcal{X}, \mathit{\Delta}; \boldsymbol{\theta}) = \prod_{j=1}^{J} \Big\{ (1 - Y_j^*) \prod_{i=1}^{n_j} B_{ij} + Y_j^* \prod_{i=1}^{n_j} (C_{ij} + D_{ij} - B_{ij}) \Big\} \, , \tag{A.2}$$

where $B_{ij} = A_{ij}(\Delta_{i,j}, X_{i,j}) f_X(X_{i,j})$,

$$C_{ij} = \Delta_{i,j} [p_0 \{1 - p_{\boldsymbol{\theta}}(X_{i,j})\} + p_1 p_{\boldsymbol{\theta}}(X_{i,j})] f_X(X_{i,j}) \, ,$$
$$D_{ij} = (1 - \Delta_{i,j}) [(1 - p_0)\{1 - p_{\boldsymbol{\theta}}(X_{i,j})\} + (1 - p_1) p_{\boldsymbol{\theta}}(X_{i,j})] f_X(X_{i,j}) \, .$$

Next, following (5.3), we integrate (A.2) w.r.t. the missing $X_{i,j}$'s, which gives

$$\mathcal{L}(\boldsymbol{\theta} | \mathcal{Y}, \mathit{\Delta}, \mathcal{X}_{\text{obs}}) = \prod_{j=1}^{J} \Big[(1 - Y_j^*) f_{0,\boldsymbol{\theta}}(\mathcal{X}_{j,\text{obs}}, \mathit{\Delta}_j) + Y_j^* \{ f_{X\Delta,\boldsymbol{\theta}}(\mathcal{X}_{j,\text{obs}}, \mathit{\Delta}_j) - f_{0,\boldsymbol{\theta}}(\mathcal{X}_{j,\text{obs}}, \mathit{\Delta}_j) \} \Big] \, ,$$

where $\mathcal{X}_{j,\text{obs}}$ denotes the observed $X_{i,j}$'s for group j, and where we used the notation

$$\begin{aligned}
f_{0,\boldsymbol{\theta}}(\mathcal{X}_{j,\text{obs}}, \Delta_j) = &\prod_{i=1}^{n_j} \Bigg[\Delta_{i,j} p_0\{1 - p_{\boldsymbol{\theta}}(X_{i,j})\} f_X(X_{i,j}) \\
&+ (1-\Delta_{i,j})(1-p_0)\int_{-\infty}^{\infty}\{1-p_{\boldsymbol{\theta}}(x)\} f_X(x)\,dx\Bigg], \\
f_{X\Delta,\boldsymbol{\theta}}(\mathcal{X}_{j,\text{obs}}, \Delta_j) = &\prod_{i=1}^{n_j}\Bigg(\Delta_{i,j}[p_0\{1-p_{\boldsymbol{\theta}}(X_{i,j})\} + p_1 p_{\boldsymbol{\theta}}(X_{i,j})] f_X(X_{i,j}) \\
&+ (1-\Delta_{i,j})\int_{-\infty}^{\infty}(1-p_0)\{1-p_{\boldsymbol{\theta}}(x)\} f_X(x) + (1-p_1)p_{\boldsymbol{\theta}}(x) f_X(x)\,dx\Bigg).
\end{aligned}$$

A.7 Discussion about the assumptions from section 5.2

The assumptions on the parametric model in Assumption **P** are standard in the parametric literature, and the other smoothness assumptions are similar to those discussed in Appendix A.3, except perhaps for Assumption (P9). However, using Assumptions (P3), (P5) and (P6), it is easy to check that Assumption (P9) holds for example when $f_X\pi$ is compactly supported or in many cases where f_X belongs to the exponential family (e.g. normal, exponential, gamma and chi-squared distributions). Indeed, using Assumptions (P5) and (P6), we have $\int_{\mathbb{R}} \sqrt{F(x-c_x)}\,dx \leq \int_{\mathbb{R}} \sqrt{f_X(x-c_x)}\,dx$. If f_X is unimodal so that it increases (resp., decreases) on the left (resp., right) hand side of its mode M (which is satisfied by many densities in the exponential family), we have that for all $x \in (M+1,\infty)$ and $c_x \in [-1,1]$, $f_X(x-c_x) \leq f_X(x-1)$ and for all $x \in (-\infty, M-1]$ and $c_x \in [-1,1]$, $f_X(x-c_x) \leq f_X(x+1)$. Therefore, if $\sqrt{f_X}$ is integrable, which is often the case in the exponential family, and using Assumption (P3), we have $\int_{\mathbb{R}} \sqrt{f_X(x-c_x)}\,dx \leq \int_{M-1}^{M+1} \sqrt{f_X(x-c_x)}\,dx + \int_{M+1}^{\infty} \sqrt{f_X(x-1)}\,dx + \int_{-\infty}^{M-1} \sqrt{f_X(x+1)}\,dx \leq 2\|\sqrt{f_X}\|_{L_\infty} + 2\int_{\mathbb{R}} \sqrt{f_X(x)}\,dx \leq 2\sqrt{f_{\max}} + 2\int_{\mathbb{R}} \sqrt{f_X(x)}\,dx < \infty$.

A.8 Calculations for section 5.3

Extending our arguments from section 4.2, the φ_j's in ϕ and Φ can be computed in practice by taking $\widehat{\varphi}_j(\widehat{q}_0) = 1/\int_A \boldsymbol{B}^\top(\boldsymbol{x})\widehat{\boldsymbol{H}}^{-1}\widehat{\boldsymbol{G}}_j\widehat{\boldsymbol{H}}^{-1}\boldsymbol{B}(\boldsymbol{x})\,d\boldsymbol{x}$, where $\widehat{\boldsymbol{H}} = \lambda\boldsymbol{\mathcal{D}}/n + \int_A \boldsymbol{B}(\boldsymbol{x})\boldsymbol{B}^\top(\boldsymbol{x})\widehat{f}_{\boldsymbol{X}obs}(\boldsymbol{x})\,d\boldsymbol{x}$ and $\widehat{\boldsymbol{G}}_j = \int_A \widehat{g}_{\text{pilot}}(\boldsymbol{x})\{\widehat{q}_0^{1-n_j} - \widehat{g}_{\text{pilot}}(\boldsymbol{x})\}\boldsymbol{B}(\boldsymbol{x})\boldsymbol{B}^\top(\boldsymbol{x})\widehat{f}_{\boldsymbol{X}obs}(\boldsymbol{x})\,d\boldsymbol{x}$, where $\widehat{g}_{\text{pilot}}$ is defined as at (5.18) but with the φ_j's all equal to 1, and $\widehat{f}_{\boldsymbol{X}obs}(\boldsymbol{x})$ denotes the standard kernel density estimator computed from the $\boldsymbol{X}_{i,j}$'s for which $\Delta_{i,j} = 1$.

B Additional simulation results

We also performed simulations where we grouped the data in groups of unequal sizes n_j. Specifically we generated data from models (i), (ii) and (iii) under missing mechanism (1), and grouped them by creating randomly $[n/4]$ groups of size $n_j = 2$ and $[n/10]$ groups of size $n_j = 5$, which refer to as grouping 1, and $[n/8]$ groups of size $n_j = 2$, $[n/16]$ groups of size $n_j = 4$, $[n/20]$ groups of size $n_j = 5$ and $[n/40]$ groups of size $n_j = 10$, which we refer to as grouping 2, with $n = 2000$ or 5000. We know from section 4.2 that when the group sizes are unequal, the groups should be given different weights. To illustrate the importance of assigning weights according to the n_j's, we compared our weighted spline estimator $\widehat{p}_s$, with weights φ_j computed as in section 4.2, with the unweighted version of our spline estimator, which we denote by $\widehat{p}_{\text{unw},s}$ and which is defined in the same way as $\widehat{p}_s$, except that we take there all φ_j's equal to 1. Again to illustrate the loss incurred by the fact of having missing data, before removing randomly some of the $X_{i,j}$'s from our data, we also computed our weighted and unweighted spline estimators applied to the full data. We denote these two estimators by $\widehat{p}_{\text{oracle},s}$ and $\widehat{p}_{\text{unw,oracle},s}$, respectively. As in section 6.2, they are obtained by setting $\widehat{p}_0 = \widehat{p}_1 \equiv 1$ and all $\Delta_{i,j}$'s equal to 1.

Table B.1 shows, for each model and each setting, 10^4 times the mean and the standard

Table B.1: $10^4\times$ Mean (Standard deviation) of 200 ISE values for four estimators of p, obtained from 200 samples simulated from models (i), (ii) and (iii) in the case where the groups are of unequal sizes, for two types of grouping, under missing mechanism (1).

		Grouping 1			
Model	n	$\widehat{p}_{\text{unw},s}$	$\widehat{p}_s$	$\widehat{p}_{\text{unw,oracle},s}$	$\widehat{p}_{\text{oracle},s}$
(i)	$2\cdot 10^3$	11.82(9.66)	10.22(9.13)	11.96(10.98)	9.06(8.44)
	$5\cdot 10^3$	5.54(5.13)	4.43(3.93)	5.26(4.71)	4.39(3.83)
(ii)	$2\cdot 10^3$	17.86(14.37)	14.62(12.24)	17.08(14.34)	13.78(11.53)
	$5\cdot 10^3$	7.50(6.15)	5.96(4.98)	7.05(5.18)	5.74(4.28)
(iii)	$2\cdot 10^3$	16.82(9.44)	15.38(8.18)	16.86(8.51)	15.24(7.36)
	$5\cdot 10^3$	7.92(4.13)	7.05(3.73)	8.05(4.37)	6.90(3.76)
		Grouping 2			
Model	n	$\widehat{p}_{\text{unw},s}$	$\widehat{p}_s$	$\widehat{p}_{\text{unw,oracle},s}$	$\widehat{p}_{\text{oracle},s}$
(i)	$2\cdot 10^3$	20.12(17.86)	14.11(12.80)	21.82(23.29)	13.61(12.38)
	$5\cdot 10^3$	9.27(8.63)	7.04(6.77)	10.21(7.99)	7.32(5.98)
(ii)	$2\cdot 10^3$	41.57(44.44)	18.57(16.93)	45.85(53.33)	17.11(14.27)
	$5\cdot 10^3$	15.86(13.40)	7.60(6.28)	17.38(15.58)	6.87(5.66)
(iii)	$2\cdot 10^3$	19.02(10.77)	16.29(8.97)	19.34(9.45)	16.43(7.94)
	$5\cdot 10^3$	10.43(5.99)	8.67(4.68)	10.98(5.62)	9.20(5.00)

deviation of the ISE computed from 200 simulated samples. Figures B.1 and B.2 depict the quartile curves for models (iii) and (ii), respectively. These results illustrate the advantage of using a weighted version of our estimator when the group sizes are unequal.

Finally, Table B.2 compares our parametric estimator $p_{\widehat{\theta}}$, our local linear estimator $\widehat{p}_{\text{LL}}$, our spline estimator $\widehat{p}_s$ and the naive spline estimator $\widehat{p}_{\text{naive},s}$ which does not make any correction for missingness, when the missing rate increases. Specifically we compare the results for missing mechanisms (1) an (2) and for models (i) to (iii), when $n_j = 4$ and $n = 2000$ or 5000. Unsurprisingly, we see that as the missing rate increases, the quality of all four estimators degrade. However, the estimator that is the most affected by increasing the missing rate is the naive estimator. Indeed, the reason this estimator is not consistent is because, when there are missing data, its bias does not tend to zero as n increases.

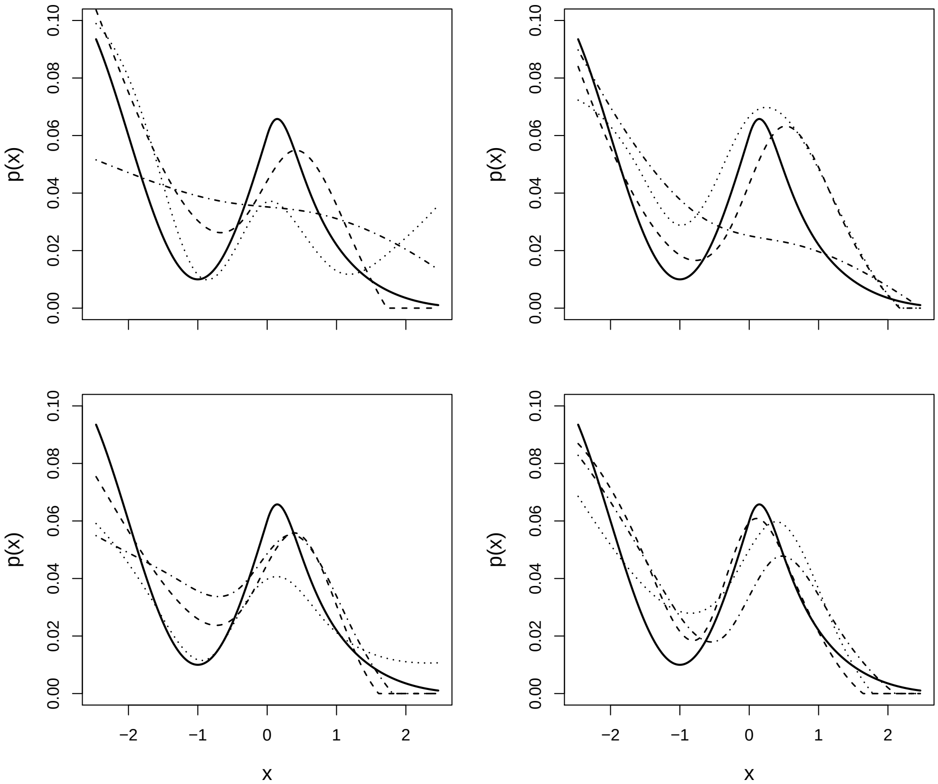

Figure B.1: True curve (—) and the 1st (- - -), 2nd ($\cdots$) and 3rd ($- \cdot -\cdot$) quartile curves for $\widehat{p}_{\mathrm{unw},s}$ (left) and $\widehat{p}_s$ (right), obtained from 200 samples coming from model (iii) when $[n/4]$ groups are of size $n_j = 2$ and $[n/10]$ groups of size $n_j = 5$, $n = 2000$ (row 1) and $n = 5000$ (row 2).

When the missing rate increases, the bias increases because of the non consistency, and the variance increases because the effective sample size reduces. By contrast, our estimators are consistent and so when the missing rate increases, the degradation in performance is more limited.

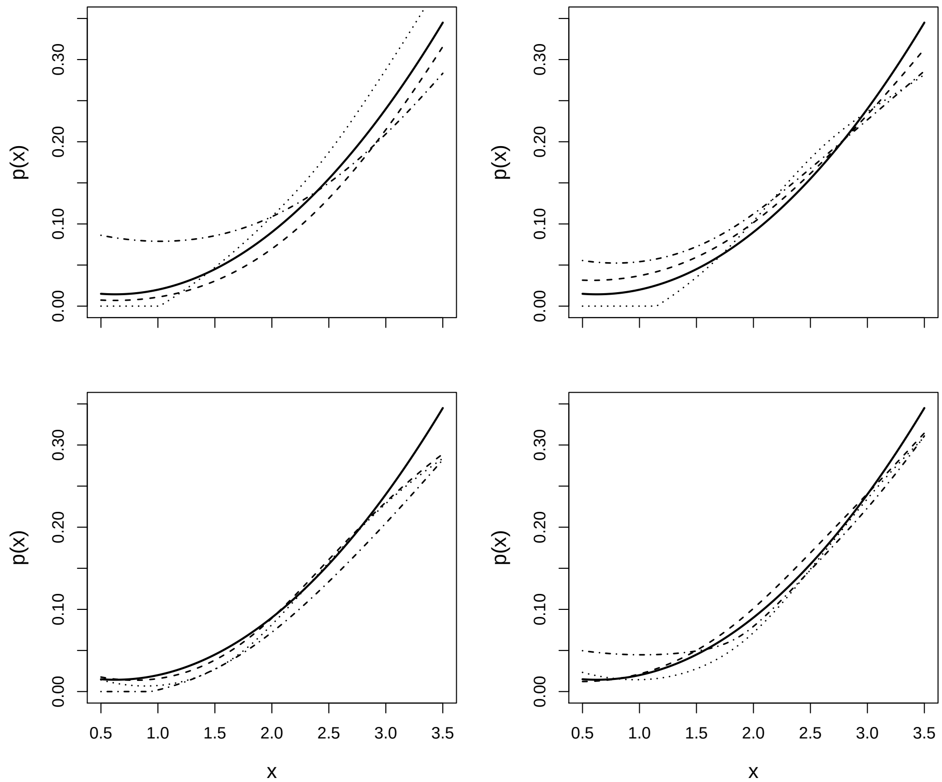

Figure B.2: True curve (—) and the 1st (- - -), 2nd ($\cdots$) and 3rd ($-\cdot-\cdot$) quartile curves for $\widehat{p}_{\text{unw},s}$ (left) and $\widehat{p}_s$ (right), obtained from 200 samples coming from model (ii) when $[n/8]$ groups are of size $n_j = 2$, $[n/16]$ groups are of size $n_j = 4$, $[n/20]$ groups are of size $n_j = 5$ and $[n/40]$ groups of size $n_j = 10$, $n = 2000$ (row 1) and $n = 5000$ (row 2).

C Proof of Theorem 1

In this section, we derive the proof of Theorem 1. Recall from (3.15) that $\widehat{p}_s(x) = \{1 - \widehat{g}_s(x)\}/\{1 + (\widehat{p}_1\widehat{p}_0^{-1} - 1)\widehat{g}_s(x)\}$, which involves three estimators: $\widehat{p}_0, \widehat{p}_1$ and $\widehat{g}_s$. To establish our result, we derive properties of $\widehat{g}_s, \widehat{p}_0, \widehat{p}_1$ and of $\widehat{g}_s^0$, where $\widehat{g}_s^0(x)$ denotes the version of $\hat{g}_s(x)$ with $\widehat{q}_0$ replaced by q_0, i.e.

$$\widehat{g}_s^0(x) = \underset{s(x)\in S_d(\boldsymbol{t})}{\operatorname{argmin}} \left[\phi(q_0) \sum_{j=1}^{J} \sum_{i=1}^{n_j} \{q_0^{1-n_j} Z_j^* - s(X_{i,j})\}^2 \varphi_j(q_0)\Delta_{i,j} + \lambda \int_a^b \{s^{(\ell)}(x)\}^2\, dx \right]. \quad \text{(C.1)}$$

Table B.2: $10^4\times$ Mean (Standard deviation) of 200 ISE values for four estimators of p, obtained from 200 samples simulated from models (i) to (iii) and missing mechanisms (1) and (2), when the group sizes are equal to $n_j = 4$ and when $n = 2000$ or 5000.

		$n = 2000$			
Model	Missing	$p_{\widehat{\theta}}$	$\widehat{p}_{\mathrm{LL}}$	$\widehat{p}_s$	$\widehat{p}_{\mathrm{naive},s}$
(i)	(1)	9.42(6.86)	18.62(19.24)	12.71(11.44)	36.85(24.61)
	(2)	11.39(8.99)	22.48(22.91)	15.86(12.68)	86.71(62.72)
(ii)	(1)	10.33(9.65)	25.87(22.69)	16.78(16.12)	53.72(38.24)
	(2)	14.21(16.07)	39.33(43.08)	26.74(31.31)	124.59(87.62)
(iii)	(1)	1.95(1.54)	17.76(11.42)	16.60(8.95)	24.82(13.65)
	(2)	2.29(1.77)	22.63(16.99)	19.03(11.55)	42.35(27.44)
		$n = 5000$			
Model	Missing	$p_{\widehat{\theta}}$	$\widehat{p}_{\mathrm{LL}}$	$\widehat{p}_s$	$\widehat{p}_{\mathrm{naive},s}$
(i)	(1)	5.91(4.64)	9.61(7.59)	6.27(5.52)	26.46(15.87)
	(2)	7.80(6.16)	12.13(10.49)	8.61(8.83)	66.44(34.91)
(ii)	(1)	5.62(4.51)	11.32(8.78)	7.26(6.75)	37.45(29.41)
	(2)	7.07(5.14)	15.41(13.13)	9.88(8.97)	99.21(42.81)
(iii)	(1)	1.19(0.90)	7.93(4.93)	8.23(4.75)	14.80(8.41)
	(2)	1.57(1.09)	10.27(6.11)	10.78(5.99)	30.04(17.23)

A number of results concerning these estimators will be provided in sections C.2 and C.3. Since the proof for $\widehat{g}_s^0$ is very technical, many auxiliary results are relegated to section D.

Throughout sections C and D, whenever we write an inequality for a random variable without any further specification, we mean that the inequality holds almost surely. We write "almost surely" explicitly in cases where it is useful to emphasize it.

Moreover, in addition to the notations introduced at page 20 of Delaigle et al. (2017), throughout we use the notation $\|\boldsymbol{v}\|_q = \Big(\sum_i |v_i|^q\Big)^{1/q}$ for $1 \le q < \infty$, $\|\boldsymbol{v}\|_\infty = \max_i |v_i|$ for any column vector $\boldsymbol{v}$, $\|\boldsymbol{u}\|_\infty = \sum_i |u_i|$ for any row vector $\boldsymbol{u}$, $\|\boldsymbol{A}\|_\infty = \max_i \sum_j |(\boldsymbol{A})_{ij}|$ for any matrix $\boldsymbol{A}$.

C.1 Proof of Theorem 1

Proof. Recall that $\widehat{p}_s(x) = \{1-\widehat{g}_s(x)\}/\{1+(\widehat{p}_1\widehat{p}_0^{-1}-1)\widehat{g}_s(x)\}$. Applying Taylor's expansion of first order in the Peano form to the function $1/\{1+(\widehat{p}_1\widehat{p}_0^{-1}-1)x\}$ around $x=a$, there exists a function h such that

$$\begin{aligned}\{1+(\widehat{p}_1\widehat{p}_0^{-1}-1)x\}^{-1} &= \{1+(\widehat{p}_1\widehat{p}_0^{-1}-1)a\}^{-1} \\ &\quad -\{1+(\widehat{p}_1\widehat{p}_0^{-1}-1)a\}^{-2}(\widehat{p}_1\widehat{p}_0^{-1}-1)(x-a) \\ &\quad +h(x)(x-a)\,,\end{aligned}$$

$\lim_{x\to a} h(x) = 0$, and $h(x)(x-a) = o(|x-a|)$ as $x \to a$. Note that $h(x)(x-a) = 0$ if $x-a=0$, and using Theorem C.1 at page 18 and Lemma C.5 at page 24, we have $\widehat{g}_s(x) - g(x) = o_p(1)$.

Then applying the above calculation and Lemma 2.12 of Van der Vaart (2000) to $\{1+(\widehat{p}_1\widehat{p}_0^{-1}-1)\widehat{g}_s(x)\}^{-1}$, we have

$$\begin{aligned}\{1+(\widehat{p}_1\widehat{p}_0^{-1}-1)\widehat{g}_s(x)\}^{-1} &= \{1+(\widehat{p}_1\widehat{p}_0^{-1}-1)g(x)\}^{-1} \\ &\quad -\{1+(\widehat{p}_1\widehat{p}_0^{-1}-1)g(x)\}^{-2}(\widehat{p}_1\widehat{p}_0^{-1}-1)\{\widehat{g}_s(x)-g(x)\} \\ &\quad +o_p\{(\widehat{g}_s(x)-g(x)\}\,.\end{aligned}$$

Let $a_n = (\widehat{p}_1\widehat{p}_0^{-1}-1)g(x) - (p_1p_0^{-1}-1)g(x)$, and note that using Lemma C.3 at page 16 and Lemma C.4 at page 17, $a_n = O_p(n^{-1/2})$. Then using the same technique as above, for $i=1,2$,

$$\begin{aligned}\{1+(\widehat{p}_1\widehat{p}_0^{-1}-1)g(x)\}^{-i} &= \{1+(p_1p_0^{-1}-1)g(x)+a_n\}^{-i} \\ &= \{1+(p_1p_0^{-1}-1)g(x)\}^{-i} - i\{1+(p_1p_0^{-1}-1)g(x)\}^{-i-1}a_n \\ &\quad +o_p(a_n) \\ &= \{1+(p_1p_0^{-1}-1)g(x)\}^{-i} + O_p(n^{-1/2})\,.\end{aligned}$$

Thus,

$$
\begin{aligned}
&\{1+(\widehat{p}_1\widehat{p}_0^{-1}-1)\widehat{g}_s(x)\}^{-1}\\
&=\{1+(p_1p_0^{-1}-1)g(x)\}^{-1}+O_p(n^{-1/2})\\
&-[\{1+(p_1p_0^{-1}-1)g(x)\}^{-2}+O_p(n^{-1/2})]\{p_1p_0^{-1}-1+O_p(n^{-1/2})\}\{\widehat{g}_s(x)-g(x)\}\\
&+o_p\{(\widehat{g}_s(x)-g(x)\}\\
&=\{1+(p_1p_0^{-1}-1)g(x)\}^{-1}-\{1+(p_1p_0^{-1}-1)g(x)\}^{-2}(p_1p_0^{-1}-1)\{\widehat{g}_s(x)-g(x)\}\\
&+o_p\{(\widehat{g}_s(x)-g(x)\}+O_p(n^{-1/2})\,.
\end{aligned}
$$

Then we have

$$
\begin{aligned}
\widehat{p}_s(x) &= \{1-g(x)+g(x)-\widehat{g}_s(x)\}\big[\{1+(p_1p_0^{-1}-1)g(x)\}^{-1}\\
&\quad-\{1+(p_1p_0^{-1}-1)g(x)\}^{-2}(p_1p_0^{-1}-1)\{\widehat{g}_s(x)-g(x)\}\\
&\quad+o_p\{\widehat{g}_s(x)-g(x)\}+O_p(n^{-1/2})\big]\\
&= p(x)+\{g(x)-\widehat{g}_s(x)\}\big[\{1+(p_1p_0^{-1}-1)g(x)\}^{-1}\\
&\quad+\{1+(p_1p_0^{-1}-1)g(x)\}^{-2}(p_1p_0^{-1}-1)\{1-g(x)\}\big]+o_p\{\widehat{g}_s(x)-g(x)\}\\
&\quad+O_p(n^{-1/2})\\
&= p(x)+\{g(x)-\widehat{g}_s(x)\}\big[p_1p_0^{-1}\{1+(p_1p_0^{-1}-1)g(x)\}^{-2}\big]+o_p\{\widehat{g}_s(x)-g(x)\}\\
&\quad+O_p(n^{-1/2})\,.
\end{aligned}
$$

Let us examine $\{1+(p_1p_0^{-1}-1)g(x)\}^{-2}$. Since $p(x)=\{1-g(x)\}/\{1+(p_1p_0^{-1}-1)g(x)\}$, we have $g(x)=\{1-p(x)\}/\{1+(p_1p_0^{-1}-1)p(x)\}$. Then

$$1+(p_1p_0^{-1}-1)g(x)=1+\frac{(p_1p_0^{-1}-1)\{1-p(x)\}}{1+(p_1p_0^{-1}-1)p(x)}=\frac{p_1p_0^{-1}}{1+(p_1p_0^{-1}-1)p(x)}\,.$$

Thus,

$$\frac{1}{1+(p_1p_0^{-1}-1)g(x)}=\frac{p_0+(p_1-p_0)p(x)}{p_1}\,.$$

Recalling from (3.6) that $\pi(x) = p_0 + (p_1 - p_0)p(x)$, we have $\{1 + (p_1 p_0^{-1} - 1)g(x)\}^{-2} = \pi^2(x)/p_1^2$. Therefore, we have

$$\begin{aligned}\widehat{p}_s(x) - p(x) &= \{g(x) - \widehat{g}_s(x)\}\pi^2(x)p_0^{-1}p_1^{-1} + o_p\{\widehat{g}_s(x) - g(x)\} + O_p(n^{-1/2}) \\ &= \big[\{g(x) - \widehat{g}_s^0(x)\} + \{\widehat{g}_s^0(x) - \widehat{g}_s(x)\}\big]\pi^2(x)p_0^{-1}p_1^{-1} \\ &\quad + o_p\{\widehat{g}_s(x) - \widehat{g}_s^0(x)\} + o_p\{\widehat{g}_s^0(x) - g(x)\} + O_p(n^{-1/2}),\end{aligned}$$

with $\widehat{g}_s^0$ defined at (C.1).

In Theorem C.1 at page 18, we prove that $g(x) - \widehat{g}_s^0(x) = AE(x) + o_p(\delta^{d+1}) + o_p(n^{-1/2}\delta^{-1/2})$, where

$$AE(x) = B_\delta(x) + \frac{\phi}{n\delta}\boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{N}^\top\boldsymbol{\Phi}\boldsymbol{\Delta}(\underline{g} - \boldsymbol{Q}\boldsymbol{Z}^*)$$

and the expectation and the variance of $AE(x)$ are derived there. In Lemma C.5 at page 24, we prove that $\widehat{g}_s(x) = \widehat{g}_s^0(x) + o_p(n^{-1/2}\delta^{-1/2})$. Moreover, using Assumptions (A6) that $K \to \infty$ as $n \to \infty$ and (A7) that $\delta \asymp K^{-1}$, we have that $\delta \to 0$ as $n \to \infty$. Thus, $O_p(n^{-1/2}) = o_p(n^{-1/2}\delta^{-1/2})$, which proves that

$$\widehat{p}_s(x) - p(x) = AE(x)\xi(x) + o_p\big(\delta^{d+1}\big) + o_p\big\{(n\delta)^{-1/2}\big\}. \tag{C.2}$$

Next we prove asymptotic normality. It follows from (C.2) and the definition of $AE(x)$ that

$$\widehat{p}_s(x) - p(x) = B_\delta(x)\xi(x) + T_{n,\delta}(x)\xi(x) + o_p\big(\delta^{d+1}\big) + o_p\big\{(n\delta)^{-1/2}\big\},$$

where $T_{n,\delta}(x) = (n\delta)^{-1}\phi\boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{N}^\top\boldsymbol{\Phi}\boldsymbol{\Delta}(\underline{g} - \boldsymbol{Q}\boldsymbol{Z}^*)$. In Lemma D.22, we prove that $T_{n,\delta}(x)/V_{n,\delta}^{1/2}(x) \xrightarrow{D} N(0,1)$ as $n \to \infty$. This proves (4.1), where $\Psi_{n,\delta}(x) = T_{n,\delta}(x)/V_{n,\delta}^{1/2}(x)$.

□

C.2 Consistency of the estimators of $\widehat{q}_0, \bar{\Delta}, \widehat{p}_0$ and $\widehat{p}_1$

The following lemmas establish the convergence rates of $\widehat{q}_0, \bar{\Delta}, \widehat{p}_0$ and $\widehat{p}_1$.

Lemma C.1. *Under Assumption (A9), $\widehat{q}_0 - q_0 = O_p(n^{-1/2})$, where $\widehat{q}_0$ is defined in Appendix A.*

Proof. The second last line of the proof of Lemma A.2. of Delaigle and Meister (2011) implies that for any sequence $\{\epsilon_n\}_n \to 0$ as $n \to \infty$,

$$\mathbb{P}(|\widehat{q}_0 - q_0| > \epsilon_n) \leq \mathcal{M} n^{-1} \epsilon_n^{-2}\{1 + o(1)\},$$

where $\mathcal{M}$ is a finite positive constant that dose not depend on n and the $o(1)$ goes to zero as n goes to infinity.

Thus, letting $\epsilon_n = Cn^{-1/2}$ for some positive constant C that does not depend on n, we have

$$\lim_{C\to\infty} \limsup_{n\to\infty} \mathbb{P}(|\widehat{q}_0 - q_0| > Cn^{-1/2}) = \lim_{C\to\infty} C^{-2}\mathcal{M} = 0.$$

Thus, $\widehat{q}_0 - q_0 = O_p(n^{-1/2})$. □

Lemma C.2. *Under Assumption (A9), $\bar{\Delta} - \mu_\Delta = O_p(n^{-1/2})$, where μ_Δ and $\bar{\Delta}$ are defined at page 8 in Delaigle et al. (2017).*

Proof. Recalling from page 8 in Delaigle et al. (2017) that $\bar{\Delta} = n^{-1}\sum_{j=1}^{J}\sum_{i=1}^{n_j} \Delta_{i,j}$, we have $\mathbb{E}(\bar{\Delta}) = n^{-1}\sum_{j=1}^{J}\sum_{i=1}^{n_j} \mu_\Delta$ where $\mu_\Delta = \mathbb{E}(\Delta_{i,j})$. Note that

$$\begin{aligned}
\mathrm{var}(\bar{\Delta}) &= \mathrm{var}\left(n^{-1}\sum_{j=1}^{J}\sum_{i=1}^{n_j}\Delta_{i,j}\right) \\
&= n^{-2}\sum_{j=1}^{J}\sum_{i=1}^{n_j}\mathrm{var}(\Delta_{i,j})
\end{aligned}$$

$$= n^{-1}\mu_\Delta(1-\mu_\Delta)$$
$$= O(n^{-1}).$$

Then we have $\bar{\Delta} - \mu_\Delta = O_p(n^{-1/2})$. □

Lemma C.3. *Under Assumptions (A4) and (A9), as long as $c_0 < c_p$, we have that $\widehat{p}_0 - p_0 = O_p(n^{-1/2})$ and $\widehat{p_0}^{-1} - p_0^{-1} = O_p(n^{-1/2})$, where c_0, p_0 and $\widehat{p}_0$ are defined at page 8 in Delaigle et al. (2017) and c_p is defined in Assumption (A4).*

Proof. Recall from page 8 in Delaigle et al. (2017) that $\widehat{p}_0 = \max(\tilde{p}_0, c_0)$ where $\tilde{p}_0 = \sum_{j=1}^J \sum_{i=1}^{n_j} \Delta_{i,j} Z_j^* / \sum_{j=1}^J \sum_{i=1}^{n_j} Z_j^*$ and $c_0 > 0$ is a small number. We first show that $\tilde{p}_0 - p_0 = O_p(n^{-1/2})$.

Letting $\tilde{p}_{\Delta Z} = \sum_{j=1}^J \sum_{i=1}^{n_j} \Delta_{i,j} Z_j^* / \sum_{j=1}^J \sum_{i=1}^{n_j} q_0^{n_j}$ and $\tilde{p}_Z = \sum_{j=1}^J \sum_{i=1}^{n_j} Z_j^* / \sum_{j=1}^J \sum_{i=1}^{n_j} q_0^{n_j}$, we have $\tilde{p}_0 = \tilde{p}_{\Delta Z}/\tilde{p}_Z$.

Using (3.1) and (3.5) in Delaigle et al. (2017), we have

$$\mathbb{E}(\tilde{p}_{\Delta Z}) = \frac{\sum_{j=1}^J \sum_{i=1}^{n_j} \mathbb{P}(\Delta_{i,j} = 1 | Z_j^* = 1)\mathbb{P}(Z_j^* = 1)}{\sum_{j=1}^J \sum_{i=1}^{n_j} q_0^{n_j}} = p_0 ,$$

and

$$\mathbb{E}(\tilde{p}_Z) = \frac{\sum_{j=1}^J \sum_{i=1}^{n_j} \mathbb{P}(Z_j^* = 1)}{\sum_{j=1}^J \sum_{i=1}^{n_j} q_0^{n_j}} = 1 .$$

Furthermore, using Assumption (A9), we have

$$\mathrm{var}(\tilde{p}_{\Delta Z}) = \frac{\sum_{j=1}^J \sum_{i=1}^{n_j} \mathrm{var}(\Delta_{i,j} Z_j^*)}{(\sum_{j=1}^J \sum_{i=1}^{n_j} q_0^{n_j})^2} = O(n^{-1}) , \mathrm{var}(\tilde{p}_Z) = \frac{\sum_{j=1}^J \sum_{i=1}^{n_j} \mathrm{var}(Z_j^*)}{(\sum_{j=1}^J \sum_{i=1}^{n_j} q_0^{n_j})^2} = O(n^{-1}) .$$

Therefore, we have $\tilde{p}_Z = 1 + O_p(n^{-1/2})$ and $\tilde{p}_{\Delta Z} = p_0 + O_p(n^{-1/2})$.

Letting $a_n = \tilde{p}_Z - 1$, we have $\tilde{p}_Z^{-1} = 1/(1 + a_n)$. Applying a Taylor expansion of zeroth order and Lemma 2.12 in Van der Vaart (2000) to $(1 + a_n)^{-1}$ around $a_n = 0$, we have $\tilde{p}_Z^{-1} = (1 + a_n)^{-1} = 1 + O_p(a_n) = 1 + O_p(n^{-1/2})$. Thus, $\tilde{p}_0 - p_0 = \tilde{p}_{\Delta Z}/\tilde{p}_Z - p_0 = \{p_0 + O_p(n^{-1/2})\}\{1 + O_p(n^{-1/2})\} - p_0 = O_p(n^{-1/2})$. Noting that $\widehat{p}_0 = \max(\tilde{p}_0, c_0)$, using Assumption (A4), since $c_0 < c_p$ we have that $|p_0 - \widehat{p}_0| \le |p_0 - \tilde{p}_0|$. Then $|p_0 - \widehat{p}_0| \le |p_0 - \tilde{p}_0| = O_p(n^{-1/2})$.

Using similar arguments, we also have $\widehat{p_0}^{-1} - p_0^{-1} = O_p(n^{-1/2})$. □

Lemma C.4. *Under Assumptions (A4) and (A9), as long as $c_0 < c_p$, we have that $\widehat{p}_1 - p_1 = O_p(n^{-1/2})$ and $\widehat{p_1}^{-1} - p_1^{-1} = O_p(n^{-1/2})$, where c_0, $\widehat{p}_1$ and p_1 are defined at page 8 in Delaigle et al. (2017) and c_p is defined in Assumption (A4).*

Proof. Recall from page 8 of Delaigle et al. (2017) that $\widehat{p}_1 = \max(\tilde{p}_1, c_0)$ where $c_0 > 0$ is a small number and $\tilde{p}_1 = (\bar{\Delta} - \widehat{p}_0\widehat{q}_0)/(1 - \widehat{q}_0)$. Using Lemmas C.1, C.2 and C.3, we have

$$\tilde{p}_1 = \frac{\bar{\Delta} - \widehat{p}_0\widehat{q}_0}{1 - \widehat{q}_0} = \frac{\mu_\Delta + O_p(n^{-1/2}) - \{p_0 + O_p(n^{-1/2})\}\{q_0 + O_p(n^{-1/2})\}}{1 - q_0 + O_p(n^{-1/2})}\,.$$

Letting $a_n = \{(1 - \widehat{q}_0) - (1 - q_0)\}/(1 - q_0)$, we have $\{(1 - q_0)(1 + a_n)\}^{-1} = (1 - \widehat{q}_0)^{-1}$ and by Lemma C.1, $a_n = O_p(n^{-1/2})$. Applying a Taylor expansion of zeroth order and Lemma 2.12 in Van der Vaart (2000) to $(1 + a_n)^{-1}$ around $a_n = 0$, we have $(1 - \widehat{q}_0)^{-1} = (1 - q_0)^{-1}\{1 + O_p(n^{-1/2})\}$. Then

$$\tilde{p}_1 = \frac{\mu_\Delta - p_0 q_0 + O_p(n^{-1/2})}{1 - q_0} + O_p(n^{-1/2}) = p_1 + O_p(n^{-1/2})\,.$$

Therefore, using Assumption (A4) and arguments similar to those used the proof of Lemma C.3, since $c_0 < c_p$, we have $\widehat{p}_1 - p_1 = O_p(n^{-1/2})$ and $\widehat{p_1}^{-1} - p_1^{-1} = O_p(n^{-1/2})$. □

C.3 Asymptotic properties of the oracle estimator $\widehat{g}_s^0$ and $\widehat{g}_s - \widehat{g}_s^0$

The following theorem describes the asymptotic behavior of the oracle version $\widehat{g}_s^0(x)$ of our estimator of g, defined at (C.1). Before we state it, we note that, similarly to the expression at (3.14) for $\widehat{g}_s$, $\widehat{g}_s^0(x)$ can be expressed as

$$\widehat{g}_s^0(x) = \phi(q_0)\boldsymbol{N}^\top(x)\{\phi(q_0)\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}(q_0)\boldsymbol{\Delta}\boldsymbol{\mathcal{N}} + \lambda\boldsymbol{D}_\ell\}^{-1}\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}(q_0)\boldsymbol{\Delta}\boldsymbol{Q}\boldsymbol{Z}^*\,, \tag{C.3}$$

where $\boldsymbol{Q} = \mathrm{diag}(q_0^{1-n_1} \overset{\times n_1}{\cdots}, \cdots, q_0^{1-n_J} \overset{\times n_J}{\cdots})$ and $\boldsymbol{\Phi}(q_0) = \mathrm{diag}\big(\varphi_1(q_0) \overset{\times n_1}{\cdots}, \cdots, \varphi_J(q_0) \overset{\times n_J}{\cdots}\big)$. This expression will be useful to prove our results.

Theorem C.1. *Under Assumption* **A***, we have, for all* $x \in [a, b]$,

$$g(x) - \widehat{g}_s^0(x) = AE(x) + o_p\big(\delta^{d+1}\big) + o_p\Big\{\sqrt{(n\delta)^{-1}}\Big\}\,,$$

where $AE(x)$ *is a random variable such that* $\mathbb{E}\{AE(x)\} = B_\delta(x)$ *and* $\mathrm{var}\{AE(x)\} = V_{n,\delta}(x) + o\{(n\delta)^{-1}\}$*, with* $B_\delta(x)$ *and* $V_{n,\delta}(x)$ *defined in* (4.2) *and* (4.3)*, respectively.*

Proof. To prove this result, note that the spline estimator $\widehat{g}_s^0$ involves several matrices whose dimension goes to infinity as n goes to infinity. By contrast, we cannot generally expand g in the same way, as it does not need to be exactly a spline function itself. To calculate the difference between them, a common approach is to use the following result, which comes from Equation (2.7) of Barrow and Smith (1978): for any function $g \in C^{d+1}([a, b])$,

$$\inf_{s \in S_d(\underline{t})} \|g(x) - s(x) - b_g(x)\|_{L_\infty[a,b]} = o(\delta^{d+1})\,, \tag{C.4}$$

where

$$b_g(x) = \sum_{j=0}^{K} \mathbb{1}_{[t_j, t_{j+1})}(x)\frac{g^{(d+1)}(t_j)}{(d+1)!}\delta_j^{d+1}B_{d+1}\left(\frac{x - t_j}{t_{j+1} - t_j}\right),$$

with B_{d+1} as in (4.4).

Note that under Assumption (A1), in our case, $g \in C^{d+1}([a, b])$. Therefore, there exists a function $s_g \in S_d(\underline{t})$ such that

$$\|g(x) - s_g(x) - b_g(x)\|_{L_\infty[a,b]} = o(\delta^{d+1})\,. \tag{C.5}$$

The advantage of this result is that, since s_g and $\widehat{g}_s^0$ are in the same function space $S_d(\underline{t})$, calculating the difference between them is simpler, as we shall see later in the proofs, whereas the approximation error of g by $s_g + b_g$ is negligible. Note because they also prove other results, Barrow and Smith (1978) assumed that the knots are generated from a positive density function. However, one can check that to prove the equation (2.7), we need only Assumption (A7). The approximation of g by $s_g + b_g$ is commonly used when establishing asymptotic properties of a spline estimator; see for example Zhou et al. (1998).

To prove the theorem, we write

$$g(x) - \widehat{g}_s^0(x) = g(x) - s_g(x) + s_g(x) - \widehat{g}_s^0(x)\,, \tag{C.6}$$

and below we refer to $g(x) - s_g(x)$ and $s_g(x) - \widehat{g}_s^0(x)$ as, respectively, the first and the second term of (C.6).

First term of (C.6). From the above discussion, we have, for all $x \in [a, b]$,

$$g(x) - s_g(x) = \sum_{j=0}^{K} \mathbb{1}_{[t_j,t_{j+1})}(x) g^{(d+1)}(t_j) h_j(x) + o(\delta^{d+1})\,, \tag{C.7}$$

where $h_j(x) = 1/\{(d+1)!\}\delta_j^{d+1} B_{d+1}\{(x - t_j)/(t_{j+1} - t_j)\}$, for $j = 0, \ldots, K$ and $x \in [a, b]$. Next, we show that $g(x) - s_g(x) = B_\delta(x) + o(\delta^{d+1})$ for all $x \in [a, b]$, where $B_\delta(x)$ is defined in Theorem 1.

Note that $|\sum_{j=0}^{K} \mathbb{1}_{[t_j,t_{j+1})}(x) g^{(d+1)}(t_j) h_j(x) - \sum_{j=0}^{K} \mathbb{1}_{[t_j,t_{j+1})}(x) g^{(d+1)}(x) h_j(x)| \leq \sum_{j=0}^{K} \mathbb{1}_{[t_j,t_{j+1})}(x) |h_j(x)| |g^{(d+1)}(t_j) - g^{(d+1)}(x)|$. For each $K \in \mathbb{Z}^+$ and $x \in [a, b]$, $x \in [t_j, t_{j+1})$ holds only for one $j = 0, \ldots, K$. If $x \notin [t_j, t_{j+1})$, $\mathbb{1}_{[t_j,t_{j+1})}(x)|g^{(d+1)}(t_j) - g^{(d+1)}(x)| = 0$. If $x \in [t_j, t_{j+1})$, $|x - t_j| < |t_{j+1} - t_j| = \delta_j \leq \delta$, where $\delta \to 0$ as $n \to \infty$. Since $g^{(d+1)}$ is continuous on $[a, b]$ under Assumption (A1), we have that for every $\epsilon > 0$, there exists $n_0 \in \mathbb{N}$ such that δ is sufficiently small for the following to hold: $\sum_{j=0}^{K} \mathbb{1}_{[t_j,t_{j+1})}(x) |h_j(x)| |g^{(d+1)}(t_j) - g^{(d+1)}(x)| <$

$\epsilon \sum_{j=0}^{K} \mathbb{1}_{[t_j,t_{j+1})}(x)|h_j(x)|$ for $n \geq n_0$. Therefore, we conclude that for all $x \in [a,b]$,

$$\sum_{j=0}^{K} \mathbb{1}_{[t_j,t_{j+1})}(x) g^{(d+1)}(t_j) h_j(x)$$
$$= g^{(d+1)}(x) \sum_{j=0}^{K} \mathbb{1}_{[t_j,t_{j+1})}(x) h_j(x) + o\left\{ \sum_{j=0}^{K} \mathbb{1}_{[t_j,t_{j+1})}(x)|h_j(x)| \right\}. \tag{C.8}$$

Now, Lehmer (1940) has shown that

$$\min_{t\in[0,1]} B_{d+1}(t) \begin{cases} = -\left(\sum_{i=1}^{\infty} i^{-(d+1)}\right) \frac{2(d+1)!}{(2\pi)^{d+1}}, & \text{when } d+1 \text{ modulo 4 is } 0\,, \\ > -\frac{2(d+1)!}{(2\pi)^{d+1}}, & \text{otherwise}\,, \end{cases} \tag{C.9}$$

and

$$\max_{t\in[0,1]} B_{d+1}(t) \begin{cases} = \left(\sum_{i=1}^{\infty} i^{-(d+1)}\right) \frac{2(d+1)!}{(2\pi)^{d+1}}, & \text{when } d+1 \text{ modulo 4 is } 2\,, \\ < \frac{2(d+1)!}{(2\pi)^{d+1}}, & \text{otherwise}\,, \end{cases} \tag{C.10}$$

where $\sum_{i=1}^{\infty} i^{-(d+1)} \equiv \zeta_d$ is finite for all $d > 0$.

Note that, for any given j and for all $x \in [t_j, t_{j+1})$, $(x-t_j)/(t_{j+1}-t_j) \in [0,1]$. Therefore, using the upper bounds above, under Assumption (A7), we have for all $x \in [a,b]$,

$$\sum_{j=0}^{K} \mathbb{1}_{[t_j,t_{j+1})}(x)|h_j(x)| = O(\delta^{d+1}) \tag{C.11}$$

and

$$\sum_{j=0}^{K} \mathbb{1}_{[t_j,t_{j+1})}(x) h_j(x) = O(\delta^{d+1})\,.$$

Thus, from (C.7), (C.8), (C.11), under Assumption (A2), and using the definition of $B_\delta(x)$ in Theorem 1,

$$g(x) - s_g(x) = B_\delta(x) + o(\delta^{d+1})\,, \tag{C.12}$$

and $B_\delta(x) = O(\delta^{d+1})$.

Second term of (C.6). Regarding the second term, $s_g(x) - \widehat{g}_s^0(x)$, of (C.6), since it is a spline function, s_g can be written in a matrix form similar to $\widehat{g}_s^0$, in such a way that the property of the difference can be studied by investigating the properties of matrices.

In the rest of this section, we decompose the second term of (C.6) into several terms which we study one by one. For simplicity of notation, in the whole section C, we use ϕ and $\mathbf{\Phi}$ to represent the $\phi(q_0)$ and $\mathbf{\Phi}(q_0)$, respectively.

Since $s_g(x) \in S_d(\boldsymbol{t})$, we can write $s_g(x) = \boldsymbol{N}^\top(x)\boldsymbol{\eta}$, where $\boldsymbol{\eta}$ is a vector of coefficients. If we let $\underline{s}_g = \{s_g(X_{1,1}), \ldots, s_g(X_{n_J,J})\}^\top$, we have $\underline{s}_g = \boldsymbol{\mathcal{N}}\boldsymbol{\eta}$, where $\boldsymbol{\mathcal{N}}$ is defined under (C.3). Let $\widehat{\boldsymbol{G}}_{n,\varphi} = (n\delta)^{-1}\{\phi\boldsymbol{\mathcal{N}}^\top\mathbf{\Phi}\mathbf{\Delta}\boldsymbol{\mathcal{N}}\}$ and $\widehat{\boldsymbol{H}}_{n,\lambda,\varphi} = \widehat{\boldsymbol{G}}_{n,\varphi} + (n\delta)^{-1}\lambda\boldsymbol{D}_\ell$. Let the event $\mathcal{E}_{n,1} = \{\widehat{\boldsymbol{G}}_{n,\varphi}$ is invertible$\}$. It follows from Corollary D.7 at page 43 that we have

$$\lim_{n\to\infty} \mathbb{P}(\mathcal{E}_{n,1}) = 1\,. \tag{C.13}$$

Now, on $\mathcal{E}_{n,1}$, we have that for all $x \in [a, b]$,

$$s_g(x) = \boldsymbol{N}^\top(x)\widehat{\boldsymbol{G}}_{n,\varphi}^{-1}\{(n\delta)^{-1}\phi\boldsymbol{\mathcal{N}}^\top\mathbf{\Phi}\mathbf{\Delta}\boldsymbol{\mathcal{N}}\}\boldsymbol{\eta} = \frac{\phi}{n\delta}\boldsymbol{N}^\top(x)\widehat{\boldsymbol{G}}_{n,\varphi}^{-1}\boldsymbol{\mathcal{N}}^\top\mathbf{\Phi}\mathbf{\Delta}\underline{s}_g\,.$$

Then, using (C.3), on $\mathcal{E}_{n,1}$, we have that for all $x \in [a, b]$,

$$s_g(x) - \widehat{g}_s^0(x) = \frac{\phi}{n\delta}\boldsymbol{N}^\top(x)\widehat{\boldsymbol{G}}_{n,\varphi}^{-1}\boldsymbol{\mathcal{N}}^\top\mathbf{\Phi}\mathbf{\Delta}\underline{s}_g - \frac{\phi}{n\delta}\boldsymbol{N}^\top(x)\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}^{-1}\boldsymbol{\mathcal{N}}^\top\mathbf{\Phi}\mathbf{\Delta}\boldsymbol{Q}\boldsymbol{Z}^*\,. \tag{C.14}$$

Next, let $\boldsymbol{G}_n = \mathbb{E}(\widehat{\boldsymbol{G}}_{n,\varphi})$. We shall see below that this matrix does not depend on the $\varphi_j(q_0)$'s. Then using the fact that $N_{i,d+1}(x) = 0$ for $x \notin [a, b]$ and all $i = -d, \ldots, K$, the elements

$$(\boldsymbol{G}_n)_{ij} \;=\; \mathbb{E}\Big\{(\widehat{\boldsymbol{G}}_{n,\varphi})_{ij}\Big\} = \mathbb{E}\left[\frac{1}{n\delta}\{\phi\boldsymbol{\mathcal{N}}^\top\mathbf{\Phi}\mathbf{\Delta}\boldsymbol{\mathcal{N}}\}_{ij}\right]$$

$$
\begin{aligned}
&= \frac{\phi}{n\delta}\sum_{t=1}^{J}\sum_{s=1}^{n_t}\varphi_t(q_0)\mathbb{E}\{\Delta_{s,t}N_{i,d+1}(X_{s,t})N_{j,d+1}(X_{s,t})\}\\
&= \frac{\phi}{n\delta}\sum_{t=1}^{J}\sum_{s=1}^{n_t}\varphi_t(q_0)\mathbb{E}\big[\mathbb{E}\{\Delta_{s,t}N_{i,d+1}(X_{s,t})N_{j,d+1}(X_{s,t})|X_{s,t}\}\big]\\
&= \frac{\phi}{n\delta}\sum_{t=1}^{J}\sum_{s=1}^{n_t}\varphi_t(q_0)\mathbb{E}\big[\mathbb{E}(\Delta_{s,t}|X_{s,t})N_{i,d+1}(X_{s,t})N_{j,d+1}(X_{s,t})\big]\\
&= \frac{\phi}{n\delta}\sum_{t=1}^{J}\sum_{s=1}^{n_t}\varphi_t(q_0)\int_a^b \pi(x)N_{i,d+1}(x)N_{j,d+1}(x)f_X(x)\,dx\\
&= \frac{n}{\delta\sum_{t=1}^{J}n_t\varphi_t(q_0)}\times\frac{\sum_{t=1}^{J}n_t\varphi_t(q_0)}{n}\int_a^b \pi(x)N_{i,d+1}(x)N_{j,d+1}(x)f_X(x)\,dx\\
&= \delta^{-1}\int_a^b \pi(x)N_{i,d+1}(x)N_{j,d+1}(x)f_X(x)\,dx\,,
\end{aligned}
$$

for $i,j=-d,\dots,K$, where we use the definition of ϕ under (C.1).

Recalling the definition of $\boldsymbol{H}_{n,\lambda}$ at (4.6) and the definition of $\boldsymbol{G}_n$ above, we have $\boldsymbol{H}_{n,\lambda} = \boldsymbol{G}_n + (n\delta)^{-1}\lambda\boldsymbol{D}_\ell$. Letting $\underline{g} = \{g(X_{1,1}),\dots,g(X_{n_J,J})\}^\top$, we have

$$
\begin{aligned}
&\phi\Big\{\frac{1}{n\delta}\boldsymbol{N}^\top(x)\widehat{\boldsymbol{G}}_{n,\varphi}^{-1}\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}\boldsymbol{\Delta}\underline{s}_g - \frac{1}{n\delta}\boldsymbol{N}^\top(x)\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}^{-1}\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}\boldsymbol{\Delta}\boldsymbol{Q}\boldsymbol{Z}^*\Big\}\\
&= \phi\Big\{\frac{1}{n\delta}\boldsymbol{N}^\top(x)\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}^{-1}\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}\widehat{\boldsymbol{G}}_{n,\varphi}^{-1}\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}\boldsymbol{\Delta}\underline{s}_g - \frac{1}{n\delta}\boldsymbol{N}^\top(x)\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}^{-1}\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}\boldsymbol{\Delta}\boldsymbol{Q}\boldsymbol{Z}^*\Big\}\\
&= \phi\Big\{\frac{1}{n\delta}\boldsymbol{N}^\top(x)\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}^{-1}\{\widehat{\boldsymbol{G}}_{n,\varphi}+\lambda\boldsymbol{D}_\ell/(n\delta)\}\widehat{\boldsymbol{G}}_{n,\varphi}^{-1}\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}\boldsymbol{\Delta}\underline{s}_g - \frac{1}{n\delta}\boldsymbol{N}^\top(x)\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}^{-1}\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}\boldsymbol{\Delta}\boldsymbol{Q}\boldsymbol{Z}^*\Big\}\\
&= \phi\Big\{\frac{1}{n\delta}\boldsymbol{N}^\top(x)\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}^{-1}\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}\boldsymbol{\Delta}\underline{s}_g + \frac{\lambda}{(n\delta)^2}\boldsymbol{N}^\top(x)\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}^{-1}\boldsymbol{D}_\ell\widehat{\boldsymbol{G}}_{n,\varphi}^{-1}\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}\boldsymbol{\Delta}\underline{s}_g\\
&\quad - \frac{1}{n\delta}\boldsymbol{N}^\top(x)\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}^{-1}\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}\boldsymbol{\Delta}\boldsymbol{Q}\boldsymbol{Z}^*\Big\}\\
&= \phi\Big\{\frac{1}{n\delta}\boldsymbol{N}^\top(x)\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}^{-1}\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}\boldsymbol{\Delta}(\underline{s}_g - \boldsymbol{Q}\boldsymbol{Z}^*) + \frac{\lambda}{(n\delta)^2}\boldsymbol{N}^\top(x)\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}^{-1}\boldsymbol{D}_\ell\widehat{\boldsymbol{G}}_{n,\varphi}^{-1}\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}\boldsymbol{\Delta}\underline{s}_g\Big\}\\
&= \phi\Big\{\frac{1}{n\delta}\boldsymbol{N}^\top(x)\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}^{-1}\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}\boldsymbol{\Delta}(\underline{g} - \boldsymbol{Q}\boldsymbol{Z}^*) + \frac{1}{n\delta}\boldsymbol{N}^\top(x)\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}^{-1}\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}\boldsymbol{\Delta}(\underline{s}_g - \underline{g})
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{\lambda}{(n\delta)^2}\boldsymbol{N}^\top(x)\widehat{\boldsymbol{H}}^{-1}_{n,\lambda,\varphi}\boldsymbol{D}_\ell\widehat{\boldsymbol{G}}^{-1}_{n,\varphi}\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}\boldsymbol{\Delta}\underline{s}_g\Big\} \\
&=\phi\Big\{\frac{1}{n\delta}\boldsymbol{N}^\top(x)\boldsymbol{H}^{-1}_{n,\lambda}\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}\boldsymbol{\Delta}(\underline{g}-\boldsymbol{Q}\boldsymbol{Z}^*)+\frac{1}{n\delta}\boldsymbol{N}^\top(x)(\widehat{\boldsymbol{H}}^{-1}_{n,\lambda,\varphi}-\boldsymbol{H}^{-1}_{n,\lambda})\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}\boldsymbol{\Delta}(\underline{g}-\boldsymbol{Q}\boldsymbol{Z}^*) \\
&\quad+\frac{1}{n\delta}\boldsymbol{N}^\top(x)\widehat{\boldsymbol{H}}^{-1}_{n,\lambda,\varphi}\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}\boldsymbol{\Delta}(\underline{s}_g-\underline{g})+\frac{\lambda}{(n\delta)^2}\boldsymbol{N}^\top(x)\widehat{\boldsymbol{H}}^{-1}_{n,\lambda,\varphi}\boldsymbol{D}_\ell\widehat{\boldsymbol{G}}^{-1}_{n,\varphi}\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}\boldsymbol{\Delta}\underline{s}_g\Big\}.
\end{aligned}
$$

Recalling that $\underline{s}_g=\boldsymbol{N}\boldsymbol{\eta}$, $\widehat{\boldsymbol{G}}_{n,\varphi}=\{\phi\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}\boldsymbol{\Delta}\boldsymbol{\mathcal{N}}\}/(n\delta)$ and, by (E.8), $\boldsymbol{D}_\ell=\nabla_l^\top\boldsymbol{R}\nabla_l$, we can write

$$
\begin{aligned}
\frac{\phi\lambda}{(n\delta)^2}\boldsymbol{N}^\top(x)\widehat{\boldsymbol{H}}^{-1}_{n,\lambda,\varphi}\boldsymbol{D}_\ell\widehat{\boldsymbol{G}}^{-1}_{n,\varphi}\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}\boldsymbol{\Delta}\underline{s}_g &= \frac{\lambda}{n\delta}\boldsymbol{N}^\top(x)\widehat{\boldsymbol{H}}^{-1}_{n,\lambda,\varphi}\boldsymbol{D}_\ell\widehat{\boldsymbol{G}}^{-1}_{n,\varphi}\{(n\delta)^{-1}\phi\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}\boldsymbol{\Delta}\boldsymbol{\mathcal{N}}\}\boldsymbol{\eta} \\
&= \frac{\lambda}{n\delta}\boldsymbol{N}^\top(x)\widehat{\boldsymbol{H}}^{-1}_{n,\lambda,\varphi}\boldsymbol{D}_\ell\boldsymbol{\eta} \\
&= \frac{\lambda}{n\delta}\boldsymbol{N}^\top(x)\widehat{\boldsymbol{H}}^{-1}_{n,\lambda,\varphi}\nabla_l^\top\boldsymbol{R}\nabla_l\boldsymbol{\eta}\,.
\end{aligned}
$$

Thus, from (C.14), on the event $\mathcal{E}_{n,1}$, we have that for all $x\in[a,b]$,

$$
s_g(x)-\widehat{g}^0_s(x)=\mathrm{dif}_{sg}(x)\,,
$$

where

$$
\begin{aligned}
\mathrm{dif}_{sg}(x) &= \frac{\phi}{n\delta}\boldsymbol{N}^\top(x)\boldsymbol{H}^{-1}_{n,\lambda}\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}\boldsymbol{\Delta}(\underline{g}-\boldsymbol{Q}\boldsymbol{Z}^*) \\
&\quad+\frac{\phi}{n\delta}\boldsymbol{N}^\top(x)(\widehat{\boldsymbol{H}}^{-1}_{n,\lambda,\varphi}-\boldsymbol{H}^{-1}_{n,\lambda})\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}\boldsymbol{\Delta}(\underline{g}-\boldsymbol{Q}\boldsymbol{Z}^*) \\
&\quad+\frac{\phi}{n\delta}\boldsymbol{N}^\top(x)\widehat{\boldsymbol{H}}^{-1}_{n,\lambda,\varphi}\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}\boldsymbol{\Delta}(\underline{s}_g-\underline{g}) \\
&\quad+\frac{\lambda}{n\delta}\boldsymbol{N}^\top(x)\widehat{\boldsymbol{H}}^{-1}_{n,\lambda,\varphi}\nabla_l^\top\boldsymbol{R}\nabla_l\boldsymbol{\eta}\,. \qquad \text{(C.15)}
\end{aligned}
$$

Combing (C.12) and (C.15), on event $\mathcal{E}_{n,1}$, we have that for all $x\in[a,b]$,

$$
g(x)-\widehat{g}^0_s(x) = B_\delta(x)+o(\delta^{d+1})+\mathrm{dif}_{sg}(x)\,. \qquad \text{(C.16)}
$$

Thus, using (C.13) and Lemma D.1, we have for all $x \in [a,b]$,

$$\lim_{n\to\infty} \mathbb{P}\{g(x) - \widehat{g}_s^0(x) = B_\delta(x) + o(\delta^{d+1}) + \mathrm{dif}_{sg}(x)\} = 1\,.$$

Applying Lemmas D.23, D.24 and D.26 with (C.15), for all $x \in [a,b]$,

$$\mathrm{dif}_{sg} = \frac{\phi}{n}\boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{N}^\top\boldsymbol{\Phi}\boldsymbol{\Delta}(\underline{g} - \boldsymbol{Q}\boldsymbol{Z}^*) + o_p(\delta^{d+1}) + o_p\{\sqrt{(n\delta)^{-1}}\}\,.$$

Then using Corollary D.4, we have that for all $x \in [a,b]$,

$$g(x) - \widehat{g}^0(x) = B_\delta(x) + \frac{\phi}{n}\boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{N}^\top\boldsymbol{\Phi}\boldsymbol{\Delta}(\underline{g} - \boldsymbol{Q}\boldsymbol{Z}^*) + o_p(\delta^{d+1}) + o_p\{\sqrt{(n\delta)^{-1}}\}\,.$$

Letting

$$AE(x) = B_\delta(x) + \frac{\phi}{n}\boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{N}^\top\boldsymbol{\Phi}\boldsymbol{\Delta}(\underline{g} - \boldsymbol{Q}\boldsymbol{Z}^*)\,,$$

and applying Lemma D.22, the result in Theorem C.1 follows. □

Lemma C.5. *Under Assumption* **A***, we have for all* $x \in [a,b]$, $\widehat{g}_s(x) = \widehat{g}_s^0(x) + o_p\{\sqrt{(n\delta)^{-1}}\}$.

Proof. For $q \in (0,1)$, let $\widehat{\boldsymbol{H}}(q) = \boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}(q)\boldsymbol{\Delta}\boldsymbol{\mathcal{N}}/(n\delta) + \lambda\phi^{-1}(q)\boldsymbol{D}_\ell/(n\delta)$. Then recalling the definition of $\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}$ at page 21, we have $\widehat{\boldsymbol{H}}(q_0) = \phi^{-1}(q_0)\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}$. Under Assumption (A12), the φ_j's are uniformly bounded on $(0,1)$. Thus, applying the same arguments as in Lemma D.14 and Corollary D.15 on $\phi(q)\widehat{\boldsymbol{H}}(q)$ and then combining with the fact that $\phi(q)$ is uniformly bounded, letting the event $\mathcal{E}_{n,2} = \{\widehat{\boldsymbol{H}}(q)$ is invertible for all $q \in (0,1)\}$, we have

$$\lim_{n\to\infty} \mathbb{P}(\mathcal{E}_{n,2}) = 1\,. \tag{C.17}$$

Now, for $q \in (0,1)$, let

$$\widehat{g}(q) = (n\delta)^{-1}\boldsymbol{N}^\top(x)\widehat{\boldsymbol{H}}^{-1}(q)\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}(q)\boldsymbol{\Delta}\boldsymbol{Q}(q)\boldsymbol{Z}^*\,, \tag{C.18}$$

where $\boldsymbol{Q}(q) = \mathrm{diag}(q^{1-n_1} \overset{\times n_1}{\cdots}, \cdots, q^{1-n_J} \overset{\times n_J}{\cdots})$. Then recalling (C.3) and the definition of $\widehat{g}_s(x)$ below (3.15), we have, for all $x \in [a,b]$,

$$\widehat{g}_s^0(x) = \widehat{g}(q_0) \quad \text{and} \quad \widehat{g}_s(x) = \widehat{g}(\widehat{q}_0)\,. \tag{C.19}$$

Thus, if $\widehat{g}(q)$ is differentiable at q_0 and $\widehat{g}'(q_0)$ is bounded, then we can apply Lemma C.1 and a Taylor expansion to $\widehat{g}(q)$ around q_0 to obtain the result. This is what we investigate next.

For the differentiability, recalling the definition of the matrices under (3.13), we have on the event $\mathcal{E}_{n,2}$, for all $q \in (0,1)$,

$$\widehat{g}(q) = (n\delta)^{-1} \sum_{i=-d}^{K} \sum_{j=-d}^{K} N_{i,d+1}(x) \{\widehat{\boldsymbol{H}}^{-1}(q)\}_{ij} \sum_{t=1}^{J} \sum_{s=1}^{n_t} \varphi_t(q) \Delta_{s,t} N_{j,d+1}(X_{s,t}) q^{1-n_t} Z_t^*\,.$$

Using the representation of an inverse matrix in terms of adjoint and determinant, we find that the $\{\widehat{\boldsymbol{H}}^{-1}(q)\}_{ij}$'s are all summations, products and quotients of linear functions of the $\varphi_j(q)$'s. Therefore, $\widehat{g}(q)$ is a differentiable function of the $\varphi_j(q)$'s and the q^{1-n_j}'s. Since the $\varphi_j(q)$'s and the q^{1-n_j}'s are all differentiable at q_0, we deduce that, on the event $\mathcal{E}_{n,2}$, $\widehat{g}(q)$ is differentiable at q_0.

Then, as in the proof of Theorem 1 at page 12, on the event $\mathcal{E}_{n,2}$, we can apply a Taylor expansion of first order to $\widehat{g}(\widehat{q}_0)$ around q_0. In particular, there exists a function h such that

$$\widehat{g}(\widehat{q}_0) = \widehat{g}(q_0) + \widehat{g}'(q_0)(\widehat{q}_0 - q_0) + h(\widehat{q}_0)(\widehat{q}_0 - q_0)\,,$$

and $\lim_{\widehat{q}_0 \to q_0} h(\widehat{q}_0) = 0$. Let $g_n = \widehat{g}'(q_0)(\widehat{q}_0 - q_0) + h(\widehat{q}_0)(\widehat{q}_0 - q_0)$. Using (C.17) with Lemma D.1, we have

$$\lim_{n\to\infty} \mathbb{P}\{\widehat{g}(\widehat{q}_0) - \widehat{g}(q_0) = g_n\} = 1\,. \tag{C.20}$$

Then using Lemma C.1 and Lemma 2.12 of Van der Vaart (2000), we have

$$g_n \;=\; \widehat{g}'(q_0)(\widehat{q}_0 - q_0) + o_p(|\widehat{q}_0 - q_0|) \tag{C.21}$$

and

$$\widehat{q}_0 - q_0 = O_p(n^{-1/2}) = o_p\{\sqrt{(n\delta)^{-1}}\}\,, \tag{C.22}$$

where the last equality comes from the fact that under Assumption (A6), $K \to \infty$ as $n \to \infty$, and under Assumption (A7), $\delta^{-1} \asymp K$.

Next we study $\widehat{g}'(q_0)$. Recalling (C.18) and that $\widehat{\boldsymbol{H}}(q_0) = \phi^{-1}(q_0)\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}$, and using the product rule for matrix-by-scalar derivatives, we have

$$\begin{aligned}
\widehat{g}'(q_0) \;=\; & -\frac{\phi^2(q_0)}{n\delta}\boldsymbol{N}^\top(x)\widehat{\boldsymbol{H}}^{-1}_{n,\lambda,\varphi}\frac{\partial \widehat{\boldsymbol{H}}(q)}{\partial q}\bigg|_{q=q_0}\widehat{\boldsymbol{H}}^{-1}_{n,\lambda,\varphi}\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}(q_0)\boldsymbol{\Delta}\boldsymbol{Q}(q_0)\boldsymbol{Z}^* \\
& +\frac{\phi(q_0)}{n\delta}\boldsymbol{N}^\top(x)\widehat{\boldsymbol{H}}^{-1}_{n,\lambda,\varphi}\boldsymbol{\mathcal{N}}^\top\frac{\partial \boldsymbol{\Phi}(q)}{\partial q}\bigg|_{q=q_0}\boldsymbol{\Delta}\boldsymbol{Q}(q_0)\boldsymbol{Z}^* \\
& +\frac{\phi(q_0)}{n\delta}\boldsymbol{N}^\top(x)\widehat{\boldsymbol{H}}^{-1}_{n,\lambda,\varphi}\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}(q_0)\boldsymbol{\Delta}\frac{\partial \boldsymbol{Q}(q)}{\partial q}\bigg|_{q=q_0}\boldsymbol{Z}^*\,.
\end{aligned}$$

Let A_n, B_n and C_n denote the three terms of the last equation in the order where they appear. For $i = -d, \ldots, K$, let $T_{1,i}$, $T_{2,i}$, $T_{3,i}$ be the $(i{+}d{+}1)$th element of $\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}(q_0)\boldsymbol{\Delta}\boldsymbol{Q}(q_0)\boldsymbol{Z}^*$, $\boldsymbol{\mathcal{N}}^\top[\{\partial\boldsymbol{\Phi}(q)\}/\{\partial q\}]|_{q=q_0}\boldsymbol{\Delta}\boldsymbol{Q}(q_0)\boldsymbol{Z}^*$ and $\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}(q_0)\boldsymbol{\Delta}[\{\partial\boldsymbol{Q}(q)\}/\{\partial q\}]|_{q=q_0}\boldsymbol{Z}^*$, respectively.

By B-spline Property 1 and letting $0 \le i_x \le K$ be the index such that $x \in [t_{i_x}, t_{i_x+1})$ for each $x \in [a, b]$, we have $0 < N_{i,d+1}(x) \le 1$ for all $i \in [i_x - d, i_x]$ and $N_{i,d+1}(x) = 0$, otherwise. According to the discussion at page 36, $\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}$ is a centered $2(d{+}1)-$banded matrix. Noting that similarly to $\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}$, $[\{\partial\widehat{\boldsymbol{H}}(q)\}/\{\partial q\}]|_{q=q_0}$ is also a centered $2(d+1)-$banded matrix; we have $([\{\partial\widehat{\boldsymbol{H}}(q)\}/\{\partial q\}]|_{q=q_0})_{jk} \ne 0$ only if $|j-k| > d+1$ for all $j, k = -d, \ldots, K$. Then using B-spline Property 1, recalling the definition of $\phi(q_0)$ under (C.1) and under Assumption (A10), we have for all $x \in [a, b]$,

$$\begin{aligned}
& |A_n| \\
&= \left|\frac{\phi^2(q_0)}{n\delta}\sum_{i=i_x-d}^{i_x-1}\sum_{j=-d}^{K}\sum_{k=\max(-d,j-d-1)}^{\min(j+d+1,K)}\sum_{\ell=-d}^{K} N_{i,d+1}(x)(\widehat{\boldsymbol{H}}^{-1}_{n,\lambda,\varphi})_{ij}\left(\frac{\partial \widehat{\boldsymbol{H}}(q)}{\partial q}\bigg|_{q=q_0}\right)_{jk}\right.
\end{aligned}$$

$$
\begin{aligned}
&\quad \times (\widehat{\boldsymbol{H}}^{-1}_{n,\lambda,\varphi})_{k\ell}T_{1,\ell}\Bigg| \\
&\le \frac{\phi_1^{-2}}{n\delta}\max_{j,k=-d,\dots,K}\left|\left(\frac{\partial \widehat{\boldsymbol{H}}(q)}{\partial q}\bigg|_{q=q_0}\right)_{jk}\right|\max_{\ell=-d,\dots,K}|T_{1,\ell}| \\
&\quad \times \sum_{i=i_x-d}^{i_x}\sum_{j=-d}^{K}\sum_{k=\max(-d,j-d-1)}^{\min(j+d+1,K)}\sum_{\ell=-d}^{K}|(\widehat{\boldsymbol{H}}^{-1}_{n,\lambda,\varphi})_{ij}||(\widehat{\boldsymbol{H}}^{-1}_{n,\lambda,\varphi})_{k\ell}|\,, \qquad (C.23)\\
|B_n| &= \left|\frac{\phi(q_0)}{n\delta}\sum_{i=i_x-d}^{i_x}\sum_{j=-d}^{K}N_{i,d+1}(x)(\widehat{\boldsymbol{H}}^{-1}_{n,\lambda,\varphi})_{ij}T_{2,j}\right| \\
&\le \frac{\phi_1^{-1}}{n\delta}\max_{j=-d,\dots,K}|T_{2,j}|\sum_{i=i_x-d}^{i_x}\sum_{j=-d}^{K}|(\widehat{\boldsymbol{H}}^{-1}_{n,\lambda,\varphi})_{ij}|\,, \qquad (C.24)
\end{aligned}
$$

and

$$
\begin{aligned}
|C_n| &= \left|\frac{\phi(q_0)}{n\delta}\sum_{i=i_x-d}^{i_x}\sum_{j=-d}^{K}N_{i,d+1}(x)(\widehat{\boldsymbol{H}}^{-1}_{n,\lambda,\varphi})_{ij}T_{3,j}\right| \\
&\le \frac{\phi_1^{-1}}{n\delta}\max_{j=-d,\dots,K}|T_{3,j}|\sum_{i=i_x-d}^{i_x}\sum_{j=-d}^{K}|(\widehat{\boldsymbol{H}}^{-1}_{n,\lambda,\varphi})_{ij}|\,. \qquad (C.25)
\end{aligned}
$$

Let $\tilde{g}_n = |A_n| + |B_n| + |C_n|$. Then the first term in (C.21) satisfies

$$
|\widehat{g'}(q_0)(\widehat{q}_0 - q_0)| \le |\tilde{g}_n(\widehat{q}_0 - q_0)|\,. \qquad (C.26)
$$

Now we calculate the order of $\tilde{g}_n$. Recalling the definition of $\widehat{\boldsymbol{H}}(q)$ at the beginning of the proof, the definition of $\boldsymbol{\mathcal{N}}$ and $\boldsymbol{\Phi}(q)$ at page 11 in Delaigle et al. (2017) and the definition of $\phi(q)$ at page 10 in Delaigle et al. (2017), we have $\widehat{\boldsymbol{H}}(q) = \sum_{t=1}^{J}\sum_{s=1}^{n_t}\boldsymbol{N}(X_{s,t})\varphi_t(q)\Delta_{s,t}\boldsymbol{N}^\top(X_{s,t})/(n\delta) + \lambda\sum_{t=1}^{J}n_t\varphi_t(q)\boldsymbol{D}_\ell/(n^2\delta)$. Thus,

$$
\frac{\partial \widehat{\boldsymbol{H}}(q)}{\partial q}\bigg|_{q=q_0} = \frac{1}{n\delta}\sum_{t=1}^{J}\sum_{s=1}^{n_t}\boldsymbol{N}(X_{s,t})\varphi_t'(q_0)\Delta_{s,t}\boldsymbol{N}^\top(X_{s,t}) + \frac{\sum_{t=1}^{J}n_t\varphi_t'(q_0)}{n}\lambda\boldsymbol{D}_\ell/(n\delta)\,.
$$

Note that under Assumptions (A10) and (A12), for all $t = 1,\dots,J$, $\varphi_t(q_0) > \phi_1 > 0$ and $0 \le |\varphi_t'(q_0)| < \infty$. Moreover, for all $s = 1,\dots,n_j$ and $t = 1,\dots,J$, all the elements in $\boldsymbol{N}(X_{s,t})$ and $\Delta_{s,t}$ are nonnegative. Then using the definition of $\widehat{\boldsymbol{G}}_{n,\varphi}$ at page 21, there exists a constant $0 \leq C_1 = \max_{t=1,\ldots,J} |\varphi_t'(q)| < \infty$ such that

$$
\begin{aligned}
&\max_{j,k=-d,\ldots,K} \left| \left(\frac{\partial \widehat{\boldsymbol{H}}(q)}{\partial q} \bigg|_{q=q_0} \right)_{jk} \right| \\
&\leq \frac{C_1}{\phi_1 n\delta} \Bigg\{ \max_{j,k=-d,\ldots,K} \sum_{t=1}^{J} \sum_{s=1}^{n_t} \left| N_{j,d+1}(X_{s,t}) \varphi_t(q_0) \Delta_{s,t} N_{k,d+1}(X_{s,t}) \right| \\
&\quad + \lambda \phi^{-1}(q_0) \max_{j,k=-d,\ldots,K} |(\boldsymbol{D}_\ell)_{jk}| \Bigg\} \\
&= \frac{C_1}{\phi_1 n\delta} \Bigg\{ \max_{j,k=-d,\ldots,K} \left| \sum_{t=1}^{J} \sum_{s=1}^{n_t} N_{j,d+1}(X_{s,t}) \varphi_t(q_0) \Delta_{s,t} N_{k,d+1}(X_{s,t}) \right| \\
&\quad + \lambda \phi^{-1}(q_0) \max_{j,k=-d,\ldots,K} |(\boldsymbol{D}_\ell)_{jk}| \Bigg\} \\
&\leq \frac{C_1 \phi_2}{\phi_1} \Bigg\{ \max_{j,k=-d,\ldots,K} |(\widehat{\boldsymbol{G}}_{n,\varphi})_{jk}| + \lambda/(n\delta) \max_{j,k=-d,\ldots,K} |(\boldsymbol{D}_\ell)_{jk}| \Bigg\}. \qquad \text{(C.27)}
\end{aligned}
$$

Note that for any symmetric real matrix $\boldsymbol{A} \in \mathbb{R}^{b \times b}$ where b is a positive integer, $\max_{i,j=1,\ldots,b} |(\boldsymbol{A})_{ij}| \leq e_{\boldsymbol{A},m}$, where $e_{\boldsymbol{A},m} = \max\{|e_{\boldsymbol{A}}|$ s.t. $e_{\boldsymbol{A}}$ is an eigenvalue of $\boldsymbol{A}\}$ (see Horn and Johnson, 1985, page 315, equation (6,2)). Then using Lemma D.12, we have $e_{\boldsymbol{D},m} = e_{\boldsymbol{D},K+d+1}$. Thus, $\max_{j,k=-d,\ldots,K} |(\boldsymbol{D}_\ell)_{jk}| \leq c_4 \delta^{-2\ell+1}$ where $0 < c_4 < \infty$ does not depend on n. Similarly, using (D.12), we have

$$
\lim_{n\to\infty} \mathbb{P}\Big\{ \max_{j,k=-d,\ldots,K} |(\widehat{\boldsymbol{G}}_{n,\varphi})_{jk}| \leq e_{\widehat{\boldsymbol{G}},\max} \Big\} = 1 \,. \qquad \text{(C.28)}
$$

Now, letting the event $\mathcal{E}_{n,3} = \{\max_{j,k=-d,\ldots,K} |(\widehat{\boldsymbol{G}}_{n,\varphi})_{jk}| \leq e_{\widehat{\boldsymbol{G}},\max}\}$, we have on event $\mathcal{E}_{n,3}$,

$$
\max_{j,k=-d,\ldots,K} \left| \left(\frac{\partial \widehat{\boldsymbol{H}}(q)}{\partial q} \bigg|_{q=q_0} \right)_{jk} \right| \leq A_{n,1} \,, \qquad \text{(C.29)}
$$

where $A_{n,1} = C_1\phi_2/\phi_1\{e_{\widehat{\boldsymbol{G}},\max} + \lambda/(n\delta) c_4 \delta^{-2\ell+1}\}$. Using Lemma D.6, under Assumption (A7) that $\delta^{-1} \asymp K$ and Assumption (A11), we have

$$A_{n,1} = O_p(1)\,. \tag{C.30}$$

Next, we investigate $\max_{\ell=-d,\ldots,K} |T_{1,\ell}|$. Let $\tilde{T}_{1,\ell} = \{\boldsymbol{\mathcal{N}}^\top \boldsymbol{\Phi}(q_0)\boldsymbol{\Delta}\underline{g}\}_\ell$ for $\ell = -d, \ldots, K$. Using the definition of T_ℓ at the beginning of the proof of Lemma D.23, we have $T_{1,\ell} = T_\ell - \tilde{T}_{1,\ell}$. We know from (D.51) that

$$\max_{\ell=-d,\ldots,K} |T_\ell| = O_p(\sqrt{n\delta \log K})\,. \tag{C.31}$$

It remains to study $\max_{\ell=-d,\ldots,K} |\tilde{T}_{1,\ell}|$.

We have $\tilde{T}_{1,\ell} = \sum_{j=1}^{J} \sum_{i=1}^{n_j} a_{i,j}(\ell)$, where $a_{i,j}(\ell) = \varphi_j(q_0)\Delta_{i,j} N_{\ell,d+1}(X_{i,j}) g(X_{i,j})$. For each fixed $\ell = -d, \ldots, K$, the $a_{i,j}(\ell)$'s are independent random variables. Therefore, using B-spline Property 4,

$$\begin{aligned}
\mathrm{var}(\tilde{T}_{1,\ell}) =& \sum_{j=1}^{J} \sum_{i=1}^{n_j} \varphi_j^2(q_0) \mathrm{var}\{\Delta_{i,j} N_{\ell,d+1}(X_{i,j}) g(X_{i,j})\} \\
\le& \sum_{j=1}^{J} \sum_{i=1}^{n_j} \varphi_j^2(q_0) \mathbb{E}[\{\Delta_{i,j} N_{\ell,d+1}(X_{i,j}) g(X_{i,j})\}^2] \\
=& \sum_{j=1}^{J} \sum_{i=1}^{n_j} \varphi_j^2(q_0) \mathbb{E}\{\mathbb{E}(\Delta_{i,j}|X_{i,j}) N_{\ell,d+1}^2(X_{i,j}) g^2(X_{i,j})\} \\
\le& \sum_{j=1}^{J} \sum_{i=1}^{n_j} \varphi_j^2(q_0) \|\pi\|_{L_\infty[a,b]} \|g\|_{L_\infty[a,b]}^2 \|f_X\|_{L_\infty[a,b]} \|N_{r,d+1}\|_{L_\infty[a,b]} \int_a^b N_{\ell,d+1}(x)\,dx \\
\le& n\delta c_v\,,
\end{aligned}$$

where $c_v = \phi_2^2 \|\pi\|_{L_\infty[a,b]} \|g\|_{L_\infty[a,b]}^2 \|f_X\|_{L_\infty[a,b]} \|N_{r,d+1}\|_{L_\infty[a,b]} \in (0, \infty)$. Furthermore, there exists a positive finite constant $\mathcal{M}$, such that for all ℓ, $|a_{i,j}(\ell)| \le \mathcal{M}$. Now, under Assumption (A7), $\delta^{-1} \asymp K$. Thus, using Assumption (A6), we have $\mathcal{M}\sqrt{\log K/(c_v n\delta)} = o(1)$.

Therefore, the conditions in Lemma D.3 are fulfilled for the sequence $\tilde{T}_{1,\ell}$, and we have $\max_{\ell=-d,\ldots,K}|\tilde{T}_{1,\ell}| = \max_{\ell=-d,\ldots,K}|\mathbb{E}(\tilde{T}_{1,\ell})| + O_p(\sqrt{n\delta\log K})$.

Note that

$$\begin{aligned}
\max_{\ell=-d,\ldots,K}|\mathbb{E}(\tilde{T}_{1,\ell})| &= \max_{\ell=-d,\ldots,K}\left|\sum_{j=1}^{J}\sum_{i=1}^{n_j}\varphi_j(q_0)\mathbb{E}\{\mathbb{E}(\Delta_{i,j}|X_{i,j})N_{\ell,d+1}(X_{i,j})g(X_{i,j})\}\right| \\
&= \max_{\ell=-d,\ldots,K}\left|\sum_{j=1}^{J}\sum_{i=1}^{n_j}\varphi_j(q_0)\int_a^b \pi(x)N_{\ell,d+1}(x)g(x)f_X(x)\,dx\right| \\
&\le \max_{\ell=-d,\ldots,K}\left|\sum_{j=1}^{J}\sum_{i=1}^{n_j}\varphi_j(q_0)\|\pi\|_{L_\infty[a,b]}\|g\|_{L_\infty[a,b]}\|f_X\|_{L_\infty[a,b]}\int_a^b N_{\ell,d+1}(x)\,dx\right| \\
&= O(n\delta)\,.
\end{aligned}$$

Therefore, $\max_{\ell=-d,\ldots,K}|\tilde{T}_{1,\ell}| = O(n\delta) + O_p(\sqrt{n\delta\log K})$. Combining with (C.31), we deduce that $\max_{\ell=-d,\ldots,K}|T_{1,\ell}| = O(n\delta) + O_p(\sqrt{n\delta\log K})$. Under Assumption (A7), we have $\delta^{-1} \asymp K$. Then under Assumption (A6), we have $\sqrt{n\delta\log K} = n\delta\sqrt{\log K/(n\delta)} \asymp n\delta\sqrt{K\log K/n} = o(n\delta)$. Thus,

$$\max_{\ell=-d,\ldots,K}|T_{1,\ell}| = O_p(n\delta)\,. \tag{C.32}$$

Now let us calculate $\max_{\ell=-d,\ldots,K}|T_{2,\ell}|$ and $\max_{\ell=-d,\ldots,K}|T_{3,\ell}|$. Similarly to the arguments above (C.27), by definition of $T_{1,\ell}$ and $T_{2,\ell}$ at page 26, we have

$$\begin{aligned}
\max_{\ell=-d,\ldots,K}|T_{2,\ell}| &= \max_{\ell=-d,\ldots,K}\left|\sum_{t=1}^{J}\sum_{s=1}^{n_t}N_{\ell,d+1}(X_{s,t})\varphi_t'(q_0)\Delta_{s,t}q_0^{1-n_t}Z_t^*\right| \\
&\le \max_{\ell=-d,\ldots,K}\frac{C_1}{\phi_1}\left|\sum_{t=1}^{J}\sum_{s=1}^{n_t}N_{\ell,d+1}(X_{s,t})\varphi_t(q_0)\Delta_{s,t}q_0^{1-n_t}Z_t^*\right| \\
&= \frac{C_1}{\phi_1}\max_{\ell=-d,\ldots,K}\left|\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}(q_0)\boldsymbol{\Delta}\boldsymbol{Q}(q_0)\boldsymbol{Z}^*\right| = O\big(\max_{\ell=-d,\ldots,K}|T_{1,j}|\big) \\
&= O_p(n\delta)\,.
\end{aligned} \tag{C.33}$$

Similarly, using the definition of $T_{3,\ell}$ and $T_{1,\ell}$ at page 26, and Assumption (A9), we find that

$$
\begin{aligned}
\max_{\ell=-d,\ldots,K}|T_{3,\ell}| &= \max_{\ell=-d,\ldots,K}\left|\sum_{t=1}^{J}\sum_{s=1}^{n_t}N_{\ell,d+1}(X_{s,t})\varphi_t(q_0)\Delta_{s,t}(1-n_t)q_0^{-n_t}Z_t^*\right| \\
&\leq \sup_{t=1,\ldots,J}(1-n_t)q_0^{-1}\max_{\ell=-d,\ldots,K}\left|\sum_{t=1}^{J}\sum_{s=1}^{n_t}N_{\ell,d+1}(X_{s,t})\varphi_t(q_0)\Delta_{s,t}q_0^{1-n_t}Z_t^*\right| \\
&= \sup_{t=1,\ldots,J}(1-n_t)q_0^{-1}\max_{\ell=-d,\ldots,K}\left|\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}(q_0)\boldsymbol{\Delta Q}(q_0)\boldsymbol{Z}^*\right| = O\big(\max_{\ell=-d,\ldots,K}|T_{1,j}|\big) \\
&=O_p(n\delta)\,.
\end{aligned} \tag{C.34}
$$

Now we go back to the upper bound for A_n in (C.23). Using (C.29) and Lemma D.18, if $\mathcal{E}_{n,3}\cap\mathcal{E}_n$ holds, then

$$
|A_n| \;\leq\; \frac{\phi_1^{-2}}{n\delta}A_{n,1}\max_{\ell=-d,\ldots,K}|T_{1,\ell}|\sum_{i=i_x-d}^{i_x}\sum_{j=-d}^{K}\sum_{k=\max(-d,j-d-1)}^{\min(j+d+1,K)}\sum_{\ell=-d}^{K}\frac{16}{\tilde{c}_1^2}\nu^{|i-j|+|k-\ell|}\,.
$$

Next we study the right hand side of the above equation, which we denote by $\tilde{A}_n$. To do this, we use a technieque similar to that used in the proof of Lemma 6.4 of Zhou et al. (1998). Specifically, for all $k=\max(j-d-1,-d),\ldots,\min(K,j+d+1)$, we have $|k-\ell|\geq|j-\ell|-d-1$ for $j,\ell=-d,\ldots,K$. Recalling from Lemma D.18 that $0<\inf_{n\in\mathbb{N}}\nu\leq\sup_{n\in\mathbb{N}}\nu<1$, we have $\nu^{|k-\ell|}\leq\nu^{|j-\ell|-d-1}$. Then,

$$
\begin{aligned}
&\sum_{i=i_x-d}^{i_x}\sum_{j=-d}^{K}\sum_{\ell=-d}^{K}\nu^{|i-j|+|j-\ell|} \\
&=\sum_{i=i_x-d}^{i_x}\sum_{\ell=-d}^{K}\left(\sum_{j=\min(i,\ell)+1}^{\max(i,\ell)}\nu^{|i-\ell|}+\sum_{j=-d}^{\min(i,\ell)}\nu^{|i-\ell|+2\min(i,\ell)-2j}+\sum_{j=\max(i,\ell)+1}^{K}\nu^{|i-\ell|+2j-2\max(i,\ell)}\right) \\
&=\sum_{i=i_x-d}^{i_x}\sum_{\ell=-d}^{K}\nu^{|i-\ell|}\left(|i-\ell|+\sum_{j=0}^{\min(i,\ell)+d}\nu^{2j}+\sum_{j=1}^{K-\max(i,\ell)}\nu^{2j}\right) \\
&=O(1)\,.
\end{aligned}
$$

Combining with (C.30) and (C.32), we conclude that $\tilde{A}_n = O_p(1)$.

Similarly, using the definition of B_n in (C.24) and Lemma D.18, on the event $\mathcal{E}_n$, we have $|B_n| \leq \tilde{B}_n$, where $\tilde{B}_n = \phi_1^{-1}/(n\delta) \max_{\ell=-d,\ldots,K} |T_{2,j}| \sum_{i=i_x-d}^{i_x} \sum_{j=-d}^{K} 4/\tilde{c}_1 \nu^{|i-j|}$. Using (C.33), we have $\tilde{B}_n = O_p(1)$. Moreover, using the dinition of C_n in (C.25) and Lemma D.18, we find that if $\mathcal{E}_n$ holds, $|C_n| \leq \tilde{C}_n$, where $\tilde{C}_n = \phi_1^{-1}/(n\delta) \max_{\ell=-d,\ldots,K} |T_{3,\ell}| \sum_{i=i_x-d}^{i_x} \sum_{j=-d}^{K} 4/\tilde{c}_1 \nu^{|i-j|} = O_p(1)$ if we use (C.34).

Thus, combining the bound on $|A_n|, |B_n|$ and $|C_n|$ with (C.26), if $\mathcal{E}_{n,3} \cap \mathcal{E}_n$ holds, we get $|\widehat{g}'(q_0)(\widehat{q}_0 - q_0)| \leq (|\tilde{A}_n| + |\tilde{B}_n| + |\tilde{C}_n|)(\widehat{q}_0 - q_0)$. Moreover, using (C.22) and the above calculations, $(|\tilde{A}_n| + |\tilde{B}_n| + |\tilde{C}_n|)(\widehat{q}_0 - q_0) = o_p\{\sqrt{(n\delta)^{-1}}\}$. Then combining with (C.17) and (D.26), using Lemma D.1 and Corollary D.4, we conclude that

$$|\widehat{g}'(q_0)(\widehat{q}_0 - q_0)| = o_p\{\sqrt{(n\delta)^{-1}}\}\,.$$

Combining with (C.21), we deduce that $g_n = o_p\{\sqrt{(n\delta)^{-1}}\}$. Then using (C.20), (C.19) and Corollary D.4, the result follows. □

D Auxiliary lemmas for Theorem C.1

In this section, we provide the auxiliary lemmas that are useful to prove Theorem C.1.

D.1 Side results

Lemma D.1. *Let A_n and B_n be two events depending on a real number $n > 0$. If $\lim_{n\to\infty} \mathbb{P}(A_n) = 1$ and, for all n, A_n implies B_n, then $\lim_{n\to\infty} \mathbb{P}(B_n) = 1$.*

Proof. We have $\mathbb{P}(B_n) \geq \mathbb{P}(A_n)$ for all n. Then $\lim_{n\to\infty} \mathbb{P}(B_n) \geq \lim_{n\to\infty} \mathbb{P}(A_n) = 1$. □

Lemma D.2. *Let $A_{n,1}, \ldots, A_{n,m}$ be a set of events depending on a real number $n > 0$, where m is a finite positive integer. If $\lim_{n\to\infty} \mathbb{P}(A_{n,i}) = 1$ for $i = 1, \ldots, m$, then $\lim_{n\to\infty} \mathbb{P}(\cap_{i=1}^{m} A_{n,i}) = 1$.*

Proof. Note that $\mathbb{P}(\cap_{i=1}^m A_{n,i}) = 1 - \mathbb{P}(\cup_{i=1}^m A_{n,i}^{\mathrm{c}})$. Since $\mathbb{P}(\cup_{i=1}^m A_{n,i}^{\mathrm{c}}) \leq \sum_{i=1}^m \mathbb{P}(A_{n,i}^{\mathrm{c}})$ and m is finite, we have

$$\lim_{n\to\infty} \mathbb{P}(\cup_{i=1}^m A_{n,i}^{\mathrm{c}}) \leq \sum_{i=1}^m \lim_{n\to\infty} \mathbb{P}(A_{n,i}^{\mathrm{c}}) = 0\,.$$

Thus, $\lim_{n\to\infty} \mathbb{P}(\cap_{i=1}^m A_{n,i}) = 1 - \lim_{n\to\infty} \mathbb{P}(\cup_{i=1}^m A_{n,i}^{\mathrm{c}}) = 1$. □

Lemma D.3. *Let $T_{n,i} = \sum_{j=1}^n a_{j,i}$, $i = 1, \ldots, m$ be a sequence of random variables depending on a positive integer n, with m a positive integer, where $m \to \infty$ as $n \to \infty$, and for all $i = 1, \ldots, m$, $a_{1,i}, \ldots, a_{n,i}$ are independent random variables. If there exist some positive real sequences $\mathcal{M}_n, c_{v,n}$ and a constant $\tilde{n} > 0$, such that for all $n \geq \tilde{n}$,*

- *for all $i = 1, \ldots, m$ and $j = 1, \ldots, n$, $|a_{j,i}| \leq \mathcal{M}_n$;*
- *for all $i = 1, \ldots, m$, $var(T_{n,i}) \leq c_{v,n}$;*
- *$\mathcal{M}_n \sqrt{\log m / c_{v,n}} \leq 1$,*

then $\max_{i=1,\ldots,m} |T_{n,i}| = \max_{i=1,\ldots,m} |\mathbb{E}(T_{n,i})| + O_p(\sqrt{c_{v,n} \log m})$.

Proof. Note that if $|a_{j,i}| \leq \mathcal{M}_n$, we have $|a_{j,i} - \mathbb{E}(a_{j,i})| \leq 2\mathcal{M}_n$. Thus, under the conditions in the statement, we have for all $n \geq \tilde{n}$, and $i = 1, \ldots, m$, $a_{1,i} - \mathbb{E}(a_{1,i}), \ldots, a_{n,i} - \mathbb{E}(a_{n,i})$ are independent random variables with zero mean and $|a_{j,i} - \mathbb{E}(a_{j,i})| \leq 2\mathcal{M}_n$. Moreover, $\sum_{j=1}^n \mathrm{var}\{a_{j,i} - \mathbb{E}(a_{j,i})\} \leq c_{v,n}$. Then applying Bernstein's inequality (see page 193 of Pollard, 2012) to $T_{n,i}$ for each fixed $n \geq \tilde{n}$ and $i = 1, \ldots, m$, we have for any $\xi > 0$,

$$\mathbb{P}(|T_{n,i} - \mathbb{E}(T_{n,i})| \geq \xi) \leq 2 \exp\left\{ - \frac{(1/2)\xi^2}{c_{v,n} + (2/3)\mathcal{M}_n \xi} \right\}.$$

Selecting $\xi = \sqrt{\tilde{c} c_{v,n} \log m}$ for some positive constant $\tilde{c}$ that dose not depend on n, we have for all $n \geq \tilde{n}$,

$$\mathbb{P}\Big\{ \max_{i=1,\ldots,m} |T_{n,i} - \mathbb{E}(T_{n,i})| \geq \xi \Big\} = \mathbb{P}\Big\{ |T_{n,i} - \mathbb{E}(T_{n,i})| \geq \xi \text{ for at least one } i = 1, \ldots, m \Big\}$$

$$
\begin{aligned}
&\leq 2m \exp\left\{ -\frac{(1/2)\xi^2}{c_{v,n} + (2/3)\mathcal{M}_n \xi} \right\} \\
&= 2m \exp\left\{ -\frac{(1/2)\tilde{c}\log m}{1 + (2/3)\mathcal{M}_n \sqrt{\tilde{c}\log m / c_{v,n}}} \right\} \\
&= 2m^{1-\tilde{c}/C} \,,
\end{aligned}
$$

where $C = 2\{1 + (2/3)\mathcal{M}_n\sqrt{\tilde{c}\log m/c_{v,n}}\}$. Since $m \geq 1$, and for all $n \geq \tilde{n}$, $\mathcal{M}_n\sqrt{\log m/c_{v,n}} \leq 1$, we have for all $n \geq \tilde{n}$,

$$
\mathbb{P}\Big\{ \max_{i=1,\dots,m} |T_{n,i} - \mathbb{E}(T_{n,i})| \geq \sqrt{\tilde{c} c_{v,n} \log m} \Big\} \leq 2m^{1-\tilde{c}/2\{1+(2/3)\sqrt{\tilde{c}}\}} \,.
$$

Note that $\tilde{c}/2\{1 + (2/3)\sqrt{\tilde{c}}\} \to \infty$ as $\tilde{c} \to \infty$. Therefore,

$$
\lim_{\tilde{c}\to\infty} \limsup_{n\to\infty} \mathbb{P}\Big\{ \max_{i=1,\dots,m} |T_{n,i} - \mathbb{E}(T_{n,i})| \geq \sqrt{\tilde{c} c_{v,n} \log m} \Big\} = 0 \,,
$$

so that $\max_{i=1,\dots,m} |T_{n,i} - \mathbb{E}(T_{n,i})| = O_p(\sqrt{c_v \log m})$. Using the fact that $\big| \max_{i=1,\dots,m} |T_{n,i}| - \max_{i=1,\dots,m} |\mathbb{E}(T_{n,i})| \big| \leq \max_{i=1,\dots,m} |T_{n,i} - \mathbb{E}(T_{n,i})|$, the result follows. □

Corollary D.4. *Let A_n, B_n be two random variables depending on a real number n. Suppose that $\lim_{n\to\infty} \mathbb{P}(|A_n| \leq |B_n|) = 1$. If $B_n = o_p(a_n)$, then $A_n = o_p(a_n)$.*

Proof. Note that $B_n = o_p(a_n)$ means that for any $\xi > 0$,

$$
\lim_{n\to\infty} \mathbb{P}(|B_n/a_n| < \xi) = 1 \,.
$$

Then applying Lemma D.2, we have for any $\xi > 0$,

$$
\lim_{n\to\infty} \mathbb{P}\{(|B_n/a_n| < \xi) \cap (|A_n| \leq |B_n|)\} = 1 \,.
$$

For any $\xi > 0$ and all n, the intersection of the event $|B_n/a_n| < \xi$ and the event $|A_n| \leq |B_n|$ implies the event that $|A_n/a_n| < \xi$. Thus, applying Lemma D.1, we have for any $\xi > 0$,

$$
\lim_{n\to\infty} \mathbb{P}(|A_n/a_n| < \xi) = 1 \,.
$$

□

D.2 Properties of matrices used in the proof of Theorem 1

In this subsection, we study some properties of the matrices $\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}$, $\widehat{\boldsymbol{G}}_{n,\varphi}$, $\boldsymbol{G}_n$ defined at page 21 and $\boldsymbol{H}_{n,\lambda}$ defined in (4.6). As we shall see they all have a similar structure, which will be useful to derive our results.

Recalling from the definition of $\widehat{\boldsymbol{G}}_{n,\varphi}$ at page 21 and $\boldsymbol{\Phi}, \boldsymbol{\Delta}, \boldsymbol{\mathcal{N}}$ at page 11 in Delaigle et al. (2017), we have, for $i, j = -d, \dots, K$,

$$(\widehat{\boldsymbol{G}}_{n,\varphi})_{ij} = \left\{ \frac{\phi \boldsymbol{\mathcal{N}}^\top \boldsymbol{\Phi} \boldsymbol{\Delta} \boldsymbol{\mathcal{N}}}{n\delta} \right\}_{ij} = \frac{\phi}{n\delta} \sum_{t=1}^{J} \sum_{s=1}^{n_t} \varphi_t(q_0) \Delta_{s,t} N_{i,d+1}(X_{s,t}) N_{j,d+1}(X_{s,t}) \,, \tag{D.1}$$

where, from (E.2), we have, for $i = -d, \dots, K$,

$$N_{i,d+1}(x) \neq 0 \,, \text{ iff } x \in (t_i, t_{i+d+1}) \,.$$

We deduce that

$$\text{if } \ |i-j| > d+1 \,, \quad N_{i,d+1}(x) N_{j,d+1}(x) = 0 \ \text{ for all } \ x \in \mathbb{R} \,,$$

which implies that

$$\text{if } \ |i-j| > d+1 \,, \quad (\widehat{\boldsymbol{G}}_{n,\varphi})_{ij} = 0 \,.$$

A matrix of that type is called a centered $2(d+1)-$banded matrix (see Demko et al., 1984 for example). Similarly,

$$\begin{aligned} (\boldsymbol{G}_n)_{ij} &= \delta^{-1} \int_a^b \pi(x) N_{i,d+1}(x) N_{j,d+1}(x) f_X(x)\, dx \\ &= 0 \ \text{ if } \ |i-j| > d+1 \,. \end{aligned} \tag{D.2}$$

Thus, $\boldsymbol{G}_n$ is also a centered $2(d+1)-$banded matrix.

Furthermore, recalling the definition under (C.3), $(\boldsymbol{D}_\ell)_{ij} = \int_a^b N^{(\ell)}_{i,d+1}(x) N^{(\ell)}_{j,d+1}(x)\, dx$. By recursively repeating (E.4) ℓ times where $0 \le \ell < d$, we see that for $i = -d, \dots, K$,

$N^{(\ell)}_{i,d+1}(x)$ is a linear combination of $N_{i,d+1-\ell}(x), \ldots, N_{i+\ell,d+1-\ell}(x)$. Again from (E.2), for $k = -d, \ldots, K+\ell$, if $x \notin (t_k, t_{k+d+1-\ell})$, $N_{k,d+1-\ell}(x) = 0$. Therefore, for $0 \le \ell < d$, $i = -d, \ldots, K$,

$$N^{(\ell)}_{i,d+1}(x) = 0 \ \text{ if } \ x \notin (t_i, t_{i+d+1})\,. \tag{D.3}$$

Then for all $x \in [a,b]$,

$$N^{(\ell)}_{i,d+1}(x) N^{(\ell)}_{j,d+1}(x) = 0 \ \text{ if } \ |i-j| > d+1\,,$$

so that

$$(\boldsymbol{D}_\ell)_{ij} = 0 \ \text{ if } \ |i-j| > d+1\,.$$

Since $\widehat{\boldsymbol{G}}_{n,\varphi}$, $\boldsymbol{G}_n$ and $\boldsymbol{D}_\ell$ are all centered $2(d+1)-$banded matrices, $\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}$ and $\boldsymbol{H}_{n,\lambda}$ which are the linear combinations of them, are also centered $2(d+1)-$banded matrices.

The next three lemmas concern the eigenvalues of $\boldsymbol{G}_n$, $\widehat{\boldsymbol{G}}_{n,\varphi}$ and the difference between them. The techniques used for proving them are adapted straightforwardly from Lemma 6.2 of Zhou et al. (1998) and Lemma 1 of Holland (2017), where the authors have investigated the eigenvalue problem of the matrix $\boldsymbol{\mathcal{N}}^\top \boldsymbol{\mathcal{N}}$.

Lemma D.5 (Eigenvalues of $\boldsymbol{G}_n$)**.** *Let $e_{\boldsymbol{G},\min}$ and $e_{\boldsymbol{G},\max}$ denote the minimum and maximum eigenvalues of $\boldsymbol{G}_n$, respectively. Under Assumptions (A5) to (A10),*

$$\begin{aligned} e_{\boldsymbol{G},\min} &= c_1\,, \\ e_{\boldsymbol{G},\max} &= c_2\,, \end{aligned} \tag{D.4}$$

where c_1, c_2 are constants such that $0 < \pi_{\min} f_{\min} (d+1)^{-1} C_d^2 M^{-1} \le c_1 \le c_2 \le f_{\max}$ for all n and $n \to \infty$, with $\pi_{\min}, f_{\min}$ and $f_{\max}$ defined in Assumptions (A5) and (A8), M a constant defined in Assumption (A7) and C_d a constant defined in Lemma E.1 at page 85.

Proof. By definition of the eigenvalues of symmetric matrices, $e_{\boldsymbol{G},\min} = \min_{\boldsymbol{v}^\top \boldsymbol{v}=1} \boldsymbol{v}^\top \boldsymbol{G}_n \boldsymbol{v}$ and $e_{\boldsymbol{G},\max} = \max_{\boldsymbol{v}^\top \boldsymbol{v}=1} \boldsymbol{v}^\top \boldsymbol{G}_n \boldsymbol{v}$ where $\boldsymbol{v}$ denotes a column vector of length $K+d+1$. Let

$s(x) = \boldsymbol{N}^\top(x)\boldsymbol{v}$. Then recalling (D.2), we have for all vectors $\boldsymbol{v} = (v_{-d}, \dots, v_K)^\top \in \mathbb{R}^{K+d+1}$,

$$\boldsymbol{v}^\top \boldsymbol{G}_n \boldsymbol{v} = \delta^{-1} \int_a^b \pi(x) \boldsymbol{v}^\top \boldsymbol{N}(x) \boldsymbol{N}^\top(x) \boldsymbol{v} f_X(x)\, dx = \delta^{-1} \int_a^b \pi(x) s^2(x) f_X(x)\, dx\,.$$

Note that $s(x)$ is a spline function in $S_d(\boldsymbol{t})$ with coefficient vector $\boldsymbol{v}$. Thus, applying Lemma E.1 with $q = 2$, we have

$$(d+1)^{-1} C_d^2 \sum_{i=-d}^{K} v_i^2 (t_{i+d+1} - t_i) \le \int_a^b s^2(x)\, dx \le (d+1)^{-1} \sum_{i=-d}^{K} v_i^2 (t_{i+d+1} - t_i)\,. \tag{D.5}$$

Under Assumptions (A5) and (A8), we deduce that

$$\begin{aligned} &\pi_{\min} f_{\min} (d+1)^{-1} C_d^2 \sum_{i=-d}^{K} v_i^2 (t_{i+d+1} - t_i) \\ &\le \int_a^b \pi(x) s^2(x) f_X(x)\, dx \le f_{\max} (d+1)^{-1} \sum_{i=-d}^{K} v_i^2 (t_{i+d+1} - t_i)\,. \end{aligned}$$

Using Assumption (A7), we deduce that

$$\pi_{\min} f_{\min} (d+1)^{-1} C_d^2 M^{-1} \delta \sum_{i=-d}^{K} v_i^2 \le \int_a^b \pi(x) s^2(x) f_X(x)\, dx \le f_{\max} \delta \sum_{i=-d}^{K} v_i^2\,.$$

Therefore, for any $\boldsymbol{v}$ such that $\boldsymbol{v}^\top \boldsymbol{v} = 1$,

$$\pi_{\min} f_{\min} (d+1)^{-1} C_d^2 M^{-1} \le \boldsymbol{v}^\top \boldsymbol{G}_n \boldsymbol{v} \le f_{\max}\,.$$

It follows that we can find some constants $0 < \pi_{\min} f_{\min} (d+1)^{-1} C_d^2 M^{-1} \le c_1 \le c_2 \le f_{\max}$ such that,

$$\begin{aligned} e_{\boldsymbol{G},\min} &= \min\nolimits_{\boldsymbol{v}^\top \boldsymbol{v} = 1} \boldsymbol{v}^\top \boldsymbol{G}_n \boldsymbol{v} = c_1\,, \\ e_{\boldsymbol{G},\max} &= \max\nolimits_{\boldsymbol{v}^\top \boldsymbol{v} = 1} \boldsymbol{v}^\top \boldsymbol{G}_n \boldsymbol{v} = c_2\,. \end{aligned}$$

□

Lemma D.6 (Eigenvalues of $\widehat{\boldsymbol{G}}_{n,\varphi}$)**.** *Let $e_{\widehat{\boldsymbol{G}},\min}$ and $e_{\widehat{\boldsymbol{G}},\max}$ be the minimum and maximum eigenvalues of $\widehat{\boldsymbol{G}}_{n,\varphi}$, respectively. Under Assumptions (A5) to (A10), we have*

$$e_{\widehat{\boldsymbol{G}},\min} = c_1\{1 + o_p(1)\}\,, \tag{D.6}$$

$$e_{\widehat{\boldsymbol{G}},\max} = c_2\{1 + o_p(1)\}\,, \tag{D.7}$$

where c_1, c_2 are defined as in Lemma D.5.

Proof. Again, by the definition of the eigenvalues of symmetric matrices, $e_{\widehat{\boldsymbol{G}},\min} = \min_{\boldsymbol{v}^\top\boldsymbol{v}=1} \boldsymbol{v}^\top \widehat{\boldsymbol{G}}_{n,\varphi}\boldsymbol{v}$ and $e_{\widehat{\boldsymbol{G}},\max} = \max_{\boldsymbol{v}^\top\boldsymbol{v}=1} \boldsymbol{v}^\top \widehat{\boldsymbol{G}}_{n,\varphi}\boldsymbol{v}$, where $\boldsymbol{v}$ denotes a vector of length $K+d+1$. We will study those eigenvalues using Lemma D.5 and the upper bound of the difference between the eigenvalues of $\widehat{\boldsymbol{G}}_{n,\varphi}$ and $\boldsymbol{G}_n$.

Recalling (D.1) and the definition of $\boldsymbol{G}_n$, for all vectors $\boldsymbol{v} \in \mathbb{R}^{K+d+1}$, we have

$$\begin{aligned}
&|\boldsymbol{v}^\top(\widehat{\boldsymbol{G}}_{n,\varphi} - \boldsymbol{G}_n)\boldsymbol{v}| = \left|\sum_{i=-d}^{K}\sum_{j=-d}^{K} v_i\{(\widehat{\boldsymbol{G}}_{n,\varphi})_{ij} - (\boldsymbol{G}_n)_{ij}\}v_j\right| \\
&= \left|\frac{\phi}{n\delta}\sum_{t=1}^{J}\sum_{s=1}^{n_t}\varphi_t(q_0)\sum_{k=0}^{K}\sum_{i=-d}^{K}\sum_{j=-d}^{K} v_i v_j\left[N_{i,d+1}(X_{s,t})\Delta_{s,t}N_{j,d+1}(X_{s,t})\mathbb{1}_{[t_k,t_k+1]}(X_{s,t})\right.\right. \\
&\qquad \left.\left. - \mathbb{E}\{N_{i,d+1}(X_{s,t})\Delta_{s,t}N_{j,d+1}(X_{s,t})\mathbb{1}_{[t_k,t_k+1]}(X_{s,t})\}\right]\right|.
\end{aligned}$$

Using B-spline Property 1, for each $k = 0, \ldots, K$, $N_{i,d+1}(X_{s,t})\mathbb{1}_{[t_k,t_k+1]}(X_{s,t}) = 0$ if $i \notin [k-d, k]$ for all $X_{s,t} \in [a, b]$. Thus, for all vectors $\boldsymbol{v} \in \mathbb{R}^{K+d+1}$,

$$\begin{aligned}
&|\boldsymbol{v}^\top(\widehat{\boldsymbol{G}}_{n,\varphi} - \boldsymbol{G}_n)\boldsymbol{v}| \\
&\leq \sum_{k=0}^{K}\sum_{i=k-d}^{k}\sum_{j=k-d}^{k}|v_i v_j|\left|\sum_{t=1}^{J}\sum_{s=1}^{n_t}\frac{\phi}{n\delta}\varphi_t(q_0)\left[\Delta_{s,t}N_{i,d+1}(X_{s,t})N_{j,d+1}(X_{s,t})\mathbb{1}_{[t_k,t_k+1]}(X_{s,t})\right.\right. \\
&\qquad \left.\left. - \mathbb{E}\{\Delta_{s,t}N_{i,d+1}(X_{s,t})N_{j,d+1}(X_{s,t})\mathbb{1}_{[t_k,t_k+1]}(X_{s,t})\}\right]\right|.
\end{aligned} \tag{D.8}$$

To study the upper bound of $|\boldsymbol{v}^\top(\widehat{\boldsymbol{G}}_{n,\varphi} - \boldsymbol{G}_n)\boldsymbol{v}|$, we first apply Bernstein's inequality (see page 193 of Pollard, 2012) to $\sum_{t=1}^{J}\sum_{s=1}^{n_t}|\widehat{A}_{ij,k}(s,t) - A_{ij,k}(s,t)|$, where

$$\widehat{A}_{ij,k}(s,t) = \frac{\phi}{n\delta}\varphi_t(q_0)\Delta_{s,t}N_{i,d+1}(X_{s,t})N_{j,d+1}(X_{s,t})\mathbb{1}_{[t_k,t_k+1]}(X_{s,t})$$

and

$$A_{ij,k}(s,t) = \frac{\phi}{n\delta}\varphi_t(q_0)\mathbb{E}\{\Delta_{s,t}N_{i,d+1}(X_{s,t})N_{j,d+1}(X_{s,t})\mathbb{1}_{[t_k,t_k+1]}(X_{s,t})\}\,,$$

for $i, j = -d, \dots, K$, $k = 0, \dots, K$.

Using the definition of ϕ below (C.1), and using Assumption (A10), we have $\phi = n/\{\sum_{j=1}^{J} n_j\varphi_j(q_0)\} \leq \phi_1^{-1}n/(\sum_{j=1}^{J} n_j) = \phi_1^{-1} < \infty$. Then under Assumption (A10) and using B-spline Property 1, we have, for $s = 1, \dots, n_t$, $t = 1, \dots, J$,

$$|\widehat{A}_{ij,k}(s,t) - A_{ij,k}(s,t)| \leq \frac{\phi_2\phi_1^{-1}}{n\delta}\,,$$

almost surely.

Moreover, under Assumption (A10), since the $(\Delta_{s,t}, X_{s,t})$'s are i.i.d., we have

$$\begin{aligned}
&\sum_{t=1}^{J}\sum_{s=1}^{n_t}\mathrm{var}\{\widehat{A}_{ij,k}(s,t) - A_{ij,k}(s,t)\}\\
&= \frac{\phi^2}{(n\delta)^2}\sum_{t=1}^{J}\sum_{s=1}^{n_t}\varphi_t^2(q_0)\mathrm{var}\{\Delta_{s,t}N_{i,d+1}(X_{s,t})N_{j,d+1}(X_{s,t})\mathbb{1}_{[t_k,t_k+1]}(X_{s,t})\}\\
&\leq \frac{\phi\phi_2}{n\delta^2}\times\frac{\phi}{n}\sum_{t=1}^{J}\sum_{s=1}^{n_t}\varphi_t(q_0)\mathbb{E}\{\Delta_{s,t}^2N_{i,d+1}^2(X_{s,t})N_{j,d+1}^2(X_{s,t})\mathbb{1}_{[t_k,t_k+1]}(X_{s,t})\}\\
&= \frac{\phi\phi_2}{n\delta^2}\times\frac{\phi}{n}\sum_{t=1}^{J}\sum_{s=1}^{n_t}\varphi_t(q_0)\mathbb{E}\{\mathbb{E}(\Delta_{s,t}|X_{s,t})N_{i,d+1}^2(X_{s,t})N_{j,d+1}^2(X_{s,t})\mathbb{1}_{[t_k,t_k+1]}(X_{s,t})\}\\
&\leq \frac{\phi\phi_2}{n\delta^2}\times\frac{n}{\sum_{t=1}^{J} n_t\varphi_t(q_0)}\times\frac{\sum_{t=1}^{J} n_t\varphi_t(q_0)}{n}\int_a^b \pi(x)N_{i,d+1}^2(x)N_{j,d+1}^2(x)f_X(x)\,dx\\
&= \frac{\phi\phi_2}{n\delta^2}\int_a^b \pi(x)N_{i,d+1}^2(x)N_{j,d+1}^2(x)f_X(x)\,dx
\end{aligned}$$

$$
\begin{aligned}
&\leq \frac{\phi\phi_2\pi_{\max}f_{\max}}{n\delta^2}\int_a^b N_{j,d+1}(x)\,dx \\
&\leq \frac{\phi\phi_2\pi_{\max}f_{\max}}{n\delta^2}\frac{t_{j+d+1}-t_j}{d+1} \\
&\leq \phi\phi_2\pi_{max}f_{\max}n^{-1}\delta^{-1}\,,
\end{aligned}
$$

where we used B-spline properties, and where the last inequality comes from Assumption (A7).

It follows from those calculations that for all $i,j = -d,\ldots,K$, $k = 0,\ldots,K$, $\widehat{A}_{ij,k}(s,t)-A_{ij,k}(s,t)$ are independent random variables with mean zero and bounded range: $|\widehat{A}_{ij,k}(s,t)-A_{ij,k}(s,t)| \leq n^{-1}\delta^{-1}\mathcal{M}$, where $\mathcal{M} = \phi_2\phi_1^{-1}$, and $s = 1,\ldots,n_t$ and $t = 1,\ldots,J$. Moreover, the sum of the variances of the $\widehat{A}_{ij,k}(s,t)-A_{ij,k}(s,t)$'s is

$$
\sum_{t=1}^{J}\sum_{s=1}^{n_t}\mathrm{var}\{\widehat{A}_{ij,k}(s,t)-A_{ij,k}(s,t)\} \leq c_v n^{-1}\delta^{-1}\,,
$$

where $c_v = \phi\phi_2\pi_{\max}f_{\max}$. Thus, we can apply Bernstein's inequality to $\sum_{t=1}^{J}\sum_{s=1}^{n_t}\{\widehat{A}_{ij,k}(s,t)-A_{ij,k}(s,t)\}$. Specifically, for all $i,j = -d,\ldots,K$, $k = 0,\ldots,K$, and each $\xi > 0$, we have

$$
\mathbb{P}\bigg\{\bigg|\sum_{t=1}^{J}\sum_{s=1}^{n_t}\{\widehat{A}_{ij,k}(s,t)-A_{ij,k}(s,t)\}\bigg| \geq \xi\bigg\} \leq 2\exp\bigg\{-\frac{(1/2)n\delta\xi^2}{c_v+(1/3)\mathcal{M}\xi}\bigg\}. \tag{D.9}
$$

Now recall (D.8). If $\big|\sum_{t=1}^{J}\sum_{s=1}^{n_t}\{\widehat{A}_{ij,k}(s,t)-A_{ij,k}(s,t)\}\big| < \xi$ for all i,j,k and any $\xi > 0$, then for all $\boldsymbol{v} = (v_{-d},\ldots,v_K)^\top$ such that $\boldsymbol{v}^\top\boldsymbol{v} = 1$, we have

$$
\begin{aligned}
|\boldsymbol{v}^\top(\widehat{\boldsymbol{G}}_{n,\varphi}-\boldsymbol{G}_n)\boldsymbol{v}| &< \xi\sum_{k=0}^{K}\sum_{i=k-d}^{k}\sum_{j=k-d}^{k}|v_i||v_j| \\
&\leq \xi\sum_{k=0}^{K}(d+1)^{1/2}\bigg\{\sum_{i=k-d}^{k}v_i^2\bigg\}^{1/2}(d+1)^{1/2}\bigg\{\sum_{j=k-d}^{k}v_j^2\bigg\}^{1/2} = (d+1)^2\xi\,,
\end{aligned}
\tag{D.10}
$$

where the second last inequality comes from Cauchy-Schwarz's inequality.

Thus, we have

$$\begin{aligned}
&\mathbb{P}\Big\{\max_{\boldsymbol{v}^\top\boldsymbol{v}}\big|\boldsymbol{v}^\top\widehat{\boldsymbol{G}}_{n,\varphi}\boldsymbol{v}-\boldsymbol{v}^\top\boldsymbol{G}_n\boldsymbol{v}\big|<(d+1)^2\xi\Big\}\\
&\geq \mathbb{P}\left\{\left|\sum_{t=1}^{J}\sum_{s=1}^{n_t}\{\widehat{A}_{ij,k}(s,t)-A_{ij,k}(s,t)\}\right|<\xi \ \text{ for all } \ i,j=-d,\dots,K,k=0,\dots,K\right\}.
\end{aligned}$$

That is,

$$\begin{aligned}
&\mathbb{P}\{\max_{\boldsymbol{v}^\top\boldsymbol{v}}\big|\boldsymbol{v}^\top\widehat{\boldsymbol{G}}_{n,\varphi}\boldsymbol{v}-\boldsymbol{v}^\top\boldsymbol{G}_n\boldsymbol{v}\big|\geq(d+1)^2\xi\}\\
&\leq \mathbb{P}\left\{\left|\sum_{t=1}^{J}\sum_{s=1}^{n_t}\{\widehat{A}_{ij,k}(s,t)-A_{ij,k}(s,t)\}\right|\geq\xi\right.\\
&\qquad\left.\text{for at least one triplet } (i,j,k), \text{ where } \ i,j=-d,\dots,K,k=0,\dots,K\right\}.
\end{aligned}$$

Note again that if $i,j\notin[k-d,k]$, $i,j=-d,\dots,K$, $k=0,\dots,K$, then $\widehat{A}_{ij,k}(s,t)-A_{ij,k}(s,t)=0$ almost surely for all $s=1,\dots,n_t$ and $t=1,\dots,J$. Thus, almost surely, $\big|\sum_{t=1}^{J}\sum_{s=1}^{n_t}\{\widehat{A}_{ij,k}(s,t)-A_{ij,k}(s,t)\}\big|=0$ if $i,j\notin[k-d,k]$, for $i,j=-d,\dots,K$ and $k=0,\dots,K$. Then, using (D.9), we have for any $\xi>0$,

$$\begin{aligned}
&\mathbb{P}\{\max_{\boldsymbol{v}^\top\boldsymbol{v}}\big|\boldsymbol{v}^\top\widehat{\boldsymbol{G}}_{n,\varphi}\boldsymbol{v}-\boldsymbol{v}^\top\boldsymbol{G}_n\boldsymbol{v}\big|\geq(d+1)^2\xi\}\\
&\leq \mathbb{P}\left\{\left|\sum_{t=1}^{J}\sum_{s=1}^{n_t}\{\widehat{A}_{ij,k}(s,t)-A_{ij,k}(s,t)\}\right|\geq\xi\right.\\
&\qquad\left.\text{for at least one triplet } (i,j,k), \text{ where } \ i,j\in[k-d,k],k=0,\dots,K\right\}\\
&\leq \sum_{k=0}^{K}\sum_{i=k-d}^{k}\sum_{j=k-d}^{k}\mathbb{P}\left\{\left|\sum_{t=1}^{J}\sum_{s=1}^{n_t}\{\widehat{A}_{ij,k}(s,t)-A_{ij,k}(s,t)\}\right|\geq\xi\right\}\\
&< 2(d+1)^2(K+1)\exp\left\{-\frac{(1/2)n\delta\xi^2}{c_v+(1/3)\mathcal{M}\xi}\right\},
\end{aligned}$$

where $\mathcal{M} = \phi_2\phi_1^{-1} < \infty$ and $c_v = \phi\phi_2\pi_{\max}f_{\max} < \infty$.

Select $\xi = \sqrt{\tilde{c}\log K/(n\delta)}$ for some constant $0 < \tilde{c} < \infty$. Recalling that under Assumption (A6) $K\log K/n \to 0$ as $n \to \infty$, and under Assumption (A7), $\delta^{-1} \asymp K$, we have $\log K/(n\delta) \to 0$ as $n \to \infty$. Thus,

$$\begin{aligned}
&\mathbb{P}\bigg\{\max_{\boldsymbol{v}^\top\boldsymbol{v}}\big|\boldsymbol{v}^\top\widehat{\boldsymbol{G}}_{n,\varphi}\boldsymbol{v} - \boldsymbol{v}^\top\boldsymbol{G}_n\boldsymbol{v}\big| \geq (d+1)^2\sqrt{\frac{\tilde{c}\log K}{n\delta}}\bigg\}\\
&\leq 2(d+1)^2(K+1)\exp\left\{-\frac{(1/2)n\delta\tilde{c}\log K/(n\delta)}{c_v + (1/3)\mathcal{M}\sqrt{\tilde{c}\log K/(n\delta)}}\right\}\\
&= O\bigg(K\exp\bigg[-\frac{\tilde{c}\log K}{2\{c_v + (1/3)\mathcal{M}\sqrt{\tilde{c}\log K/(n\delta)}\}}\bigg]\bigg)\\
&= O\bigg(K^{1-\tilde{c}/\tilde{C}}\bigg),
\end{aligned}$$

where $\tilde{C} = 2c_v + o(1)$. Therefore, by selecting $\tilde{c} > \tilde{C}$, we have

$$\mathbb{P}\bigg\{\max_{\boldsymbol{v}^\top\boldsymbol{v}}\big|\boldsymbol{v}^\top\widehat{\boldsymbol{G}}_{n,\varphi}\boldsymbol{v} - \boldsymbol{v}^\top\boldsymbol{G}_n\boldsymbol{v}\big| \geq (d+1)^2\tilde{c}\sqrt{\frac{\log K}{n\delta}}\bigg\} = o(1)\,.$$

This implies that

$$\max_{\boldsymbol{v}^\top\boldsymbol{v}}\big|\boldsymbol{v}^\top\widehat{\boldsymbol{G}}_{n,\varphi}\boldsymbol{v} - \boldsymbol{v}^\top\boldsymbol{G}_n\boldsymbol{v}\big| = O_p\{\sqrt{\log K/(n\delta)}\}\,. \tag{D.11}$$

Note that given two functions $f, g : A \to \mathbb{R}$, where $A \in \mathbb{R}^a$ for some positive integer a, using the fact that $f = f - g + g$ and $f - g \leq |f - g|$, we have

$$\sup_A f - \sup_A g \leq \sup_A |f - g|\,.$$

Similarly,

$$\sup_A g - \sup_A f \leq \sup_A |f - g|\,.$$

Thus, $|\sup_A f - \sup_A g| \leq \sup_A |f - g|$. Replacing f by $-f$, g by $-g$ and using $\sup_A(-f) = -\inf_A f$, we have $|\inf_A f - \inf_A g| \leq \sup_A |f - g|$.

Thus, we have

$$\big|\max_{\boldsymbol{v}^\top\boldsymbol{v}} \boldsymbol{v}^\top\widehat{\boldsymbol{G}}_{n,\varphi}\boldsymbol{v} - \max_{\boldsymbol{v}^\top\boldsymbol{v}} \boldsymbol{v}^\top\boldsymbol{G}_n\boldsymbol{v}\big| \le \max_{\boldsymbol{v}^\top\boldsymbol{v}} \big|\boldsymbol{v}^\top\widehat{\boldsymbol{G}}_{n,\varphi}\boldsymbol{v} - \boldsymbol{v}^\top\boldsymbol{G}_n\boldsymbol{v}\big| = O_p\{\sqrt{\log K/(n\delta)}\}$$

and

$$\big|\min_{\boldsymbol{v}^\top\boldsymbol{v}} \boldsymbol{v}^\top\widehat{\boldsymbol{G}}_{n,\varphi}\boldsymbol{v} - \min_{\boldsymbol{v}^\top\boldsymbol{v}} \boldsymbol{v}^\top\boldsymbol{G}_n\boldsymbol{v}\big| \le \max_{\boldsymbol{v}^\top\boldsymbol{v}} \big|\boldsymbol{v}^\top\widehat{\boldsymbol{G}}_{n,\varphi}\boldsymbol{v} - \boldsymbol{v}^\top\boldsymbol{G}_n\boldsymbol{v}\big| = O_p\{\sqrt{\log K/(n\delta)}\}\,.$$

Using the assumption that $K\log K/n \to 0$ as $n \to \infty$ and $\delta^{-1} \asymp K$, $O_p\{\sqrt{\log K/(n\delta)}\} = o_p(1)$. Then combining with Lemma D.5, we have $e_{\widehat{\boldsymbol{G}},\max} = c_2\{1 + o_p(1)\}$ and $e_{\widehat{\boldsymbol{G}},\min} = c_1\{1 + o_p(1)\}$. □

Corollary D.7. *Under Assumptions (A5) to (A10),* $\lim_{n\to\infty}\mathbb{P}(\widehat{\boldsymbol{G}}_{n,\varphi}$ *is invertible*$) = 1$.

Proof. According to Lemma D.6, the minimum eigenvalue of $\widehat{\boldsymbol{G}}_{n,\varphi}$ is $c_1\{1 + o_p(1)\}$, where $c_1 > 0$. That is, for any $\xi > 0$, $\lim_{n\to\infty}\mathbb{P}(|e_{\widehat{\boldsymbol{G}},\min} - c_1| \ge \xi) = 0$. Therefore, we have for any $\xi > 0$,

$$\lim_{n\to\infty}\mathbb{P}(e_{\widehat{\boldsymbol{G}},\min} \ge c_1 + \xi \;\text{ or }\; e_{\widehat{\boldsymbol{G}},\min} \le c_1 - \xi) = 0\,.$$

Selecting $\xi = (1/2)c_1$, we have

$$\lim_{n\to\infty}\mathbb{P}(e_{\widehat{\boldsymbol{G}},\min} \le c_1/2) \le \lim_{n\to\infty}\mathbb{P}(e_{\widehat{\boldsymbol{G}},\min} \ge 3c_1/2 \;\text{ or }\; e_{\widehat{\boldsymbol{G}},\min} \le c_1/2) = 0\,.$$

This implies that

$$\lim_{n\to\infty}\mathbb{P}(\text{all the eigenvalues of } \widehat{\boldsymbol{G}}_{n,\varphi} \text{ are strictly greater than } 0) = 1\,. \tag{D.12}$$

Then using Lemma D.1 at page 32 and 0.5(m) at page 14 of Horn and Johnson (1985), we have $\lim_{n\to\infty}\mathbb{P}(\widehat{\boldsymbol{G}}_{n,\varphi}$ is invertible$) = 1$. □

In the next sequences of results, we investigate properties of the eigenvalues of $\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}$ and $\boldsymbol{H}_{n,\lambda}$. Claeskens et al. (2009) have investigated properties of the eigenvalues of the matrix $\boldsymbol{\mathcal{N}}^\top\boldsymbol{\mathcal{N}}/n + \lambda \boldsymbol{D}_\ell/n$, which is quite similar to $\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}$. However, the proof of one of their crucial results (Lemma A3 in their paper) is missing and it is not clear from their references how to prove it. They stated that their Lemma A3 was adapted from Speckman (1985) and Eubank (1999), but Eubank (1999) is in a different setting and it is not clear how it can be extended to theirs. Speckman (1985) stated the same result as Lemma A3 in Claeskens et al. (2009), but the proof is in Speckman (1981), which is unpublished and does not seem to be publicly available. The most relevant reference we found is Utreras (1983), where the matrix investigated is $\nabla_\ell \boldsymbol{R}^{-1}\nabla_\ell$, with the knots of the spline equally spaced on $[a, b]$. Since $\boldsymbol{R}$ is a bounded and bounded invertible matrix, this result is a useful guide for studying $\boldsymbol{D}_\ell = \nabla_\ell \boldsymbol{R}\nabla_\ell$.

Due to the difference between the knots setting of Utreras (1983) and our setting mentioned above, we cannot extend their results to $\boldsymbol{D}_\ell$ directly. However, we can use their idea of bounding the eigenvalues of a product of matrices by the product of the eigenvalues of the matrices, and some of their results concerning the eigenvalues of ∇_ℓ, to study $\boldsymbol{D}_\ell$. In order to achieve this, Lemmas D.8 to D.10 and Corollary D.11 will be useful.

Lemma D.8 (Horn and Johnson, 1985, page 53)**.** *Suppose that $\boldsymbol{A} \in \mathbb{R}^{k\times r}$ and $\boldsymbol{B} \in \mathbb{R}^{r\times k}$ where $k \le r$ are two positive integers. Then all the eigenvalues of $\boldsymbol{AB}$ are the eigenvalues of $\boldsymbol{BA}$ with the same multiplicity, and the remaining $(r-k)$ eigenvalues of $\boldsymbol{BA}$ are all equal to 0.*

Lemma D.9 (Horn and Johnson, 1985, page 181)**.** *For any symmetric matrices $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{r\times r}$ with r a positive integer, let $e_{\boldsymbol{A},1} \le \ldots \le e_{\boldsymbol{A},r}$, $e_{\boldsymbol{B},1} \le \ldots \le e_{\boldsymbol{B},r}$ and $e_{\boldsymbol{A}+\boldsymbol{B},1} \le \ldots \le e_{\boldsymbol{A}+\boldsymbol{B},r}$ be the eigenvalues of $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{A}+\boldsymbol{B}$ in increasing order, respectively. For $i = 1, \ldots, r$, we have*

$$e_{\boldsymbol{A},i} + e_{\boldsymbol{B},1} \le e_{\boldsymbol{A}+\boldsymbol{B},i} \le e_{\boldsymbol{A},i} + e_{\boldsymbol{B},r}\,.$$

Lemma D.10 (Zhou et al., 1998, Lemma 6.5). *For any positive integer r, if $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{r\times r}$ are non-negative definite matrices in the sense that, for any vector $\boldsymbol{v} \in \mathbb{R}^r$, $\boldsymbol{v}^\top \boldsymbol{A}\boldsymbol{v} \geq 0$ and $\boldsymbol{v}^\top \boldsymbol{B}\boldsymbol{v} \geq 0$, then*

$$e_{\boldsymbol{A},\min} Tr(\boldsymbol{B}) \leq Tr(\boldsymbol{A}\boldsymbol{B}) \leq e_{\boldsymbol{A},\max} Tr(\boldsymbol{B}),$$

where Tr denotes the trace and $e_{\boldsymbol{A},\min}$ and $e_{\boldsymbol{A},\max}$ are the minimum and maximum eigenvalues of $\boldsymbol{A}$, respectively.

Corollary D.11. *If $\boldsymbol{A} \in \mathbb{R}^{r\times r}$ is a non-negative definite symmetric matrix, then for any $\boldsymbol{v} \in \mathbb{R}^r$, we have*

$$e_{\boldsymbol{A},\min} \boldsymbol{v}^\top \boldsymbol{v} \leq \boldsymbol{v}^\top \boldsymbol{A}\boldsymbol{v} \leq e_{\boldsymbol{A},\max} \boldsymbol{v}^\top \boldsymbol{v},$$

where $e_{\boldsymbol{A},\min}$ and $e_{\boldsymbol{A},\max}$ are the minimum and maximum eigenvalues of $\boldsymbol{A}$, respectively.

Next, we investigate the eigenvalues of $\boldsymbol{D}_\ell$.

Lemma D.12 (Eigenvalues of $\boldsymbol{D}_\ell$). *Let $e_{\boldsymbol{D},1} \leq e_{\boldsymbol{D},2} \leq \ldots \leq e_{\boldsymbol{D},K+d+1}$ be the eigenvalues of $\boldsymbol{D}_\ell$. Under Assumptions (A3), (A6) and (A7),*

$$e_{\boldsymbol{D},1} = \ldots = e_{\boldsymbol{D},\ell} = 0 \tag{D.13}$$

and $c_3\delta\{1 + o(1)\} \leq e_{\boldsymbol{D},i} \leq c_4\delta^{-2\ell+1}$, for $i = \ell + 1, \ldots, K + d + 1$, where c_3 and c_4 are two positive finite constants that do not depend on i and n.

Proof. Recall from (E.8) that $\boldsymbol{D}_\ell = \nabla_l^\top \boldsymbol{R} \nabla_l$, with ∇_l and $\boldsymbol{R}$ defined at page 84. We start by investigating the eigenvalues of $\boldsymbol{R}$.

Let $e_{\boldsymbol{R},\max}$ and $e_{\boldsymbol{R},\min}$ be the maximum and minimum eigenvalues of $\boldsymbol{R}$, respectively. By definition, $e_{\boldsymbol{R},\max} = \max_{\boldsymbol{v}^\top\boldsymbol{v}=1} \boldsymbol{v}^\top \boldsymbol{R}\boldsymbol{v}$ and $e_{\boldsymbol{R},\min} = \min_{\boldsymbol{v}^\top\boldsymbol{v}=1} \boldsymbol{v}^\top \boldsymbol{R}\boldsymbol{v}$, where $\boldsymbol{v}$ denotes a vector of length $K+d+1-\ell$. Recalling from page 84 that $(\boldsymbol{R})_{ij} = \int_a^b N_{i,d+1-\ell}(x) N_{j,d+1-\ell}(x)\, dx$

for $i, j = -d+\ell, \ldots, K$, we have

$$\boldsymbol{v}^\top \boldsymbol{R} \boldsymbol{v} = \int_a^b \left\{ \sum_{i=-d+\ell}^{K} v_i N_{i,d+1-\ell}(x) \right\}^2 dx \,,$$

for any $\boldsymbol{v} \in \mathbb{R}^{K+d+1-\ell}$.

Note that under Assumption (A3) which states that $d-\ell > 0$, $\sum_{i=-d+\ell}^{K} v_i N_{i,d+1-\ell}(x) \in S_{d-\ell}(\boldsymbol{t})$. Then by Lemma E.1, taking $q = 2$ and replacing there d by $d-\ell$, we have

$$(d+1-\ell)^{-1} C_{d-\ell}^2 \sum_{i=-d+\ell}^{K} v_i^2 (t_{i+d+1-\ell} - t_i) \leq \boldsymbol{v}^\top \boldsymbol{R} \boldsymbol{v} \leq (d+1-\ell)^{-1} \sum_{i=-d+\ell}^{K} v_i^2 (t_{i+d+1-\ell} - t_i) \,.$$

Therefore using Assumption (A7),

$$(d+1-\ell)^{-1} C_{d-\ell}^2 M^{-1} \delta \sum_{i=-d+\ell}^{K} v_i^2 \leq \boldsymbol{v}^\top \boldsymbol{R} \boldsymbol{v} \leq (d+1-\ell)^{-1} \delta \sum_{i=-d+\ell}^{K} v_i^2 \,.$$

Thus,

$$\delta C_{d-\ell}^2 (d+1-\ell)^{-1} M^{-1} \leq e_{\boldsymbol{R},\min} \leq e_{\boldsymbol{R},\max} \leq \delta (d+1-\ell)^{-1} \,. \tag{D.14}$$

Now for any vector $\boldsymbol{u} \in \mathbb{R}^{K+d+1}$, applying Corollary D.11 with $\boldsymbol{v} = \nabla_l \boldsymbol{u}$, we have

$$\frac{C_{d-\ell}^2 \delta}{M(d+1-\ell)} \boldsymbol{u}^\top \nabla_\ell^\top \nabla_\ell \boldsymbol{u} \leq \boldsymbol{u}^\top \nabla_\ell^\top \boldsymbol{R} \nabla_\ell \boldsymbol{u} \leq \frac{\delta}{d+1-\ell} \boldsymbol{u}^\top \nabla_\ell^\top \nabla_\ell \boldsymbol{u} \,,$$

which implies that $\nabla_\ell^\top \boldsymbol{R} \nabla_\ell - \delta C_{d-\ell}^2 (d+1-\ell)^{-1} M^{-1} \nabla_\ell^\top \nabla_\ell$ and $\delta (d+1-\ell)^{-1} \nabla_\ell^\top \nabla_\ell - \nabla_\ell^\top \boldsymbol{R} \nabla_\ell$ are non-negative definite matrices, so that their smallest eigenvalues are larger than or equal to 0.

Let $e_{\ell,1} \leq \ldots \leq e_{\ell,K+d+1}$ be the eigenvalues of $\nabla_\ell^\top \nabla_\ell$. Applying Lemma D.9 with $\boldsymbol{A} = \delta C_{d-\ell}^2 (d+1-\ell)^{-1} M^{-1} \nabla_\ell^\top \nabla_\ell$ and $\boldsymbol{B} = \nabla_\ell^\top \boldsymbol{R} \nabla_\ell - \delta C_{d-\ell}^2 (d+1-\ell)^{-1} M^{-1} \nabla_\ell^\top \nabla_\ell$, so that $\boldsymbol{A} + \boldsymbol{B} = \nabla_\ell^\top \boldsymbol{R} \nabla_\ell = \boldsymbol{D}_\ell$, we have

$$\delta C_{d-\ell}^2 (d+1-\ell)^{-1} M^{-1} e_{\ell,i} + 0 \leq e_{\boldsymbol{D},i} \,.$$

Then applying Lemma D.9 again by letting $\boldsymbol{A} = \nabla_\ell^\top \boldsymbol{R} \nabla_\ell = \boldsymbol{D}_\ell$ and $\boldsymbol{B} = \delta(d+1-\ell)^{-1}\nabla_\ell^\top \nabla_\ell - \nabla_\ell^\top \boldsymbol{R}\nabla_\ell$, we have

$$e_{\boldsymbol{D},i} + 0 \leq \delta(d+1-\ell)^{-1} e_{\ell,i}\,.$$

Therefore, for $i = 1, \ldots, K+d+1$,

$$\delta C_{d-\ell}^2 (d+1-\ell)^{-1} M^{-1} e_{\ell,i} \leq e_{\boldsymbol{D},i} \leq \delta(d+1-\ell)^{-1} e_{\ell,i}\,. \tag{D.15}$$

Now we investigate the eigenvalues $e_{\ell,i}$ of $\nabla_\ell^\top \nabla_\ell$. Note that ∇_ℓ is a matrix of dimension $(K+d+1-\ell) \times (K+d+1)$. Lemma D.8 implies that the eigenvalues of $\nabla_\ell^\top \nabla_\ell$ are those of $\nabla_\ell \nabla_\ell^\top$, plus an additional ℓ eigenvalues equal to 0. Since for any $\boldsymbol{v} \in \mathbb{R}^{K+d+1}$, one can see that $\boldsymbol{v}^\top \boldsymbol{D}_\ell \boldsymbol{v} = \boldsymbol{v}^\top \nabla_\ell^\top \boldsymbol{R} \nabla_\ell \boldsymbol{v} \geq 0$, $\boldsymbol{D}_\ell$ is non-negative definite. Thus, (D.13) follows and it remains to study the eigenvalues of $\nabla_\ell \nabla_\ell^\top$.

To do this we start by expressing $\nabla_\ell \nabla_\ell^\top$ in a way that will make our analysis simpler. It follows from the definition of ∇_ℓ under B-spline Property 3 of page 84 that

$$\begin{aligned}
(\nabla_\ell \boldsymbol{\beta})_i &= (d+1-\ell)\frac{(\nabla_{\ell-1}\boldsymbol{\beta})_i - (\nabla_{\ell-1}\boldsymbol{\beta})_{i-1}}{t_{i+d+1-\ell} - t_i}\,, \quad i = -d+\ell, \ldots, K\,,\\
&\vdots\\
(\nabla_k \boldsymbol{\beta})_i &= (d+1-k)\frac{(\nabla_{k-1}\boldsymbol{\beta})_i - (\nabla_{k-1}\boldsymbol{\beta})_{i-1}}{t_{i+d+1-k} - t_i}\,, \quad i = -d+k, \ldots, K\,,\\
&\vdots\\
(\nabla_1 \boldsymbol{\beta})_i &= d\frac{(\nabla_0\boldsymbol{\beta})_i - (\nabla_0\boldsymbol{\beta})_{i-1}}{t_{i+d} - t_i}\,, \quad i = -d+1, \ldots, K\,,\\
\nabla_0 \boldsymbol{\beta} &= \boldsymbol{\beta}\,.
\end{aligned}$$

Recall that $\delta_i = t_{i+1} - t_i$ and under Assumption (A7), $M^{-1}\delta \leq \delta_i \leq \delta$ for some constant $1 \leq M < \infty$, $i = 0, \ldots, K$. Moreover, $\delta_i = 0$ for $i = -d, \ldots, -1$. Thus, $t_{i+d+1-k} - t_i = \sum_{j=i}^{i+d-k} \delta_j \in [M^{-1}\delta, (d+1-k)\delta]$, for $i = -d+k.\ldots, K$, where $k = 1, \ldots, \ell$. Then for each for $i = -d+k, \ldots, K$ and $k = 1, \ldots, \ell$, we can find a constant $M^{-1} \le c_{i,k} \le d+1-k$ such that $t_{i+d+1-k} - t_i = c_{i,k}\delta$. For $k = 1, \ldots, \ell$, let ∇^k be a $(K+d+1-k) \times \{K+d+1-(k-1)\}$ matrix, such that for $i = -d+k, \ldots, K$, $j = -d+k, \ldots, K+1$,

$$(\nabla^k)_{ij} = \begin{cases} -\frac{d+1-k}{c_{i,k}}\delta^{-1}\,, & j = i\,, \\ \frac{d+1-k}{c_{i,k}}\delta^{-1}\,, & j = i+1\,, \\ 0\,, & \text{otherwise}\,. \end{cases}$$

Then we can see that $\nabla_\ell = \nabla^\ell \nabla^{\ell-1} \ldots \nabla^1$.

Now define a set of matrices $\boldsymbol{W}_k$ of dimension $(K+d+1-k) \times \{K+d+1-(k-1)\}$ such that for $k = 1, \ldots, \ell$, and for $i = -d+k, \ldots, K$, $j = -d+k, \ldots, K+1$,

$$(\boldsymbol{W}_k)_{ij} = \begin{cases} -1\,, & j = i\,, \\ 1\,, & j = i+1\,, \\ 0\,, & \text{otherwise}\,. \end{cases} \tag{D.16}$$

Then we have

$$\nabla^k = \delta^{-1}\boldsymbol{T}_k \boldsymbol{W}_k\,, \tag{D.17}$$

where $\boldsymbol{T}_k$ is a $(K+d+1-k) \times (K+d+1-k)$ diagonal matrix with

$$(\boldsymbol{T}_k)_{ii} = \frac{d+1-k}{c_{i,k}}\,, \quad i = -d+k, \ldots, K\,, \tag{D.18}$$

which are also the eigenvalues of $\boldsymbol{T}_k$. Then we have $\nabla_\ell = \nabla^\ell \nabla^{\ell-1} \ldots \nabla^1 = \delta^{-l}\boldsymbol{T}_\ell \boldsymbol{W}_\ell \ldots \boldsymbol{T}_1 \boldsymbol{W}_1$, and

$$\nabla_\ell \nabla_\ell^\top = \delta^{-2\ell}\boldsymbol{T}_\ell \boldsymbol{W}_\ell \ldots \boldsymbol{T}_1 \boldsymbol{W}_1 \boldsymbol{W}_1^\top \boldsymbol{T}_1^\top \ldots \boldsymbol{W}_\ell^\top \boldsymbol{T}_\ell^\top\,.$$

Note that for any $k = 1, \ldots, \ell$,

$$\boldsymbol{W}_k \boldsymbol{W}_k^\top = \begin{pmatrix} 2 & -1 & & & & \\ -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & -1 & 2 & -1 \\ & & & & -1 & 2 \end{pmatrix}$$

is a $(K+d+1-k)\times(K+d+1-k)$ tridiagonal matrix. Let $e_{W_k,1}\le\ldots\le e_{W_k,K+d+1-k}$ be the eigenvalues of $\boldsymbol{W}_k\boldsymbol{W}_k^\top$ in increasing order. Then it is known that $e_{W_k,i}=2-2\cos\{(i\pi)/(K+d+2-k)\}$, for $i=1,\ldots,K+d+1-k$; see Theorem 2.2 of Kulkarni et al. (1999) for example. As shown in Utreras (1983) (see (32) and (33) in that paper), by Taylor expansion, for all $1\le i\le K+d+1-k$ and $k=1,\ldots,\ell$,

$$c_5\frac{i^2}{(K+d+2-k)^2}\le 2-2\cos\left(\frac{i\pi}{K+d+2-k}\right)\le c_6\frac{i^2}{(K+d+2-k)^2}\,,$$

where $c_5=(1-\pi^2/12)\pi^2$ and $c_6=(1+\pi^4/3)\pi^2$. Thus, for all k,

$$c_5\frac{1}{(K+d+1)^2}\le e_{W_k,1}\le e_{W_k,K+d+1-k}\le c_6\frac{(K+d+1-k)^2}{(K+d+2-k)^2}\le c_6\,.$$

Since for vector $\boldsymbol{v}\in\mathbb{R}^{K+d+1-\ell}$,

$$\boldsymbol{v}^\top\nabla_\ell\nabla_\ell^\top\boldsymbol{v} \;=\; \delta^{-2\ell}\boldsymbol{v}^\top\boldsymbol{T}_\ell\boldsymbol{W}_\ell\ldots\boldsymbol{T}_1\boldsymbol{W}_1\boldsymbol{W}_1^\top\boldsymbol{T}_1^\top\ldots\boldsymbol{W}_\ell^\top\boldsymbol{T}_\ell^\top\boldsymbol{v}\,,$$

treating $\boldsymbol{T}_1^\top\ldots\boldsymbol{W}_\ell^\top\boldsymbol{T}_\ell^\top\boldsymbol{v}$ as a vector, and using Corollary D.11, we have

$$\begin{aligned}\boldsymbol{v}^\top\nabla_\ell\nabla_\ell^\top\boldsymbol{v} &\ge \delta^{-2\ell}c_5\frac{1}{(K+d+1)^2}\boldsymbol{v}^\top\boldsymbol{T}_\ell\boldsymbol{W}_\ell\ldots\boldsymbol{T}_1\boldsymbol{T}_1^\top\ldots\boldsymbol{W}_\ell^\top\boldsymbol{T}_\ell^\top\boldsymbol{v}\,,\\ \boldsymbol{v}^\top\nabla_\ell\nabla_\ell^\top\boldsymbol{v} &\le \delta^{-2\ell}c_6\boldsymbol{v}^\top\boldsymbol{T}_\ell\boldsymbol{W}_\ell\ldots\boldsymbol{T}_1\boldsymbol{T}_1^\top\ldots\boldsymbol{W}_\ell^\top\boldsymbol{T}_\ell^\top\boldsymbol{v}\,.\end{aligned}$$

Since $\boldsymbol{T}_1$ is a diagonal matrix, the eigenvalues of $\boldsymbol{T}_1\boldsymbol{T}_1^\top$ are equal to the square of the diagonal terms of $\boldsymbol{T}_1$. Thus, by Corollary D.11 again, and noting that $M^{-1}\le c_{i,k}\le d+1-k$ for all $k=1,\ldots,\ell$, we have, using (D.18),

$$\begin{aligned}&\boldsymbol{v}^\top\boldsymbol{T}_\ell\boldsymbol{W}_\ell\ldots\boldsymbol{T}_2\boldsymbol{W}_2\boldsymbol{W}_2^\top\boldsymbol{T}_2^\top\ldots\boldsymbol{W}_\ell^\top\boldsymbol{T}_\ell^\top\boldsymbol{v}\\ &\le \boldsymbol{v}^\top\boldsymbol{T}_\ell\boldsymbol{W}_\ell\ldots\boldsymbol{W}_2\boldsymbol{T}_1\boldsymbol{T}_1^\top\boldsymbol{W}_2^\top\ldots\boldsymbol{W}_\ell^\top\boldsymbol{T}_\ell^\top\boldsymbol{v}\\ &\le M^2d^2\boldsymbol{v}^\top\boldsymbol{T}_\ell\boldsymbol{W}_\ell\ldots\boldsymbol{T}_2\boldsymbol{W}_2\boldsymbol{W}_2^\top\boldsymbol{T}_2^\top\ldots\boldsymbol{W}_\ell^\top\boldsymbol{T}_\ell^\top\boldsymbol{v}\,.\end{aligned}$$

Repeating this procedure multiple times, we obtain

$$\boldsymbol{v}^\top \nabla_\ell \nabla_\ell^\top \boldsymbol{v} \geq \delta^{-2\ell} c_5^\ell \frac{1}{(K+d+1)^{2\ell}} \boldsymbol{v}^\top \boldsymbol{v}\,,$$
$$\boldsymbol{v}^\top \nabla_\ell \nabla_\ell^\top \boldsymbol{v} \leq \delta^{-2\ell} c_6^\ell M^{2\ell} d^{2\ell} \boldsymbol{v}^\top \boldsymbol{v}\,.$$

Thus, by taking the minimum and maximum over $\boldsymbol{v}^\top \boldsymbol{v} = 1$ on both sides of the above inequalities, respectively, we have for all $i = \ell+1, \ldots, K+d+1$,

$$c_7 \delta^{-2\ell} \frac{1}{(K+d+1)^{2\ell}} \leq e_{\ell,i} \leq c_8 \delta^{-2\ell}\,,$$

where $c_7 = c_5^\ell$ and $c_8 = c_6^\ell M^{2\ell} d^{2\ell}$.

Under Assumption (A6), we have $d/(K+1) = o(1)$ as $n \to \infty$. Using a Taylor expansion of first order to $\{1+d/(K+1)\}^{-2\ell}$ around 0, we have $\{1+d/(K+1)\}^{-2\ell} = 1 - 2\ell\{d/(K+1)\} + o\{d/(K+1)\} = 1 + o(1)$. Thus, $(K+d+1)^{-2\ell} = (K+1)^{-2\ell}\{1+d/(K+1)\}^{-2\ell} = (K+1)^{-2\ell}\{1+o(1)\}$. Using Assumption (A7), we have $(K+1)M^{-1}\delta \leq \sum_{i=0}^{K} \delta_i = b-a \leq (K+1)\delta$, i.e. $(b-a)/\delta \leq (K+1) \leq (b-a)M/\delta$. Thus, $(K+d+1)^{-2\ell} = c\delta^{2\ell}\{1+o(1)\}$ for some constant $\{(b-a)M\}^{-1} \leq c \leq (b-a)^{-1}$. Then, we have for all $i = \ell+1, \ldots, K+d+1$,

$$c_7 c\{1+o(1)\} \leq e_{\ell,i} \leq c_8 \delta^{-2\ell}\,.$$

Now combining with (D.15), we deduce that for $i = \ell+1, \ldots, K+d+1$,

$$\delta C_{d-\ell}^2 c_7 c (d+1-\ell)^{-1} M^{-1}\{1+o(1)\} \leq e_{\boldsymbol{D},i} \leq \delta^{-2\ell+1} c_8 (d+1-\ell)^{-1}\,.$$

Let $c_3 = c_7 c(d+1-\ell)^{-1} C_{d-\ell}^2 M^{-1}$, $c_4 = c_8(d+1-\ell)^{-1}$. Combining this with the result that $e_{\boldsymbol{D},1} = \ldots = e_{\boldsymbol{D},\ell} = 0$ we have proved at page 47, the result follows. $\square$

Now we use the eigenvalues of $\boldsymbol{D}_\ell$ to study those of $\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}$ and $\boldsymbol{H}_{n,\lambda}$. In what follows we will repeatedly use the quantity $(c_4 \lambda \delta^{-2\ell})/n$, where $c_4 = (1+\pi^4/3)\pi^2(Md)^{2\ell}(d+1-\ell)^{-1}$

was introduced in Lemma D.12 with M defined in Assumption (A7), and $\lambda > 0$ is the smoothing parameter. Throughout, we will use the notation

$$K_{n,\lambda} = c_4 \lambda \delta^{-2\ell}/n \,. \tag{D.19}$$

Lemma D.13 (Eigenvalues of $\boldsymbol{H}_{n,\lambda}$)**.** *Let $e_{\boldsymbol{H},\min}$ and $e_{\boldsymbol{H},\max}$ be the minimum and maximum eigenvalues of $\boldsymbol{H}_{n,\lambda}$, respectively. Under Assumptions (A6) to (A10),*

$$c_1 \leq e_{\boldsymbol{H},\min} \leq c_2 \tag{D.20}$$

and

$$c_2 \leq e_{\boldsymbol{H},\max} \leq c_2(1 + K_{n,\lambda}/c_2) \,, \tag{D.21}$$

where c_1, c_2 are defined as in Lemma D.5 and $K_{n,\lambda}$ was introduced at (D.19)*.*

Proof. By definition at (4.6), $\boldsymbol{H}_{n,\lambda} = \boldsymbol{G}_n + \lambda \boldsymbol{D}_l/(n\delta)$. From Lemma D.5, $e_{\boldsymbol{G},\min} = c_1$ and $e_{\boldsymbol{G},\max} = c_2$. Then applying Lemmas D.9 and D.12, we have $c_1 \leq e_{\boldsymbol{H},\min} \leq c_2$ and

$$c_2 \leq e_{\boldsymbol{H},\max} \leq c_2 + \frac{\lambda}{n} c_4 \delta^{-2\ell} = c_2(1 + K_{n,\lambda}/c_2) \,.$$

□

Lemma D.14 (Eigenvalues of $\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}$)**.** *Let $e_{\widehat{\boldsymbol{H}},\min}$ and $e_{\widehat{\boldsymbol{H}},\max}$ be the minimum and maximum eigenvalues of $\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}$, respectively. Under Assumptions (A5) to (A10),*

$$e_{\widehat{\boldsymbol{H}},\min} = \tilde{c}_1\{1 + o_p(1)\} \,, \tag{D.22}$$

and

$$e_{\widehat{\boldsymbol{H}},\max} = \tilde{c}_{2,n,\lambda}\{1 + o_p(1)\} \,, \tag{D.23}$$

where $c_1 \leq \tilde{c}_1 \leq c_2$ with c_1, c_2 defined as in Lemma D.5 and $c_2 \leq \tilde{c}_{2,n,\lambda} \leq c_2(1 + K_{n,\lambda}/c_2)$ with $K_{n,\lambda}$ defined at (D.19)*.*

Proof. By definition at page 21, $\widehat{\boldsymbol{H}}_{n,\lambda,\varphi} = \widehat{\boldsymbol{G}}_{n,\varphi} + \lambda \boldsymbol{D}_\ell/(n\delta)$ and $\boldsymbol{H}_{n,\lambda} = \boldsymbol{G}_n + \lambda \boldsymbol{D}_\ell/(n\delta)$. Thus, $\widehat{\boldsymbol{H}}_{n,\lambda,\varphi} - \boldsymbol{H}_{n,\lambda} = \widehat{\boldsymbol{G}}_{n,\varphi} - \boldsymbol{G}$. Then using the same arguments as in the proof of Lemma D.6, we have $e_{\widehat{\boldsymbol{H}},min} = e_{\boldsymbol{H},\min} + o_p(1)$ and $e_{\widehat{\boldsymbol{H}},\max} = e_{\boldsymbol{H},\max} + o_p(1)$. Combining with Lemma D.13, we deduce that $e_{\widehat{\boldsymbol{H}},\min} = \tilde{c}_1\{1 + o_p(1)\}$, where $c_1 \leq \tilde{c}_1 \leq c_2$. Note that $c_2 + o_p(1) = c_2\{1 + o_p(1)\}$. Letting $a_n = e_{\widehat{\boldsymbol{H}},\max} - e_{\boldsymbol{H},\max}$, we have $a_n = o_p(1)$ and

$$c_2(1 + K_{n,\lambda}/c_2) + a_n = c_2(1 + K_{n,\lambda}/c_2)\left(1 + \frac{a_n}{c_2 + K_{n,\lambda}}\right).$$

Now since $\inf_{n\in\mathbb{N}} K_{n,\lambda} \geq 0$ and $c_2 > 0$, we have $|a_n/(c_2 + K_{n,\lambda})| \leq |a_n|/c_2 = o_p(1)$. Thus, $c_2(1 + K_{n,\lambda}/c_2) + a_n = c_2(1 + K_{n,\lambda}/c_2)\{1 + o_p(1)\}$. The result then follows. □

Using arguments similar to those in Corollary D.7, we have the following corollary.

Corollary D.15. *Under Assumptions (A5) to (A10),* $\lim_{n\to\infty} \mathbb{P}(\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}$ *is invertible*$) = 1$.

Furthermore, some properties of the inverse of the matrices $\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}$ and $\boldsymbol{H}_{n,\lambda}$ are also needed to study the asymptotic properties of our estimator.

Lemma A1 of Claeskens et al. (2009) provides upper bounds for the elements of the matrix $(\boldsymbol{\mathcal{N}}^\top\boldsymbol{\mathcal{N}}/n + \lambda\boldsymbol{D}_\ell/n)^{-1}$, which is similar to our $\widehat{\boldsymbol{H}}^{-1}_{n,\lambda,\varphi}$. However, it is not clear that this lemma is valid without additional conditions. First, the spectral norm of the inverse of their matrix is $h_{\max}/h_{\min}$, where $h_{\max}$ and $h_{\min}$ are the maximum and minimum eigenvalues of $\boldsymbol{\mathcal{N}}^\top\boldsymbol{\mathcal{N}}/n + \lambda\boldsymbol{D}_\ell/n$, respectively. Using their arguments, when $K_q \geq 1$, $h_{\max}/h_{\min} \geq 1 + K_q^{2q}$, where K_q depends on λ, K, n. Since they did not assume that K_q is always finite, this term can go to infinity as n goes to infinity and the condition in Theorem 2.2 of Demko (1977) which they used is not satisfied. Moreover, in the statement of the lemma, the authors claimed that the constant c_0 involved in the upper bounds is independent of K and n. However, it appears from their proof that c_0 depends on K_q.

To find how λ, K and n influence the upper bounds of the elements of $\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}^{-1}$ and $\boldsymbol{H}_{n,\lambda}^{-1}$, we need to know the explicit form of the constant in Theorem 2.2 of Demko (1977), which Demko et al. (1984) have derived. We first present their Proposition 2.2 and then show a corollary of it.

Proposition D.16 (Proposition 2.2 of Demko et al. (1984))**.** *Let* $\boldsymbol{A} \in \mathbb{R}^{b\times b}$ *be a positive definite,* $2m$*-banded matrix, where* b *and* $m < b-1$ *are positive integers. Suppose it is bounded and boundedly invertible in the sense that, for all* $\boldsymbol{v} \in \mathbb{R}^b$, $\boldsymbol{v}^\top\boldsymbol{A}^\top\boldsymbol{A}\boldsymbol{v} \le \mathrm{const}_1\boldsymbol{v}^\top\boldsymbol{v}$ *and* $\boldsymbol{v}(\boldsymbol{A}^{-1})^\top\boldsymbol{A}^{-1}\boldsymbol{v} \le \mathrm{const}_2\boldsymbol{v}^\top\boldsymbol{v}$, *for some constants* $\mathrm{const}_1, \mathrm{const}_2 > 0$. *Suppose that* $[r,s]$ *is the smallest interval that contains all the eigenvalues of* $\boldsymbol{A}$. *Letting* $\gamma = s/r < \infty$, *we have*

$$|(\boldsymbol{A}^{-1})_{ij}| \le \max(r^{-1}, C_{r,\gamma})\left(\frac{\sqrt{\gamma}-1}{\sqrt{\gamma}+1}\right)^{|i-j|/m},$$

for all $i,j = 1,\ldots,b$ *where* $C_{r,\gamma} = (1+\gamma^{1/2})^2/(2r\gamma)$.

Corollary D.17 (Inverse band matrix)**.** *Let* $\boldsymbol{A} \in \mathbb{R}^{b\times b}$ *be a symmetric, centered* $2m-$*banded matrix, where* b *and* $m < b-1$ *are positive integers. Denote by* $[\tilde{r},\tilde{s}]$ *an interval that contains all the eigenvalues of* $\boldsymbol{A}$. *If* $0 < \tilde{r} \le \tilde{s} < \infty$, *letting* $\tilde{\gamma} = \tilde{s}/\tilde{r}$ *and* $\mu = \{(\sqrt{\tilde{\gamma}}-1)/(\sqrt{\tilde{\gamma}}+1)\}^{1/m}$, *we have that* $\boldsymbol{A}$ *is invertible, and for all* $i,j = 1,\ldots,b$,

$$|(\boldsymbol{A}^{-1})_{ij}| \le \frac{2}{\tilde{r}}\mu^{|i-j|}, \tag{D.24}$$

where $(\boldsymbol{A}^{-1})_{ij}$ *denotes the* (i,j)*th entry of* $\boldsymbol{A}^{-1}$ *and* $0 < \mu < 1$.

Proof. Let $[r,s]$ be the smallest interval that contains all the eigenvalues of $\boldsymbol{A}$. Then we have $\tilde{r} \le r \le s \le \tilde{s}$. We first apply Proposition D.16 using $[r,s]$.

To apply the proposition, we only need to show that $\boldsymbol{A}$ is positive definite, bounded and boundedly invertible. Since the interval $[r,s]$ contains all the eigenvalues of $\boldsymbol{A}$ and $0 < r \le s < \infty$, $\boldsymbol{A}$ is a positive definite matrix and invertible. Note that since $\boldsymbol{A}$

is symmetric, $\boldsymbol{v}^\top \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{v} = \boldsymbol{v}^\top \boldsymbol{A}\boldsymbol{A}\boldsymbol{v}$ and $\boldsymbol{v}^\top (\boldsymbol{A}^{-1})^\top \boldsymbol{A}^{-1}\boldsymbol{v} = \boldsymbol{v}^\top \boldsymbol{A}^{-1}\boldsymbol{A}^{-1}\boldsymbol{v}$. Moreover, by Corollary D.11, we have for all $\boldsymbol{v} \in \mathbb{R}^b$, $\boldsymbol{v}^\top \boldsymbol{A}\boldsymbol{A}\boldsymbol{v} \le s^2 \boldsymbol{v}^\top \boldsymbol{v}$ and $\boldsymbol{v}^\top \boldsymbol{A}^{-1}\boldsymbol{A}^{-1}\boldsymbol{v} \le r^{-2}\boldsymbol{v}^\top \boldsymbol{v}$. Then given that $0 < \tilde{r} \le r \le s \le \tilde{s} < \infty$, $\boldsymbol{A}$ is a bounded and boundedly invertible matrix. Thus, we can apply Proposition D.16 to obtain that

$$|(\boldsymbol{A}^{-1})_{ij}| \le \max(r^{-1}, C_{r,\gamma}) \left(\frac{\sqrt{\gamma}-1}{\sqrt{\gamma}+1} \right)^{|i-j|/m},$$

for $i, j = 1, \ldots, b$ where $C_{r,\gamma} = (1+\gamma^{1/2})^2/(2r\gamma)$.

Now by definition of γ, we have

$$\begin{aligned} \frac{(1+\gamma^{1/2})^2}{2r\gamma} &= \frac{1+2\gamma^{1/2}+\gamma}{2r\gamma} \\ &= \frac{1}{2s} + \frac{1}{\sqrt{sr}} + \frac{1}{2r} \\ &\le \frac{2}{r}. \end{aligned}$$

Then we have

$$|(\boldsymbol{A}^{-1})_{ij}| \le \frac{2}{r} \left(\frac{\sqrt{\gamma}-1}{\sqrt{\gamma}+1} \right)^{|i-j|/m},$$

for $i, j = 1, \ldots, b$. Note that $0 < \tilde{r} \le r \le s \le \tilde{s} < \infty$, and we have $2/r \le 2/\tilde{r}$ and $\gamma \le \tilde{\gamma}$. Since the function $(x-1)/(x+1)$ is non-decreasing for $0 < x < \infty$, we have

$$\frac{\sqrt{\gamma}-1}{\sqrt{\gamma}+1} \le \frac{\sqrt{\tilde{\gamma}}-1}{\sqrt{\tilde{\gamma}}+1}.$$

Then the result follows. □

Next we will apply Proposition D.16 to study the inverse of $\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}$ and $\boldsymbol{H}_{n,\lambda}$. There, we find that γ depends on n. Thus, to apply Proposition D.16, we need to have that $\sup_{n\in\mathbb{N}} \gamma < \infty$. By Lemmas D.13 and D.14, we see that in order to have $\sup_{n\in\mathbb{N}} \gamma < \infty$, we need to assume that

$$0 \le \inf_{n\in\mathbb{N}} K_{n,\lambda} \le \sup_{n\in\mathbb{N}} K_{n,\lambda} < \infty\,. \tag{D.25}$$

The details are provided in Lemmas D.18 and D.19. Recall from (D.19) that $K_{n,\lambda} = c_4\lambda\delta^{-2\ell}/n$, and c_4 is a finite positive constant independent of n. Using Assumption (A7), we have $\delta^{-1} \asymp K$. Then we see that (D.25) is equivalent to $0 \le \inf_{n\in\mathbb{N}} \lambda K^{2\ell}/n \le \sup_{n\in\mathbb{N}} \lambda K^{2\ell}/n < \infty$. Such an assumption is made in (A11), and thus we use Assumption (A11) when applying Lemma D.16 to $\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}$ and $\boldsymbol{H}_{n,\lambda}$.

Lemma D.18. *Suppose Assumptions (A5) to (A11) hold. Define the event* $\mathcal{E}_n = \{[(1/2)c_1, c_2(2 + K_{n,\lambda}/c_2)]$ *is an interval that contains all the eigenvalues of* $\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}\}$ *and* $\nu = [\{\sqrt{2c_2(2+K_{n,\lambda}/c_2)/c_1} - 1\}/\{\sqrt{2c_2(2+K_{n,\lambda}/c_2)/c_1} + 1\}]^{1/(d+1)}$, *where* c_1 *and* c_2 *are defined in Lemma D.5 and* $K_{n,\lambda}$ *is defined at* (D.19). *Then*

$$\lim_{n\to\infty} \mathbb{P}(\mathcal{E}_n) = 1\,, \tag{D.26}$$

and

$$0 < \inf_{n\in\mathbb{N}} \nu \le \sup_{n\in\mathbb{N}} \nu < 1\,. \tag{D.27}$$

Moreover, if $\mathcal{E}_n$ *holds, then for all* $i, j = -d, \ldots, K$, $|(\widehat{\boldsymbol{H}}^{-1}_{n,\lambda,\varphi})_{ij}| \le (4/\tilde{c}_1)\nu^{|i-j|}$, *where* $0 < \tilde{c}_1 < \infty$ *is defined in Lemma D.14.*

Proof. Note that $\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}$ is a centered $2(d+1)-$banded matrix. Recalling Lemma D.14 and the discussion above, we have $e_{\widehat{\boldsymbol{H}},\min} = \tilde{c}_1\{1 + o_p(1)\}$ and $e_{\widehat{\boldsymbol{H}},\max} = \tilde{c}_{2,n,\lambda}\{1 + o_p(1)\}$, where $0 < c_1 \le \tilde{c}_1 \le c_2 \le \tilde{c}_{2,n,\lambda} \le c_2(1 + K_{n,\lambda}/c_2) < \infty$ for all n and as $n \to \infty$. Then using arguments similar to those used in Corollary D.7, we have $\lim_{n\to\infty} \mathbb{P}\{e_{\widehat{\boldsymbol{H}},\min} \ge (1/2)c_1\} = 1$. Similarly, for any $\xi > 0$,

$$\lim_{n\to\infty} \mathbb{P}\big(e_{\widehat{\boldsymbol{H}},\max} \ge \tilde{c}_{2,n,\lambda} + \xi\big) \le \lim_{n\to\infty} \mathbb{P}\big(|e_{\widehat{\boldsymbol{H}},\max} - \tilde{c}_{2,n,\lambda}| \ge \xi\big) = 0\,.$$

Then selecting $\xi = c_2$, we have $\lim_{n\to\infty} \mathbb{P}(\mathcal{E}_n) = 1$.

Note that $\nu = [\{\sqrt{2c_2(2+K_{n,\lambda}/c_2)/c_1}-1\}/\{\sqrt{2c_2(2+K_{n,\lambda}/c_2)/c_1}+1\}]^{1/(d+1)}$. Now, if $\mathcal{E}_n$ holds, noting that for all n and as $n \to \infty$, $0 < c_1 \le c_2(1+K_{n,\lambda}/c_2) < \infty$, we can apply Corollary D.17 on $\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}$ to obtain that

$$|(\widehat{\boldsymbol{H}}^{-1}_{n,\lambda,\varphi})_{ij}| \le \frac{4}{\tilde{c}_1}\nu^{|i-j|}\,.$$

Now since for all n and as $n \to \infty$, $0 < c_1 \le c_2 \le c_2(1+K_{n,\lambda}/c_2) < \infty$ and $\{0 \le K_{n,\lambda} < \infty\}$, we have $2 \le \inf_{n\in\mathbb{N}} \sqrt{2c_2(2+K_{n,\lambda}/c_2)/c_1} \le \sup_{n\in\mathbb{N}} \sqrt{2c_2(2+K_{n,\lambda}/c_2)/c_1} < \infty$. Thus, $0 < \inf_{n\in\mathbb{N}} \nu \le \sup_{n\in\mathbb{N}} \nu < 1$. $\square$

Lemma D.19. *Under Assumptions (A5) to (A11), $\boldsymbol{H}_{n,\lambda}$ is invertible, and for all $i,j = -d,\ldots,K$,*

$$|(\boldsymbol{H}^{-1}_{n,\lambda})_{ij}| \le \frac{2}{c_1}\nu^{|i-j|} \quad \textit{and} \quad 0 < \inf_{n\in\mathbb{N}} \nu \le \sup_{n\in\mathbb{N}} \nu < 1\,,$$

where c_1 is defined in Lemma D.5, ν is defined in Lemma D.18.

Proof. Note that $\boldsymbol{H}_{n,\lambda} \in \mathbb{R}^{(K+d+1)\times(K+d+1)}$ is a centered $2(d+1)-$banded matrix. From Lemma D.13, we have that $[c_1, c_2(1+K_{n,\lambda}/c_2)]$ is an interval that contains all its eigenvalues. Recalling the discussion in the proof of D.13, we have that $0 < c_1 \le c_2 \le c_2(1+K_{n,\lambda}/c_2) < \infty$ for all n and $n \to \infty$. Thus, we can apply Corollary D.17 to $\boldsymbol{H}_{n,\lambda}$ and

$$|\{(\boldsymbol{H}_{n,\lambda})^{-1}\}_{ij}| \le \frac{2}{c_1}\left(\frac{\sqrt{c_2(1+K_{n,\lambda}/c_2)/c_1}-1}{\sqrt{c_2(1+K_{n,\lambda}/c_2)/c_1}+1}\right)^{|i-j|/(d+1)}\,.$$

Now the function $(x-1)/(x+1)$ is increasing with x when $x > -1$. Thus, $\{\sqrt{c_2(1+K_{n,\lambda}/c_2)/c_1}-1\}/\{\sqrt{c_2(1+K_{n,\lambda}/c_2)/c_1}+1\} \le \{\sqrt{2c_2(2+K_{n,\lambda}/c_2)/c_1}-1\}/\{\sqrt{2c_2(2+K_{n,\lambda}/c_2)/c_1}+1\}$. The result then follows. $\square$

Lemma D.20. *Under Assumptions (A5) to (A11), if $\mathcal{E}_n$ holds, then*

$$|(\widehat{\boldsymbol{H}}^{-1}_{n,\lambda,\varphi} - \boldsymbol{H}^{-1}_{n,\lambda})_{ij}| \le e_m \underline{c}\nu^{|i-j|}(|i-j| + \overline{c})\,,$$

where $e_m = \max\{|e|$ *s.t.* e *is an eigenvalue of* $\boldsymbol{G}_n - \widehat{\boldsymbol{G}}_{n,\varphi}\}$, $\underline{c} = (d+1)8c_1^{-1}\tilde{c}_1^{-1}\nu^{-d-1}$ *with* c_1 *defined in Lemma D.5,* $\tilde{c}_1$ *defined in Lemma D.14,* ν *defined in Lemma D.18 and with* $\bar{c}$ *a constant. Moreover,* $\underline{c}$ *and* $\bar{c}$ *do not depend on* i *and* j*, and for all* n *and as* $n \to \infty$, $\{0 < \nu < 1\} \cap \{0 < \underline{c} < \infty\} \cap \{0 < \bar{c} < \infty\}$ *holds.*

Proof. By definition of $\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}^{-1}$ and $\boldsymbol{H}_{n,\lambda}^{-1}$, we can write

$$\begin{aligned}
\boldsymbol{H}_{n,\lambda}^{-1} &= \boldsymbol{H}_{n,\lambda}^{-1}\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}^{-1} \\
&= \boldsymbol{H}_{n,\lambda}^{-1}(\widehat{\boldsymbol{G}}_{n,\varphi} + \lambda\boldsymbol{D}_\ell/n)\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}^{-1} \\
&= \boldsymbol{H}_{n,\lambda}^{-1}\big[(\boldsymbol{G}_n + \lambda\boldsymbol{D}_\ell/n) - (\boldsymbol{G}_n - \widehat{\boldsymbol{G}}_{n,\varphi})\big]\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}^{-1} \\
&= \boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{H}_{n,\lambda}\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}^{-1} - \boldsymbol{H}_{n,\lambda}^{-1}(\boldsymbol{G}_n - \widehat{\boldsymbol{G}}_{n,\varphi})\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}^{-1} \\
&= \widehat{\boldsymbol{H}}_{n,\lambda,\varphi}^{-1} - \boldsymbol{H}_{n,\lambda}^{-1}(\boldsymbol{G}_n - \widehat{\boldsymbol{G}}_{n,\varphi})\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}^{-1}\,.
\end{aligned}$$

Thus, $\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}^{-1} - \boldsymbol{H}_{n,\lambda}^{-1} = \boldsymbol{H}_{n,\lambda}^{-1}(\boldsymbol{G}_n - \widehat{\boldsymbol{G}}_{n,\varphi})\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}^{-1}$.

Note that $\boldsymbol{G}_n - \widehat{\boldsymbol{G}}_{n,\varphi} \in \mathbb{R}^{(K+d+1)\times(K+d+1)}$ is a centered $2(d+1)-$banded matrix. That is, $(\boldsymbol{G}_n - \widehat{\boldsymbol{G}}_{n,\varphi})_{ij} = 0$ if $|i-j| > d+1$. Thus, we have for all $i, j = -d, \ldots, K$,

$$|(\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}^{-1} - \boldsymbol{H}_{n,\lambda}^{-1})_{ij}| = \left|\sum_{\ell=-d}^{K}\sum_{k=\max(\ell-d-1,-d)}^{\min(K,\ell+d+1)}(\boldsymbol{H}_{n,\lambda}^{-1})_{ik}(\boldsymbol{G}_n - \widehat{\boldsymbol{G}}_{n,\varphi})_{k\ell}(\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}^{-1})_{\ell j}\right|.$$

Then using Lemmas D.18 and D.19, we have if $\mathcal{E}_n$ holds, then for all $i, j = -d, \ldots, K$,

$$|(\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}^{-1} - \boldsymbol{H}_{n,\lambda}^{-1})_{ij}| \le \max_{k,\ell=-d,\ldots,K}|(\boldsymbol{G}_n - \widehat{\boldsymbol{G}}_{n,\varphi})_{k\ell}|B_n\,,$$

where $B_n = \sum_{\ell=-d}^{K}\sum_{k=\max(\ell-d-1,-d)}^{\min(K,\ell+d+1)} 8c_1^{-1}\tilde{c}_1^{-1}\nu^{|i-k|+|\ell-j|}$.

Similarly to the proof of Lemma C.5, to study B_n, we use a technique similar to that used in the proof of Lemma 6.4 of Zhou et al. (1998). Specifically, for $k = \max(\ell-d-1, -d), \ldots, \min(K, \ell+d+1)$, we have $|i-k| \ge |i-\ell| - d - 1$ for $i, \ell = -d, \ldots, K$. As shown in Lemma D.18, $0 < \inf_{n\in\mathbb{N}} \nu \le \sup_{n\in\mathbb{N}} \nu < 1$, we have $\nu^{|i-k|} \le \nu^{|i-\ell|-d-1}$. Thus, for $i, j = -d, \ldots, K$,

$$\begin{aligned}
B_n \le & (d+1)8c_1^{-1}\tilde{c}_1^{-1}\nu^{-d-1}\sum_{\ell=-d}^{K}\nu^{|i-\ell|+|\ell-j|} \\
= & (d+1)8c_1^{-1}\tilde{c}_1^{-1}\nu^{-d-1}\Bigg(\sum_{\ell=\min(i,j)+1}^{\max(i,j)}\nu^{|i-j|} + \sum_{\ell=-d}^{\min(i,j)}\nu^{|i-j|+2\min(i,j)-2\ell} \\
& + \sum_{\ell=\max(i,j)+1}^{K}\nu^{|i-j|+2\ell-2\max(i,j)}\Bigg) \\
= & (d+1)8c_1^{-1}\tilde{c}_1^{-1}\nu^{-d-1}\nu^{|i-j|}\Bigg(|i-j| + \sum_{\ell=0}^{\min(i,j)+d}\nu^{2\ell} + \sum_{\ell=1}^{K-\max(i,j)}\nu^{2\ell}\Bigg). \qquad \text{(D.28)}
\end{aligned}$$

Let $\bar{c}_{ij,K} = \sum_{\ell=0}^{\min(i,j)+d}\nu^{2\ell} + \sum_{\ell=1}^{K-\max(i,j)}\nu^{2\ell}$. Since the sum of such an infinite geometric series with $0 < \inf_{n\in\mathbb{N}} \nu \le \sup_{n\in\mathbb{N}} \nu < 1$ converges, we have $\sup_{K\in\mathbb{N}} |\bar{c}_{ij,K}| < \infty$. Taking a finite constant $\bar{c} \ge \sup_{K\in\mathbb{N}} |\bar{c}_{ij,K}|$, we find that $B_n \le \underline{c}\nu^{|i-j|}(|i-j| + \bar{c})$, where $\underline{c} = (d+1)8c_1^{-1}\tilde{c}_1^{-1}\nu^{-d-1}$.

Now, we have if $\mathcal{E}_n$ holds, then for all $i, j = -d, \ldots, K$,

$$|(\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}^{-1} - \boldsymbol{H}_{n,\lambda}^{-1})_{ij}| \le \max_{k,\ell=-d,\ldots,K} |(\boldsymbol{G}_n - \widehat{\boldsymbol{G}}_{n,\varphi})_{k\ell}|\underline{c}\nu^{|i-j|}(|i-j| + \bar{c}). \qquad \text{(D.29)}$$

Using the fact that for any symmetric real matrix $\boldsymbol{A} \in \mathbb{R}^{b\times b}$ where b is a positive integer, $\max_{i,j=1,\ldots,b} |(\boldsymbol{A})_{ij}| \le e_{\boldsymbol{A},m}$, where $e_{\boldsymbol{A},m} = \max\{|e_{\boldsymbol{A}}|$ s.t. $e_{\boldsymbol{A}}$ is an eigenvalue of $\boldsymbol{A}\}$ (see Horn and Johnson, 1985, page 315, equation (6,2)), we have $\max_{k,\ell=-d,\ldots,K} |(\boldsymbol{G}_n - \widehat{\boldsymbol{G}}_{n,\varphi})_{k\ell}| \le e_m$. Then combining with (D.29), the result follows. □

Now we investigate e_m :

Lemma D.21. *Under Assumptions (A5) to (A10), $e_m = O_p(\sqrt{(n\delta)^{-1}\log K})$, where $e_m = \max\{|e|$ s.t. e is an eigenvalue of $\boldsymbol{G}_n - \widehat{\boldsymbol{G}}_{n,\varphi}\}$.*

Proof. By the definition of e_m, we have $e_m = \max(|e_{\max}|, |e_{\min}|)$, where $e_{\max}$ and $e_{\min}$ are respectively the maximum and minimum eigenvalues of $\boldsymbol{G}_n - \widehat{\boldsymbol{G}}_{n,\varphi}$. Note that $\boldsymbol{G}_n - \widehat{\boldsymbol{G}}_{n,\varphi}$ is a real symmetric matrix. Then $|e_{\max}| = |\max_{\boldsymbol{v}^\top \boldsymbol{v}=1} \boldsymbol{v}^\top(\boldsymbol{G}_n - \widehat{\boldsymbol{G}}_{n,\varphi})\boldsymbol{v}| \leq \max_{\boldsymbol{v}^\top \boldsymbol{v}=1} |\boldsymbol{v}^\top(\boldsymbol{G}_n - \widehat{\boldsymbol{G}}_{n,\varphi})\boldsymbol{v}|$. On the other hand, $|e_{\min}| = |\min_{\boldsymbol{v}^\top \boldsymbol{v}=1} \boldsymbol{v}^\top(\boldsymbol{G}_n - \widehat{\boldsymbol{G}}_{n,\varphi})\boldsymbol{v}| = |-\max_{\boldsymbol{v}^\top \boldsymbol{v}=1}\{-\boldsymbol{v}^\top(\boldsymbol{G}_n - \widehat{\boldsymbol{G}}_{n,\varphi})\boldsymbol{v}\}| \leq \max_{\boldsymbol{v}^\top \boldsymbol{v}=1} |\boldsymbol{v}^\top(\boldsymbol{G}_n - \widehat{\boldsymbol{G}}_{n,\varphi})\boldsymbol{v}|$. Thus, $e_m \leq \max_{\boldsymbol{v}^\top \boldsymbol{v}=1} |\boldsymbol{v}^\top(\boldsymbol{G}_n - \widehat{\boldsymbol{G}}_{n,\varphi})\boldsymbol{v}|$. Then using (D.11), the result follows. □

D.3 Asymptotic properties of the terms in (C.16)

Now we use the results of the matrices to study the last four terms in (C.16) one by one. Regarding the third term in (C.16), we have

Lemma D.22. *Under Assumption* **A**, *for all* $x \in [a, b]$, *we have*

$$\mathbb{E}\left\{\frac{\phi}{n\delta}\boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}\boldsymbol{\Delta}(\underline{g} - \boldsymbol{Q}\boldsymbol{Z}^*)\right\} = 0$$

and

$$\operatorname{var}\left\{\frac{\phi}{n\delta}\boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}\boldsymbol{\Delta}(\underline{g} - \boldsymbol{Q}\boldsymbol{Z}^*)\right\} = V_{n,\delta}(x) + o\{(n\delta)^{-1}\},$$

where $V_{n,\delta}(x)$ *is defined as in Theorem 1. Moreover, as* $n \to \infty$,

$$\{V_{n,\delta}(x)\}^{-1/2}\frac{\phi}{n\delta}\boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}\boldsymbol{\Delta}(\underline{g} - \boldsymbol{Q}\boldsymbol{Z}^*) \xrightarrow{D} N(0,1)\,. \tag{D.30}$$

Proof of Lemma D.22. Recalling the definition of $\boldsymbol{\Phi}, \boldsymbol{\Delta}, \boldsymbol{\mathcal{N}}$ at page 11 in Delaigle et al. (2017) and the definition of $\underline{g}$ at page 22, we can write

$$\begin{aligned}
&\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}\boldsymbol{\Delta}(\underline{g} - \boldsymbol{Q}\boldsymbol{Z}^*) \\
&= \begin{pmatrix}\boldsymbol{N}(X_{1,1}) & \boldsymbol{N}(X_{2,1}) & \dots & \boldsymbol{N}(X_{n_J,J})\end{pmatrix}\begin{pmatrix}\varphi_1(q_0)\Delta_{1,1}\{g(X_{1,1}) - q_0^{1-n_1}Z_1^*\} \\ \varphi_1(q_0)\Delta_{2,1}\{g(X_{2,1}) - q_0^{1-n_1}Z_1^*\} \\ \vdots \\ \varphi_J(q_0)\Delta_{n_J,J}\{g(X_{n_J,J}) - q_0^{1-n_J}Z_J^*\}\end{pmatrix}
\end{aligned}$$

$$= \sum_{j=1}^{J} \sum_{i=1}^{n_j} \varphi_j(q_0) \Delta_{i,j} \{ g(X_{i,j}) - q_0^{1-n_j} Z_j^* \} \boldsymbol{N}(X_{i,j}) \,, \tag{D.31}$$

so that

$$\frac{\phi}{n\delta} \boldsymbol{N}^\top(x) \boldsymbol{H}_{n,\lambda}^{-1} \boldsymbol{\mathcal{N}}^\top \boldsymbol{\Phi} \boldsymbol{\Delta} (\underline{g} - \boldsymbol{Q} \boldsymbol{Z}^*) = \sum_{j=1}^{J} T_j \,, \tag{D.32}$$

where for $j = 1, \ldots, J$,

$$T_j = \frac{\phi}{n\delta} \boldsymbol{N}^\top(x) \boldsymbol{H}_{n,\lambda}^{-1} \sum_{i=1}^{n_j} \varphi_j(q_0) \Delta_{i,j} \{ g(X_{i,j}) - q_0^{1-n_j} Z_j^* \} \boldsymbol{N}(X_{i,j}) \,. \tag{D.33}$$

Next we calculate the mean and variance of (D.30). For the mean, we have

$$\mathbb{E}\bigg\{ \frac{\phi}{n\delta} \boldsymbol{N}^\top(x) \boldsymbol{H}_{n,\lambda}^{-1} \boldsymbol{N}^\top \boldsymbol{\Phi} \boldsymbol{\Delta} (\underline{g} - \boldsymbol{Q} \boldsymbol{Z}^*) \bigg\} = \sum_{j=1}^{J} \mathbb{E}(T_j) \,,$$

where

$$\begin{aligned} \mathbb{E}(T_j) &= \frac{\phi}{n\delta} \boldsymbol{N}^\top(x) \boldsymbol{H}_{n,\lambda}^{-1} \sum_{i=1}^{n_j} \varphi_j(q_0) \mathbb{E}\big(\mathbb{E}\big[\Delta_{i,j} \{ g(X_{i,j}) - q_0^{1-n_j} Z_j^* \} \boldsymbol{N}(X_{i,j}) \big| X_{i,j} \big] \big) \\ &= \frac{\phi}{n\delta} \boldsymbol{N}^\top(x) \boldsymbol{H}_{n,\lambda}^{-1} \sum_{i=1}^{n_j} \varphi_j(q_0) \mathbb{E}\big[\{ \mathbb{E}(\Delta_{i,j} | X_{i,j}) g(X_{i,j}) - \mathbb{E}(\Delta_{i,j} q_0^{1-n_j} Z_j^* | X_{i,j}) \} \boldsymbol{N}(X_{i,j}) \big] \,. \end{aligned}$$

Now, for $i = 1, \ldots, n_j$ and $j = 1, \ldots, J$, and for all $x' \in \mathbb{R}$, we have

$$\begin{aligned} &\mathbb{E}(\Delta_{i,j} | X_{i,j} = x') g(x') - \mathbb{E}(\Delta_{i,j} q_0^{1-n_j} Z_j^* | X_{i,j} = x') \\ &= \mathbb{P}(\Delta_{i,j} = 1 | X_{i,j} = x') g(x') - q_0^{1-n_j} \mathbb{P}(\Delta_{i,j} = 1, Z_j^* = 1 | X_{i,j} = x') \\ &= \mathbb{P}(\Delta_{i,j} = 1 | X_{i,j} = x') g(x') - \mathbb{P}(\Delta = 1 | X_{i,j} = x', Z_j^* = 1) \mathbb{E}(q_0^{1-n_j} Z_j^* | X_{i,j} = x') \\ &= \mathbb{P}(\Delta_{i,j} = 1 | X_{i,j} = x') g(x') - \mathbb{P}(\Delta = 1 | X_{i,j} = x', Z_j^* = 1) \{ 1 - p(x') \} \\ &= 0 \,, \end{aligned} \tag{D.34}$$

where the second last equality comes from (3.2) and the last equality comes from (3.4).

Therefore, $\mathbb{E}(T_j) = 0$ for $j = 1, \dots, J$, so that

$$\mathbb{E}\left\{\frac{\phi}{n\delta}\boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}\boldsymbol{\Delta}(\underline{g} - \boldsymbol{Q}\boldsymbol{Z}^*)\right\} = 0\,.$$

Regarding the variance, since the $(\Delta_{i,j}, X_{i,j})$'s are independent and the Z_j^*'s are independent, we have that the T_j's are independent, which implies that

$$\mathrm{var}\left\{\frac{\phi}{n\delta}\boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}\boldsymbol{\Delta}(\underline{g} - \boldsymbol{Q}\boldsymbol{Z}^*)\right\} = \sum_{j=1}^{J}\mathrm{var}(T_j)\,,$$

where, for $j = 1, \dots, J$,

$$\begin{aligned}
\mathrm{var}(T_j) \quad &= \frac{\phi^2}{(n\delta)^2}\mathrm{var}\left[\boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}\sum_{i=1}^{n_j}\varphi_j(q_0)\Delta_{i,j}\{g(X_{i,j}) - q_0^{1-n_j}Z_j^*\}\boldsymbol{N}(X_{i,j})\right] \\
&= \frac{\phi^2}{(n\delta)^2}\varphi_j^2(q_0)\sum_{i=1}^{n_j}\mathrm{var}\big[\Delta_{i,j}\{g(X_{i,j}) - q_0^{1-n_j}Z_j^*\}\boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{N}(X_{i,j})\big] \\
&\quad + \frac{\phi^2}{(n\delta)^2}\varphi_j^2(q_0)\sum_{i\neq i'}^{n_j}\mathrm{cov}\big[\Delta_{i,j}\{g(X_{i,j}) - q_0^{1-n_j}Z_j^*\}\boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{N}(X_{i,j}), \\
&\quad \Delta_{i',j}\{g(X_{i',j}) - q_0^{1-n_j}Z_j^*\}\boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{N}(X_{i',j})\big] \\
&\equiv I_{1,j} + I_{2,j}\,,
\end{aligned} \tag{D.35}$$

where $I_{1,j}$ denotes the variance term and $I_{2,j} = \phi^2/(n\delta)^2\varphi_j^2(q_0)\sum_{i\neq i'}^{n_j} I_{2,i,i',j}$ denotes the covariance term.

We investigate the $I_{2,j}$'s first. For all $i, i' = 1, \dots, n_j$ and $j = 1, \dots, J$, we have

$$\begin{aligned}
I_{2,i,i',j} =&\mathbb{E}\Big[\Delta_{i,j}\Delta_{i',j}\{g(X_{i,j}) - q_0^{1-n_j}Z_j^*\}\{g(X_{i',j}) - q_0^{1-n_j}Z_j^*\} \\
&\times \boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{N}(X_{i,j})\boldsymbol{N}^\top(X_{i',j})\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{N}(x)\Big] \\
&- \left\{\mathbb{E}\left[\Delta_{i,j}\{g(X_{i,j}) - q_0^{1-n_j}Z_j^*\}\boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{N}(X_{i,j})\right]\right\}^2
\end{aligned}$$

and for all $i, i' = 1, \ldots, n_j$ and $j = 1, \ldots, J$, we have

$$\begin{aligned}
&\Bigg|\mathbb{E}\Bigg[\Delta_{i,j}\Delta_{i',j}\{g(X_{i,j}) - q_0^{1-n_j} Z_j^*\}\{g(X_{i',j}) - q_0^{1-n_j} Z_j^*\} \\
&\quad \times \boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{N}(X_{i,j})\boldsymbol{N}^\top(X_{i',j})\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{N}(x)\Bigg]\Bigg| \\
&\le \mathbb{E}\Bigg[\Bigg|\Delta_{i,j}\Delta_{i',j}\{g(X_{i,j}) - q_0^{1-n_j} Z_j^*\}\{g(X_{i',j}) - q_0^{1-n_j} Z_j^*\} \\
&\quad \times \boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{N}(X_{i,j})\boldsymbol{N}^\top(X_{i',j})\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{N}(x)\Bigg|\Bigg] \\
&\le \mathrm{Const}_1 \times \mathbb{E}\Big[\big|\boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{N}(X_{i,j})\big|\Big]\mathbb{E}\Big[\big|\boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{N}(X_{i',j})\big|\Big],
\end{aligned}$$

for some positive finite constant Const_1. For $t = i, i'$, we have

$$\Bigg|\mathbb{E}\Bigg[\Delta_{t,j}\{g(X_{t,j}) - q_0^{1-n_j} Z_j^*\}\boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{N}(X_{t,j})\Bigg]\Bigg| \le \mathrm{Const}_2 \times \mathbb{E}\Big[\big|\boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{N}(X_{t,j})\big|\Big], \tag{D.36}$$

for some positive finite constant Const_2.

Recall that $\boldsymbol{N}^\top(x) = \{N_{-d,d+1}(x)\ldots N_{K,d+1}(x)\}$ is a $K+d+1$ row vector and $\boldsymbol{H}_{n,\lambda,\varphi}^{-1}$ is a $(K+d+1)\times(K+d+1)$ matrix. Using B-spline Property 4 and Assumption (A7) that $\max_{i=-d,\ldots,K+d}\delta_i \le \delta$, we have

$$\begin{aligned}
\mathbb{E}\Big[\big|\boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{N}(X_{t,j})\big|\Big] &= \mathbb{E}\Bigg[\Bigg|\sum_{i=-d}^{K}\{\boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}\}_{i+d+1}N_{i,d+1}(X_{t,j})\Bigg|\Bigg] \\
&\le \mathbb{E}\Bigg[\sum_{i=-d}^{K}\big|\{\boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}\}_{i+d+1}\big|N_{i,d+1}(X_{t,j})\Bigg] \\
&= \sum_{i=-d}^{K}\big|\{\boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}\}_{i+d+1}\big|\int_a^b N_{i,d+1}(y)f_X(y)dy \\
&\le \sum_{i=-d}^{K}\big|\{\boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}\}_{i}\big|f_{\max}\frac{t_{i+d+1}-t_i}{d+1}
\end{aligned}$$

$$
\begin{aligned}
&\leq f_{\max}\delta \sum_{i=-d}^{K}\left|\sum_{j=-d}^{K} N_{j,d+1}(x)(\boldsymbol{H}_{n,\lambda}^{-1})_{ji}\right| \\
&\leq f_{\max}\delta \max_{j=-d,\ldots,K}\sum_{i=-d}^{K}|(\boldsymbol{H}_{n,\lambda}^{-1})_{ji}| \sum_{j=-d}^{K} N_{j,d+1}(x)\,. \qquad \text{(D.37)}
\end{aligned}
$$

Then using B-spline Property 1, we have $\sum_{i=-d}^{K} N_{i,d+1}(x) = 1$ for all $x \in [a,b]$ and by Lemma D.19, we have $\max_{j=-d,\ldots,K}\sum_{i=-d}^{K}|(\boldsymbol{H}_{n,\lambda}^{-1})_{ji}| = O(1)$. Thus,

$$
\mathbb{E}\left[\left|\boldsymbol{N}^{\top}(x)\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{N}(X_{t,j})\right|\right] = O(\delta)\,. \qquad \text{(D.38)}
$$

Therefore, for all $i, i' = 1,\ldots,n_j$ and $j = 1,\ldots,J$,

$$
\begin{aligned}
I_{2,i,i',j} &\leq \text{Const}_1 \times O(\delta)\times O(\delta) - \text{Const}_2^2 \times O(\delta)\times O(\delta) \\
&= O(\delta^2)\,.
\end{aligned}
$$

Recall that $\phi = n/\{\sum_{j=1}^{J} n_j\varphi(q_0)\}$. Under Assumption (A10), $\phi_2^{-1} \leq \phi \leq \phi_1^{-1}$. Then we have, for $j = 1,\ldots,J$,

$$
\begin{aligned}
I_{2,j} &= \frac{\phi^2}{(n\delta)^2}\varphi_j^2(q_0)\sum_{i\neq i'}^{n_j} I_{2,i,i',j} \\
&= \phi^2\frac{\varphi_j^2(q_0)}{(n\delta)^2}(n_j+1)n_j \times O(\delta^2) = O(n^{-2})\,. \qquad \text{(D.39)}
\end{aligned}
$$

Now for the $I_{1,j}$'s at (D.35), we have

$$
\begin{aligned}
I_{1,j} =& \frac{\phi^2}{(n\delta)^2}\varphi_j^2(q_0)\sum_{i=1}^{n_j}\mathbb{E}[\Delta_{i,j}^2\{g(X_{i,j}) - q_0^{1-n_j}Z_j^*\}^2 \\
&\times \boldsymbol{N}^{\top}(x)\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{N}(X_{i,j})\boldsymbol{N}^{\top}(X_{i,j})\boldsymbol{H}_{n,\lambda,\varphi}^{-1}\boldsymbol{N}(x)] \\
&- \frac{\phi^2}{(n\delta)^2}\varphi_j^2(q_0)\sum_{i=1}^{n_j}\left\{\mathbb{E}[\Delta_{i,j}\{g(X_{i,j}) - q_0^{1-n_j}Z_j^*\}\boldsymbol{N}^{\top}(x)\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{N}(X_{i,j})]\right\}^2
\end{aligned}
$$

$$\equiv \frac{\phi^2}{(n\delta)^2}\varphi_j^2(q_0)\sum_{i=1}^{n_j} I_{1,i,j} - \frac{\phi^2}{(n\delta)^2}\varphi_j^2(q_0)\sum_{i=1}^{n_j} I_{2,i,j}\,. \tag{D.40}$$

By (D.36) and (D.38), we have

$$I_{2,i,j} = O(\delta^2)\,.$$

Thus, using Assumption (A10),

$$\frac{\phi^2}{(n\delta)^2}\varphi_j^2(q_0)\sum_{i=1}^{n_j} I_{2,i,j} = O(n^{-2})\,. \tag{D.41}$$

For the first term in the right hand side of (D.40), we have

$$I_{1,i,j} = \mathbb{E}\bigg(\mathbb{E}\big[\Delta_{i,j}\{g(X_{i,j}) - q_0^{1-n_j}Z_j^*\}^2|X_{i,j}\big]\boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{N}(X_{i,j})\boldsymbol{N}^\top(X_{i,j})\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{N}(x)\bigg)\,.$$

Now, using (D.34), (3.3) and the definition of π at page 6 in Delaigle et al. (2017), we have

$$\begin{aligned}
&\mathbb{E}\big[\Delta_{i,j}\{g(X_{i,j}) - q_0^{1-n_j}Z_j^*\}^2|X_{i,j}\big] \\
&= g^2(X_{i,j})\mathbb{E}(\Delta_{i,j}|X_{i,j}) - 2g(X_{i,j})\mathbb{E}(\Delta_{i,j}q_0^{1-n_j}Z_j^*|X_{i,j}) + q_0^{1-n_j}\mathbb{E}\{\Delta_{i,j}q_0^{1-n_j}Z_j^*|X_{i,j}\} \\
&= \mathbb{E}\{q_0^{1-n_j}Z_j^*|X_{i,j},\Delta_{i,j}=1\}\mathbb{P}(\Delta_{i,j}=1|X_{i,j})\{q_0^{1-n_j} - g(X_{i,j})\} \\
&= g(X_{i,j})\pi(X_{i,j})\{q_0^{1-n_j} - g(X_{i,j})\}\,.\,.
\end{aligned} \tag{D.42}$$

Therefore,

$$\begin{aligned}
I_{1,i,j} =& \mathbb{E}\bigg(\Big[\pi(X_{i,j})g(X_{i,j})\{q_0^{1-n_j} - g(X_{i,j})\}\Big]\boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{N}(X_{i,j})\boldsymbol{N}^\top(X_{i,j})\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{N}(x)\bigg) \\
=& \boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}G_{g,j}\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{N}(x)\,,
\end{aligned} \tag{D.43}$$

where $\boldsymbol{G}_{g,j} = \delta^{-1}\int_a^b \pi(y)g(y)\{q_0^{1-n_j} - g(y)\}\boldsymbol{N}(y)\boldsymbol{N}^\top(y)f_X(y)\,dy$ is defined at (4.5). Then

$$\sum_{j=1}^{J}\frac{\phi^2}{(n\delta)^2}\varphi_j^2(q_0)\sum_{i=1}^{n_j} I_{1,i,j}$$

$$
\begin{aligned}
&= \frac{\phi^2}{\delta n^2} \sum_{j=1}^{J} \varphi_j^2(q_0) \sum_{i=1}^{n_j} \boldsymbol{N}^\top(x) \boldsymbol{H}_{n,\lambda}^{-1} \boldsymbol{G}_{g,j} \boldsymbol{H}_{n,\lambda}^{-1} \boldsymbol{N}(x) \\
&= \frac{n^2}{\delta\{\sum_{j=1}^J n_j \varphi_j(q_0)\}^2} \frac{\sum_{j=1}^J n_j \varphi_j^2(q_0)}{n^2} \boldsymbol{N}^\top(x) \boldsymbol{H}_{n,\lambda}^{-1} \boldsymbol{G}_{g,j} \boldsymbol{H}_{n,\lambda}^{-1} \boldsymbol{N}(x) \\
&= V_{n,\delta}(x)\,, \qquad \text{(D.44)}
\end{aligned}
$$

where $V_{n,\delta}$ is defined at (4.3). Then recalling (D.40) and combining with (D.41), we have

$$
\sum_{j=1}^{J} I_{1,j} = V_{n,\delta}(x) + O(n^{-1})\,.
$$

Using Assumptions (A6) and (A7), we have $\delta^{-1} \to \infty$ as $n \to \infty$. Then plugging this and (D.39) into (D.35), we have

$$
\sum_{j=1}^{J} \mathrm{var}(T_j) = \sum_{j=1}^{J} I_{1,j} + \sum_{j=1}^{J} I_{2,j} = V_{n,\delta}(x) + O(n^{-1}) = V_{n,\delta}(x) + o\{(n\delta)^{-1}\}\,. \qquad \text{(D.45)}
$$

That is

$$
\mathrm{var}\left\{ \frac{\phi}{n\delta} \boldsymbol{N}^\top(x) \boldsymbol{H}_{n,\lambda}^{-1} \boldsymbol{\mathcal{N}}^\top \Phi \Delta(\underline{g} - QZ^*) \right\} = V_{n,\delta}(x) + o\{(n\delta)^{-1}\}\,. \qquad \text{(D.46)}
$$

Now we show that $V_{n,\delta}(x) \asymp (n\delta)^{-1}$. Recall from the definition of $\boldsymbol{G}_{g,j}$ at (4.5) that for each $j = 1, \ldots, J$,

$$
\begin{aligned}
&\boldsymbol{N}^\top(x) \boldsymbol{H}_{n,\lambda}^{-1} \boldsymbol{G}_{g,j} \boldsymbol{H}_{n,\lambda}^{-1} \boldsymbol{N}(x) \\
&= \delta^{-1} \int_a^b \Big[\pi(y) g(y) \{q_0^{1-n_j} - g(y)\}\Big] \Big\{ \boldsymbol{N}^\top(x) \boldsymbol{H}_{n,\lambda}^{-1} \boldsymbol{N}(y) \Big\}^2 f_X(y)\, dy\,.
\end{aligned}
$$

Assumptions (A1), (A8) and (3.7) imply that $g(q_0^{1-n_j} - g) f_X$ is continuous on $[a,b]$. Furthermore, using Assumption (A5), $\pi(y)\{\boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{N}(y)\}^2$ is non-negative for any $y \in [a,b]$. Thus, by the mean value theorem for integrals, there exists $\tau \in [a,b]$ such that

$$
\boldsymbol{N}^\top(x) \boldsymbol{H}_{n,\lambda}^{-1} \boldsymbol{G}_{g,j} \boldsymbol{H}_{n,\lambda}^{-1} \boldsymbol{N}(x)
$$

$$= [g(\tau)\{q_0^{1-n_j} - g(\tau)\}f_X(\tau)]\delta^{-1}\int_a^b \pi(y)\Big\{\boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{N}(y)\Big\}^2\,dy\,.$$

Using the definition at page 21, $\delta^{-1}\int_a^b \pi(y)\boldsymbol{N}(y)\boldsymbol{N}^\top(y)dy = \boldsymbol{G}_n$. Then

$$\begin{aligned}&\boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{G}_{g,j}\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{N}(x)\\ &\qquad\qquad = [g(\tau)\{q_0^{1-n_j} - g(\tau)\}f_X(\tau)]\boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{G}_n\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{N}(x)\,.\end{aligned}$$

Applying to Corollary D.11 with Lemmas D.5 and D.13, we have

$$\frac{c_1}{\{c_2(1+K_{n,\lambda}/c_2)\}^2}\boldsymbol{N}^\top(x)\boldsymbol{N}(x) \le \boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{G}_n\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{N}(x) \le \frac{c_2}{c_1^2}\boldsymbol{N}^\top(x)\boldsymbol{N}(x)\,.$$

Now by B-spline Property 1, $N_{i,d+1}(x) \ge 0$. Thus, $\boldsymbol{N}^\top(x)\boldsymbol{N}(x) = \sum_{i=-d}^K N_{i,d+1}^2(x) \le \{\sum_{i=-d}^K N_{i,d+1}(x)\}^2 = 1$. On the other hand, for any $x \in [a,b]$, $N_{i,d+1}(x) = 0$ when $x \notin (t_i, t_{i+d+1})$ for all $i = -d,\dots,K$. Thus, we have $N_{i,d+1}(x) = 0$ if $i > i_x$ or $i < i_x - d$, where i_x is the index of the knot such that $x \in [t_{i_x}, t_{i_x+1})$. Thus, $\boldsymbol{N}^\top(x)\boldsymbol{N}(x) = \sum_{i=i_x-d}^{i_x} N_{i,d+1}^2(x) \ge \{\sum_{i=i_x-d}^{i_x} N_{i,d+1}(x)\}^2/(d+1) = \{\sum_{i=-d}^K N_{i,d+1}(x)\}^2/(d+1) = 1/(d+1)$. Therefore, for all $x \in [a,b]$,

$$\frac{c_1}{c_2^2(1+K_{n,\lambda}/c_2)^2(d+1)} \le \boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{G}_n\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{N}(x) \le \frac{c_2}{c_1^2}\,. \tag{D.47}$$

Then, since $0 < c_1 \le c_2 < \infty$ and using assumption that $0 \le \inf_{n\in\mathbb{N}} K_{n,\lambda} \le \sup_{n\in\mathbb{N}} K_{n,\lambda} < \infty$, we have

$$\boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{G}_n\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{N}(x) \asymp 1\,. \tag{D.48}$$

Under Assumptions (A9) and (A10), we have $1 \le q_0^{1-n_j} < \infty$, $\sum_{j=1}^J n_j\varphi_j^2(q_0) \asymp n$ and $\{\sum_{j=1}^J n_j\varphi_j(q_0)\}^2 \asymp n^2$. Then we have for all $x \in [a,b]$,

$$V_{n,\delta}(x) = \frac{\sum_{j=1}^J n_j\varphi_j^2(q_0)}{\delta\{\sum_{j=1}^J n_j\varphi_j(q_0)\}^2}[g(\tau)\{q_0^{1-n_j} - g(\tau)\}f_X(\tau)]\boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{G}_n\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{N}(x) \asymp n^{-1}\delta^{-1}\,.$$

Therefore, by (D.46),

$$\mathrm{var}\left\{\frac{\phi}{n\delta}\boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{\mathcal{N}}^\top\Phi\Delta(\underline{g}-QZ^*)\right\} = V_{n,\delta} + o_p(n^{-1}\delta^{-1})\,,$$

where the o_p term is negligible compared to $V_{n,\delta}$.

To summarise, we have shown that $(n\delta)^{-1}\phi\boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{\mathcal{N}}^\top\Phi\Delta(\underline{g}-QZ^*) = \sum_{j=1}^J T_j$, where the T_j's are defined at (D.33) and are independent and such that $\mathbb{E}(T_j)=0$ for all j. Letting $S_J^2 = \sum_{j=1}^J \mathrm{var}(T_j)$, it follows from (D.45) that

$$S_J^2 = V_{n,\delta}(x) + o(n^{-1}\delta^{-1})\,, \tag{D.49}$$

where, using the above calculations, $V_{n,\delta}(x) \asymp n^{-1}\delta^{-1}$.

To establish asymptotic normality, we apply Lyapounov's central limit theorem. Note that by Jensen's inequality, for any real sequence $a_i, i=1,\ldots,n$, we have $(\sum_{i=1}^n a_i)^4 \le n^3\sum_{i=1}^n a_i^4$. Then

$$\begin{aligned}
\mathbb{E}(T_j^4) &= \mathbb{E}\left(\frac{\phi^4}{(n\delta)^4}\left[\sum_{i=1}^{n_j}\varphi_j(q_0)\Delta_{i,j}\{g(X_{i,j})-q_0^{1-n_j}Z_j^*\}\boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{N}(X_{i,j})\right]^4\right)\\
&\le \mathbb{E}\left(\frac{\phi^4 n_j^3}{(n\delta)^4}\sum_{i=1}^{n_j}\varphi_j^4(q_0)\Delta_{i,j}^4\big\{g(X_{i,j})-q_0^{1-n_j}Z_j^*\big\}^4\big\{\boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{N}(X_{i,j})\big\}^4\right)\\
&= \frac{\phi^4 n_j^3}{(n\delta)^4}\sum_{i=1}^{n_j}\varphi_j^4(q_0)\mathbb{E}\Big(\big\{\boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{N}(X_{i,j})\big\}^4\mathbb{E}\big[\Delta_{i,j}^4\big\{g(X_{i,j})-q_0^{1-n_j}Z_j^*\big\}^4|X_{i,j}\big]\Big)\,,
\end{aligned}$$

where, recalling the definition of g at (3.3) and the definition of π at page 6 in Delaigle et al. (2017), we have

$$\begin{aligned}
&\mathbb{E}\big[\Delta_{i,j}^4\big\{g(X_{i,j})-q_0^{1-n_j}Z_j^*\big\}^4|X_{i,j}\big]\\
&\quad\le \mathbb{E}\big[\Delta_{i,j}\big\{4g^4(X_{i,j})+4(q_0^{1-n_j})^4(Z_j^*)^4\big\}|X_{i,j}\big]\\
&\quad= 4g^4(X_{i,j})\pi(X_{i,j})+4(q_0^{1-n_j})^3\mathbb{E}(q_0^{1-n_j}Z_j^*|X_{i,j},\Delta_{i,j}=1)\mathbb{P}(\Delta_{i,j}=1|X_{i,j})
\end{aligned}$$

$$= 4g(X_{i,j})\pi(X_{i,j})\{g^3(X_{i,j}) + (q_0^{1-n_j})^3\}\,.$$

Therefore,

$$\begin{aligned}
\mathbb{E}(T_j^4) &\leq \frac{\phi^4 n_j^3}{(n\delta)^4}\sum_{i=1}^{n_j}\varphi_j^4(q_0)\int_a^b 4g(y)\pi(y)\{g^3(y)+(q_0^{1-n_j})^3\}\big\{\boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{N}(y)\big\}^4 f_X(y)\,dy \\
&\leq \frac{4\phi^4 n_j^4\varphi_j^4(q_0)}{n^4\delta^3}\sup_{y\in[a,b]}\left[g(y)\{g^3(y)+(q_0^{1-n_j})^3\}\big\{\boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{N}(y)\big\}^2\right] \\
&\quad \times\int_a^b \delta^{-1}\pi(y)\big\{\boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{N}(y)\big\}^2 f_X(y)\,dy\,.
\end{aligned}$$

Using again the definition at page 21, $\delta^{-1}\int_a^b \pi(y)\boldsymbol{N}(y)\boldsymbol{N}^\top(y)f_X(y)\,dy = \boldsymbol{G}_n$. Then by (D.48), we have $\int_a^b \delta^{-1}\pi(y)\big\{\boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{N}(y)\big\}^2 f_X(y)\,dy \asymp 1$.

Now, for all $y\in[a,b]$, by B-spline Property 1 again, we have $0 \leq \boldsymbol{N}^\top(y)\boldsymbol{N}(y) = \sum_{i=-d}^K N_{i,d+1}^2(y) \leq \{\sum_{i=-d}^K N_{i,d+1}(y)\}^2 = 1$. Then using Lemma D.8 and the definition of the eigenvalues of symmetric matrices, we have that all the eigenvalues of $\boldsymbol{N}(y)\boldsymbol{N}^\top(y)$ are in $[0,1]$. Then using Corollary D.11 and Lemma D.13, we have under Assumption (A11),

$$\begin{aligned}
\big\{\boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{N}(y)\big\}^2 &= \boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{N}(y)\boldsymbol{N}^\top(y)\boldsymbol{H}_{n,\lambda}^{-1}\boldsymbol{N}(x) \\
&\leq \boldsymbol{N}^\top(x)\boldsymbol{H}_{n,\lambda}^{-2}\boldsymbol{N}(x) = O(1)\,.
\end{aligned}$$

Recalling that $\phi = n/(\sum_{j=1}^J n_j\varphi_j)$, which is $O(1)$, and using Assumptions (A2), (A8), (A9) and (A10), we deduce that $\mathbb{E}(T_j^4) = O(n^{-4}\delta^{-3})$. Furthermore, from (D.49), we have that $S_J^4 \asymp n^{-2}\delta^{-2}$. Then, we have $(S_J^4)^{-1}\sum_{j=1}^J \mathbb{E}(T_j^4) = O(n^{-1}\delta^{-1})$. Again using Assumptions (A6) and (A7), we have $(n\delta)^{-1}\to 0$ as $n\to\infty$. Therefore, the conditions of Lyapounov's central limit theorem are satisfied, and we have as $n\to\infty$,

$$\frac{1}{S_J}\sum_{j=1}^J T_j \xrightarrow{D} N(0,1)\,.$$

Note that by (D.49), $S_J^2/V_{n,\delta}(x) = 1 + o(1)$, which implies that, as $n \to \infty$,

$$\frac{1}{\sqrt{V_{n,\delta}(x)}} \sum_{j=1}^{J} T_j \xrightarrow{D} N(0,1)\,.$$

Recalling (D.32), this proves (D.30).

□

Next, regarding the second term in (C.15), we have

Lemma D.23. *Under Assumption* **A**, *for all* $x \in [a,b]$,

$$\frac{1}{n\delta}\boldsymbol{N}^\top(x)(\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}^{-1} - \boldsymbol{H}_{n,\lambda}^{-1})\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}\boldsymbol{\Delta}(\underline{g} - \boldsymbol{Q}\boldsymbol{Z}^*) = o_p\big\{\sqrt{(n\delta)^{-1}}\big\}\,.$$

Proof of Lemma D.23. Let T_j be the $(j+d+1)$th entry of the vector $\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}\boldsymbol{\Delta}(\underline{g} - \boldsymbol{Q}\boldsymbol{Z}^*)$, for $j = -d,\ldots,K$. Then we have, for all $x \in [a,b]$,

$$\begin{aligned}&\frac{1}{n\delta}\boldsymbol{N}^\top(x)(\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}^{-1} - \boldsymbol{H}_{n,\lambda}^{-1})\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}\boldsymbol{\Delta}(\underline{g} - \boldsymbol{Q}\boldsymbol{Z}^*)\\ &\qquad = (n\delta)^{-1}\sum_{j=-d}^{K}\sum_{i=-d}^{K} N_{i,d+1}(x)(\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}^{-1} - \boldsymbol{H}_{n,\lambda}^{-1})_{ij}T_j\,.\end{aligned}$$

By B-spline Property 1 and letting $0 \le i_x \le K$ be the index such that $x \in [t_{i_x}, t_{i_x+1})$, for each $x \in [a,b]$, we have $0 < N_{i,d+1}(x) \le 1$ if $i \in [i_x - d, i_x]$ and $N_{i,d+1}(x) = 0$, otherwise. Therefore, for all $x \in [a,b]$,

$$\begin{aligned}&\left|\frac{1}{n\delta}\boldsymbol{N}^\top(x)(\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}^{-1} - \boldsymbol{H}_{n,\lambda}^{-1})\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}\boldsymbol{\Delta}(\underline{g} - \boldsymbol{Q}\boldsymbol{Z}^*)\right|\\ &\quad \le (n\delta)^{-1}\max_{j=-d,\ldots,K}|T_j|\sum_{j=-d}^{K}\sum_{i=i_x-d}^{i_x}|(\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}^{-1} - \boldsymbol{H}_{n,\lambda}^{-1})_{ij}|\,.\end{aligned}$$

Now using Lemma D.20, we have that if $\mathcal{E}_n$ holds, then for all $x \in [a,b]$,

$$\left|\frac{1}{n\delta}\boldsymbol{N}^\top(x)(\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}^{-1} - \boldsymbol{H}_{n,\lambda}^{-1})\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}\boldsymbol{\Delta}(\underline{g} - \boldsymbol{Q}\boldsymbol{Z}^*)\right|$$

$$
\begin{aligned}
&\leq (n\delta)^{-1} \max_{j=-d,\ldots,K} |T_j| e_m \sum_{j=-d}^{K} \sum_{i=i_x-d}^{i_x} \underline{c}\nu^{|i-j|}(|i-j|+\overline{c}) \\
&\leq (n\delta)^{-1} \max_{j=-d,\ldots,K} |T_j| e_m C\,,
\end{aligned}
$$

for some finite positive constant C that does not depend on n, where the last inequality comes from the fact that an infinite geometric series $\sum_{i=0}^{\infty} \nu^i$ and $\sum_{i=0}^{\infty} i\nu^i$ are both convergent if $0 < \nu < 1$. Thus, using Lemma D.1 and (D.26), we have

$$
\lim_{n\to\infty} \mathbb{P}\left\{ \left| \frac{1}{n\delta} \boldsymbol{N}^\top(x)(\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}^{-1} - \boldsymbol{H}_{n,\lambda}^{-1})\boldsymbol{\mathcal{N}}^\top \boldsymbol{\Phi}\boldsymbol{\Delta}(\underline{g} - \boldsymbol{Q}\boldsymbol{Z}^*) \right| \leq (n\delta)^{-1} \max_{j=-d,\ldots,K} |T_j| e_m \underline{C} \right\} = 1\,, \tag{D.50}
$$

for some finite positive constant $\underline{C}$ that does not depend on n.

Now we calculate $\max_{j=-d,\ldots,K} |T_j|$. Note that for all $j = -d, \ldots, K$, $T_j = \sum_{t=1}^{J} a_{t,j}$, where $a_{t,j} = \sum_{s=1}^{n_t} \varphi_t(q_0)\Delta_{s,t} N_{j,d+1}(X_{s,t})\{g(X_{s,t}) - q_0^{1-n_t} Z_t^*\}$. For all $j = -d, \ldots, K$, $a_{1,j}, \ldots, a_{J,j}$ are independent random variables. To apply Lemma D.3 to evaluate $\max_{j=-d,\ldots,K} |T_j|$, we check the conditions of Lemma D.3. Using Assumptions (A9), (A10) and (3.3), there exists a constant $\mathcal{M}$ such that for all $t = 1, \ldots, J$ and $j = -d, \ldots, K$, $|a_{t,j}| < \mathcal{M}$.

For the variance, since the $(\Delta_{s,t} X_{s,t})$'s are independent and the Z_t^*'s are independent, for $j = -d, \ldots, K$ we have

$$
\begin{aligned}
\mathrm{var}(T_j) &= \mathrm{var}\left[\sum_{t=1}^{J} \sum_{s=1}^{n_t} \varphi_t(q_0)\Delta_{s,t} N_{j,d+1}(X_{s,t})\{g(X_{s,t}) - q_0^{1-n_t} Z_t^*\} \right] \\
&\leq \sum_{t=1}^{J} \varphi_t^2(q_0)\mathbb{E}\left(\left[\sum_{s=1}^{n_t} \Delta_{s,t} N_{j,d+1}(X_{s,t})\{g(X_{s,t}) - q_0^{1-n_t} Z_t^*\} \right]^2 \right) \\
&\leq \sum_{t=1}^{J} \varphi_t^2(q_0)\mathbb{E}\left(n_t \sum_{s=1}^{n_t} \Delta_{s,t}[N_{j,d+1}(X_{s,t})\{g(X_{s,t}) - q_0^{1-n_t} Z_t^*\}]^2 \right) \\
&= \sum_{t=1}^{J} \sum_{s=1}^{n_t} n_t \varphi_t^2(q_0)\mathbb{E}\Big(N_{j,d+1}^2(X_{s,t})\mathbb{E}[\Delta_{s,t}\{g(X_{s,t}) - q_0^{1-n_t} Z_t^*\}^2 | X_{s,t}] \Big)
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{t=1}^{J}\sum_{s=1}^{n_t} n_t\varphi_t^2(q_0)\mathbb{E}\Big[N_{j,d+1}^2(X_{s,t})g(X_{s,t})\pi(X_{s,t})\{q_0^{1-n_t} - g(X_{s,t})\}\Big] \\
&= \sum_{t=1}^{J} n_t^2\varphi_t^2(q_0)\int_a^b N_{j,d+1}^2(x)\pi(x)g(x)\{q_0^{1-n_t} - g(x)\}f_X(x)\,dx\,,
\end{aligned}
$$

where the second last equality is from (D.42).

Let $m = \sup_{t=1,\ldots,J} n_t$. Then under Assumption (A9), $m < \infty$ and $\sup_{t=1,\ldots,J} q_0^{1-n_t} \asymp 1$. Under Assumptions (A5), (A7), (A8), (A10) and by B-spline Property 1, we have, for $j = -d, \ldots, K$,

$$
\begin{aligned}
\mathrm{var}(T_j) &\le m\phi_2^2 n \sup_{t=1,\ldots,J} \|\pi g\big(q_0^{1-n_t} - g\big)\|_{L_\infty[a,b]}\|f_X\|_{L_\infty[a,b]}\|N_{j,d+1}\|_{L_\infty[a,b]}\int_a^b N_{j,d+1}(x)\,dx \\
&\le nm\phi_2^2 \sup_{t=1,\ldots,J} \|\pi g\big(q_0^{1-n_t} - g\big)\|_{L_\infty[a,b]} \times f_{\max} \times 1 \times \frac{t_{j+d+1} - t_j}{d+1} \le c_v n\delta\,,
\end{aligned}
$$

where $c_v = m\phi_2^2 \sup_{t=1,\ldots,J} \|\pi g\big(q_0^{1-n_t} - g\big)\|_{L_\infty[a,b]} f_{\max}$ is a finite positive constant that does not depend on n. Note that under Assumption (A7), $\delta^{-1} \asymp K$. Thus, using Assumption (A6), we have $\mathcal{M}\sqrt{\log K/(c_v n\delta)} = o(1)$. Then the conditions in Lemma D.3 are fulfilled for the sequence T_j, and we have $\max_{j=-d,\ldots,K}|T_j| = \max_{j=-d,\ldots,K}|\mathbb{E}(T_j)| + O_p(\sqrt{(n\delta)\log K})$.

Noting that for all $j = -d, \ldots, K$,

$$
\begin{aligned}
\mathbb{E}(T_j) &= \sum_{t=1}^{J}\sum_{s=1}^{n_t} \varphi_t(q_0)\mathbb{E}[\Delta_{s,t}N_{j,d+1}(X_{s,t})\{g(X_{s,t}) - q_0^{1-n_t}Z_t^*\}] \\
&= \sum_{t=1}^{J}\sum_{s=1}^{n_t} \varphi_t(q_0)\mathbb{E}[N_{j,d+1}(X_{s,t})\{g(X_{s,t})\mathbb{E}(\Delta_{s,t}|X_{s,t}) - \mathbb{E}(\Delta_{s,t}q_0^{1-n_t}Z_t^*|X_{s,t})\}] \\
&= 0\,,
\end{aligned}
$$

where the last equality comes from (D.34), we have

$$
\max_{j=-d,\ldots,K}|T_j| = O_p(\sqrt{n\delta\log K})\,. \tag{D.51}
$$

Recalling the rate of convergence of e_m in Lemma D.21, we have, using Assumption (A7), $(n\delta)^{-1}\max_{j=-d,\dots,K}|T_j|e_m\underline{C} = O_p\{\sqrt{(n\delta)^{-1}}\sqrt{K\log K/n}\} = o_p\{\sqrt{(n\delta)^{-1}}\}$, where we used Assumption (A6). Then observing (D.50) and using Corollary D.4, we obtain that, for all $x \in [a,b]$,

$$\left|\frac{1}{n\delta}\boldsymbol{N}^\top(x)(\widehat{\boldsymbol{H}}^{-1}_{n,\lambda,\varphi} - \boldsymbol{H}^{-1}_{n,\lambda})\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}\boldsymbol{\Delta}(\underline{g} - \boldsymbol{Q}\boldsymbol{Z}^*)\right| = o_p\{\sqrt{(n\delta)^{-1}}\}\,.$$

□

Regarding the third term in (C.15), we derive the following result.

Lemma D.24. *Under Assumption* **A**, *for all* $x \in [a,b]$,

$$\frac{1}{n\delta}\boldsymbol{N}^\top(x)\widehat{\boldsymbol{H}}^{-1}_{n,\lambda,\varphi}\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}\boldsymbol{\Delta}(\underline{s}_g - \underline{g}) = o_p(\delta^{d+1})\,.$$

Proof of Lemma D.24. Let $\tilde{T}_r$ be the $(r+d+1)$th entry of the vector $\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}\boldsymbol{\Delta}(\underline{s}_g - \underline{g})$, for $r = -d,\dots,K$. By B-spline Property 1 and letting $0 \le s_x \le K$ be the index such that $x \in [t_{s_x}, t_{s_x+1})$ for each $x \in [a,b]$, we have $0 < N_{s,d+1}(x) \le 1$ for all $s \in [s_x - d, s_x]$ and $N_{s,d+1}(x) = 0$, otherwise. Then using B-spline Property 1, we have, for all $x \in [a,b]$,

$$\begin{aligned}\left|\frac{1}{n\delta}\boldsymbol{N}^\top(x)\widehat{\boldsymbol{H}}^{-1}_{n,\lambda,\varphi}\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}\boldsymbol{\Delta}(\underline{s}_g - \underline{g})\right| &= \left|(n\delta)^{-1}\sum_{r=-d}^{K}\sum_{s=s_x-d}^{s_x} N_{s,d+1}(x)(\widehat{\boldsymbol{H}}_{n,\lambda,\varphi})_{sr}\tilde{T}_r\right| \\ &\le (n\delta)^{-1}\max_{r=-d,\dots,K}|\tilde{T}_r|\sum_{r=-d}^{K}\sum_{s=s_x-d}^{s_x}|(\widehat{\boldsymbol{H}}_{n,\lambda,\varphi})_{sr}|\,.\end{aligned}$$

Now using Lemma D.18, we have that if $\mathcal{E}_n$ holds, then for all $x \in [a,b]$,

$$\begin{aligned}\left|\frac{1}{n\delta}\boldsymbol{N}^\top(x)\widehat{\boldsymbol{H}}^{-1}_{n,\lambda,\varphi}\boldsymbol{\mathcal{N}}^\top\boldsymbol{\Phi}\boldsymbol{\Delta}(\underline{s}_g - \underline{g})\right| &\le (n\delta)^{-1}\max_{r=-d,\dots,K}|\tilde{T}_r|\frac{4}{\tilde{c}_1}\sum_{r=-d}^{K}\sum_{s=s_x-d}^{s_x}\nu^{|i-j|} \\ &\le (n\delta)^{-1}\max_{r=-d,\dots,K}|\tilde{T}_r|\frac{4}{\tilde{c}_1}C\,,\end{aligned}$$

for some finite positive constant C that does not depend on n, where the last inequality comes from the convergence of the infinite geometric series for $\sup_{n\in\mathbb{N}} |\nu| < 1$. Thus, using Lemma D.1 and (D.26), we have

$$\lim_{n\to\infty} \mathbb{P}\left\{ \left| \frac{1}{n\delta} \boldsymbol{N}^\top(x) \widehat{\boldsymbol{H}}_{n,\lambda,\varphi}^{-1} \boldsymbol{\mathcal{N}}^\top \boldsymbol{\Phi\Delta}(\underline{s}_g - \underline{g}) \right| \leq \underline{C}(n\delta)^{-1} \max_{r=-d,\dots,K} |\tilde{T}_r| \right\} = 1\,, \tag{D.52}$$

for some finite positive constant $\underline{C}$ that does not depend on n.

Note that for all $r = -d, \dots, K$, $\tilde{T}_r = \sum_{j=1}^{J} a_{j,r}$, where $a_{j,r} = \sum_{i=1}^{n_j} \varphi_j(q_0) \Delta_{i,j} N_{r,d+1}(X_{i,j}) \{s_g(X_{i,j}) - g(X_{i,j})\}$. For all $r = -d, \dots, K$, $a_{1,r}, \dots, a_{J,r}$ are independent random variables. To apply Lemma D.3 to evaluate $\max_{r=-d,\dots,K} |\tilde{T}_r|$, we check the conditions of Lemma D.3. Under Assumptions (A9), (A10) and using (C.12), we have, almost surely, $|a_{j,r}| = O(\delta^{d+1})$ as $n \to \infty$. Thus, there exist positive real numbers $\mathcal{M}_1$ and $\tilde{n}_1$, such that for all $n \geq \tilde{n}_1$, $|a_{j,r}| \leq \mathcal{M}_1 \delta^{d+1}$.

Regarding the variance of $\tilde{T}_r$, for $r = -d, \dots, K$, since the $(\Delta_{i,j}, X_{i,j})$'s are independent, under Assumptions (A7), (A8) and using B-spline Properties 1 and 4,

$$\begin{aligned}
\operatorname{var}(\tilde{T}_r) &= \operatorname{var}\left[\sum_{j=1}^{J} \sum_{i=1}^{n_j} \varphi_j(q_0) \Delta_{i,j} N_{r,d+1}(X_{i,j}) \{s_g(X_{i,j}) - g(X_{i,j})\} \right] \\
&= \sum_{j=1}^{J} \sum_{i=1}^{n_j} \varphi_j^2(q_0) \left(\operatorname{var}\left[\Delta_{i,j} N_{r,d+1}(X_{i,j}) \{s_g(X_{i,j}) - g(X_{i,j})\} \right] \right) \\
&\leq \sum_{j=1}^{J} \sum_{i=1}^{n_j} \varphi_j^2(q_0) \mathbb{E}\left[N_{r,d+1}^2(X_{i,j}) \{s_g(X_{i,j}) - g(X_{i,j})\}^2 \right] \\
&= \sum_{j=1}^{J} n_j \varphi_j^2(q_0) \int_a^b N_{r,d+1}^2(x) \{s_g(x) - g(x)\}^2 f_X(x)\, dx \\
&\leq \phi_2^2 n (\|s_g - g\|_{L_\infty[a,b]})^2 \|f_X\|_{L_\infty[a,b]} \|N_{r,d+1}\|_{L_\infty[a,b]} \int_a^b N_{r,d+1}(x)\, dx \\
&\leq \phi_2^2 n (\|s_g - g\|_{L_\infty[a,b]})^2 f_{\max} \times 1 \times \delta\,.
\end{aligned}$$

We deduce from (C.5) with (C.9) and (C.10) that

$$\begin{aligned}\|s_g - g\|_{L_\infty[a,b]} &\le \left\| \sum_{j=0}^{K} \mathbb{1}_{[t_j,t_{j+1})}(x) \frac{g^{(d+1)}(t_j)}{(d+1)!} \delta_j^{d+1} B_{d+1}\left(\frac{x-t_j}{t_{j+1}-t_j}\right) \right\|_{L_\infty[a,b]} + o(\delta^{d+1}) \\ &\le \delta^{d+1} \left\| \frac{g^{(d+1)}(x)}{(d+1)!} \right\|_{L_\infty[a,b]} \sup_{t\in[0,1]} |B_{d+1}(t)| + o(\delta^{d+1}) = O(\delta^{d+1}). \qquad \text{(D.53)}\end{aligned}$$

Therefore, we have for $r = -d, \ldots, K$,

$$\mathrm{var}(\tilde{T}_r) = O(n\delta^{2d+3}).$$

That is, there exist positive real numbers $\mathcal{M}_2$ and an $\tilde{n}_2$ such that for all $n \ge \tilde{n}_2$, $\mathrm{var}(\tilde{T}_r) \le \mathcal{M}_2 n\delta^{2d+3}$. Thus, for all $n \ge \max(\tilde{n}_1, \tilde{n}_2)$, we have $|a_{j,r}| < \mathcal{M}_1\delta^{d+1}$ and $\mathrm{var}(\tilde{T}_r) \le \mathcal{M}_2 n\delta^{2d+3}$ for some positive real numbers $\mathcal{M}_1$ and $\mathcal{M}_2$. Note that $\mathcal{M}_1\delta^{d+1}\sqrt{\log K/(\mathcal{M}_2 n\delta^{2d+3})} = (\mathcal{M}_1/\sqrt{\mathcal{M}_2})\sqrt{\log K/(n\delta)}$. Using Assumption (A7), $\delta^{-1} \asymp K$. Then using Assumption (A6), we have $\mathcal{M}_1\delta^{d+1}\sqrt{\log K/(\mathcal{M}_2 n\delta^{2d+3})} = o(1)$. Thus, there exists a positive real number $\tilde{n}_3$ such that for all $n \ge \tilde{n}_3$, $\mathcal{M}_1\delta^{d+1}\sqrt{\log K/(\mathcal{M}_2 n\delta^{2d+3})} \le 1$. Therefore, the conditions in Lemma D.3 are satisfied with $\tilde{n} = \max(\tilde{n}_1, \tilde{n}_2, \tilde{n}_3)$, $\mathcal{M}_n = \mathcal{M}_1\delta^{d+1}$ and $c_{v,n} = \mathcal{M}_2 n\delta^{2d+3}$, and we have

$$\max_{r=-d,\ldots,K} |\tilde{T}_r| = \max_{r=-d,\ldots,K} |\mathbb{E}(\tilde{T}_r)| + O_p(\sqrt{n\delta^{2d+3}\log K}). \qquad \text{(D.54)}$$

Now, we calculate $\max_{r=-d,\ldots,K} |\mathbb{E}(\tilde{T}_r)|$. Note that for all $r = -d, \ldots, K$,

$$\begin{aligned}\mathbb{E}(\tilde{T}_r) &= \sum_{j=1}^{J}\sum_{i=1}^{n_j} \varphi_j(q_0)\mathbb{E}\big[\Delta_{i,j} N_{r,d+1}(X_{i,j})\{s_g(X_{i,j}) - g(X_{i,j})\}\big] \\ &= \sum_{j=1}^{J}\sum_{i=1}^{n_j} \varphi_j(q_0)\mathbb{E}\big[\mathbb{E}(\Delta_{i,j}|X_{i,j}) N_{r,d+1}(X_{i,j})\{s_g(X_{i,j}) - g(X_{i,j})\}\big] \\ &= \sum_{j=1}^{J}\sum_{i=1}^{n_j} \varphi_j(q_0) \int_a^b \pi(x) N_{r,d+1}(x)\{s_g(x) - g(x)\} f_X(x)\, dx. \qquad \text{(D.55)}\end{aligned}$$

Using (C.7), B-spline Property 4, and under Assumptions (A5) and (A8), we have

$$\begin{aligned}
&\int_a^b \pi(x)N_{r,d+1}(x)\{s_g(x)-g(x)\}f_X(x)\,dx \\
=\;& -\int_a^b \pi(x)N_{r,d+1}(x)\sum_{j=0}^{K}\mathbb{1}_{[t_j,t_{j+1})}(x)\frac{g^{(d+1)}(t_j)}{(d+1)!}\delta_j^{d+1}B_{d+1}\left(\frac{x-t_j}{t_{j+1}-t_j}\right)f_X(x)\,dx \\
&+o(\delta^{d+1})\int_a^b N_{r,d+1}(x)\,dx \\
=\;& -\int_a^b \pi(x)N_{r,d+1}(x)\sum_{j=0}^{K}\mathbb{1}_{[t_j,t_{j+1})}(x)\frac{g^{(d+1)}(t_j)}{(d+1)!}\delta_j^{d+1}B_{d+1}\left(\frac{x-t_j}{t_{j+1}-t_j}\right)f_X(x)\,dx \\
&+o(\delta^{d+2})\,. \qquad\qquad (D.56)
\end{aligned}$$

Regarding the first term in (D.56), using Assumptions (A1) and (A2), and (3.6), we have that π is a $d+1$ times continuously differentiable function on $[a,b]$ and $\|\pi^{(k)}\|_\infty < \infty$ for $j = 0,\ldots,d+1$ and for all $d \geq 1$. By Taylor expansion, for all $j = 0,\ldots,K$ and for all $x \in [a,b]$, there exists a constant $\xi_j \in (0,1)$ such that

$$\mathbb{1}_{[t_j,t_{j+1})}(x)\pi(x) = \mathbb{1}_{[t_j,t_{j+1})}(x)\pi(t_j) + \varepsilon_{\pi_x,j}\,,$$

where $\varepsilon_{\pi_x,j} = \mathbb{1}_{[t_j,t_{j+1})}(x)(x-t_j)\pi'\{t_j+\xi_j(x-t_j)\}$. Thus, under Assumption (A7), we have $\sup_{x\in[a,b],j=0,\ldots,K}|\varepsilon_{\pi_x,j}| = O(\delta)$. Similarly, we have, for all $j = 0,\ldots,K$ and for all $x \in [a,b]$,

$$\mathbb{1}_{[t_j,t_{j+1})}(x)f_X(x) = \mathbb{1}_{[t_j,t_{j+1})}(x)f_X(t_j) + \varepsilon_{f_x,j}\,,$$

where $\sup_{x\in[a,b],j=0,\ldots,K}|\varepsilon_{f_x,j}| = O(\delta)$.

By B-spline Property 1, for all $r = -d,\ldots,K$, $N_{r,d+1}(x) \neq 0$ only if $x \in (t_r, t_{r+d+1})$, so that

$$\int_a^b \pi(x)N_{r,d+1}(x)\sum_{j=0}^{K}\mathbb{1}_{[t_j,t_{j+1})}(x)\frac{g^{(d+1)}(t_j)}{(d+1)!}\delta_j^{d+1}B_{d+1}\left(\frac{x-t_j}{t_{j+1}-t_j}\right)f_X(x)\,dx$$

$$
\begin{aligned}
&= \int_{t_r}^{t_{r+d+1}} N_{r,d+1}(x) \sum_{j=0}^{K} \pi(x) \mathbb{1}_{[t_j,t_{j+1})}(x) \frac{g^{(d+1)}(t_j)}{(d+1)!} \delta_j^{d+1} B_{d+1}\left(\frac{x-t_j}{t_{j+1}-t_j}\right) f_X(x)\, dx \\
&= \sum_{j=\max(r,0)}^{\min(r+d,K)} \int_{t_j}^{t_{j+1}} N_{r,d+1}(x) \delta_j^{d+1} \frac{\pi(t_j) f_X(t_j) g^{(d+1)}(t_j)}{(d+1)!} B_{d+1}\left(\frac{x-t_j}{t_{j+1}-t_j}\right) dx \\
&\quad + \sum_{j=\max(r,0)}^{\min(r+d,K)} \delta_j^{d+1} \frac{g^{(d+1)}(t_j)}{(d+1)!} \int_{t_j}^{t_{j+1}} \varepsilon_{x,j} N_{r,d+1}(x) B_{d+1}\left(\frac{x-t_j}{t_{j+1}-t_j}\right) dx\,, \qquad \text{(D.57)}
\end{aligned}
$$

where $\varepsilon_{x,j} = \pi(t_j)\varepsilon_{f_x,j} + f_X(t_j)\varepsilon_{\pi_x,j} + \varepsilon_{\pi_x,j}\varepsilon_{f_x,j}$, so that $\sup_{x\in[a,b],j=0,\dots,K} |\varepsilon_{x,j}| = O(\delta)$.

In order to determine the order of the first term in (D.57), we use techniques similar to those used for deriving (2.9) of Barrow and Smith (1978). For $r = -d, \dots, K$, $j = r, \dots, r+d$, using integration by parts and the fact that $B'_{d+1}(x) = (d+1)B_d(x)$ (see (2.3) of Barrow and Smith, 1978 for example), we have

$$
\begin{aligned}
&\sum_{j=\max(r,0)}^{\min(r+d,K)} \int_{t_j}^{t_{j+1}} N_{r,d+1}(x) \delta_j^{d+1} \frac{\pi(t_j) f_X(t_j) g^{(d+1)}(t_j)}{(d+1)!} B_{d+1}\left(\frac{x-t_j}{t_{j+1}-t_j}\right) dx \\
&= \sum_{j=\max(r,0)}^{\min(r+d,K)} \pi(t_j) f_X(t_j) g^{(d+1)}(t_j) \left\{ \left[N_{r,d+1}(x) \frac{\delta_j^{d+2}}{(d+2)!} B_{d+2}\left(\frac{x-t_j}{t_{j+1}-t_j}\right)\right]\Bigg|_{t_j}^{t_{j+1}} \right. \\
&\left. - \int_{t_j}^{t_{j+1}} N_{r,d+1}^{(1)}(x) \frac{\delta_j^{d+2}}{(d+2)!} B_{d+2}\left(\frac{x-t_j}{t_{j+1}-t_j}\right) dx \right\}. \qquad \text{(D.58)}
\end{aligned}
$$

Repeating the integration by parts, we find that the integrals are of the form

$$
\int_{t_j}^{t_{j+1}} N_{r,d+1}^{(d)}(x) \frac{\delta_j^{2d+1}}{(2d+1)!} B_{2d+1}\left(\frac{x-t_j}{t_{j+1}-t_j}\right) dx\,.
$$

Since $N_{r,d+1}(x)$ is a degree d spline, it is equal to a polynomial of degree d on each (t_j, t_{j+1}). Thus, $N_{r,d+1}^{(d)}(x)$ is a constant on each (t_j, t_{j+1}). Then, using (2.2) of Barrow and Smith (1978), which states that $\int_0^1 B_k(x)\, dx = 0$ for all $k \geq 1$, we deduce that these integrals are equal to zero. Further, using (2.4) of Barrow and Smith (1978) which states that $B_k(0) = B_k(1)$ for all $k \geq 2$, (D.58) reduces to

$$\sum_{j=\max(r,0)}^{\min(r+d,K)} \sum_{k=1}^{d} (-1)^{k-1} \pi(t_j) f_X(t_j) g^{(d+1)}(t_j) \frac{B_{d+1+k}(0)}{(d+1+k)!} \delta_j^{d+1+k} \{N_{r,d+1}^{(k-1)}(t_{j+1}) - N_{r,d+1}^{(k-1)}(t_j)\}$$
$$= \sum_{j=\max(r,0)}^{\min(r+d,K)} \sum_{k=1}^{d} o(\delta_j^{d+1+k}) = o(\delta^{d+2}),$$

where we used Assumption (A1) with (3.7), Assumptions (A5) and (A8), and the continuous differentiability of B-splines.

Therefore, concerning the first term in (D.57), we have

$$\sum_{j=\max(r,0)}^{\min(r+d,K)} \int_{t_j}^{t_{j+1}} N_{r,d+1}(x) \delta_j^{d+1} \frac{\pi(t_j) f_X(t_j) g^{(d+1)}(t_j)}{(d+1)!} B_{d+1}\left(\frac{x - t_j}{t_{j+1} - t_j}\right) dx = o(\delta^{d+2}).$$

Using (C.9) and (C.10), we have $\sup_{t\in[0,1]} |B_{d+1}(t)| = O(1)$. Thus, using B-spline Property 4, the second term in (D.57) can be bounded by

$$\left| \sum_{j=\max(r,0)}^{\min(r+d,K)} \delta_j^{d+1} \frac{g^{(d+1)}(t_j)}{(d+1)!} \int_{t_j}^{t_{j+1}} \varepsilon_{x,j} N_{r,d+1}(x) B_{d+1}\left(\frac{x - t_j}{t_{j+1} - t_j}\right) dx \right|$$
$$\leq \sum_{j=r}^{r+d} \delta_j^{d+1} \delta_j^{d+1} \frac{g^{(d+1)}(t_j)}{(d+1)!} \sup_{x\in[a,b], j=0,\ldots,K} |\varepsilon_{x,j}| \sup_{t\in[0,1]} |B_{d+1}(t)| \int_{t_j}^{t_{j+1}} N_{r,d+1}(x)\, dx$$
$$= O(\delta^{d+2}) \int_{t_r}^{t_{r+d+1}} N_{r,d+1}(x)\, dx$$
$$= O(\delta^{d+3}).$$

Now, we have (D.57) $= o(\delta^{d+2})$. Thus, (D.56) $= o(\delta^{d+2})$. Note that under Assumption (A9), $\sum_j^J n_j \varphi_j(q_0) = O(n)$. Using (D.55), we have for all $r = -d, \ldots, K$,

$$\mathbb{E}(\tilde{T}_r) = \sum_{j=1}^{J} \sum_{i=1}^{n_j} \varphi_j(q_0) \int_a^b \pi(x) N_{r,d+1}(x) \{s_g(x) - g(x)\} f_X(x)\, dx = o(n\delta^{d+2}). \qquad \text{(D.59)}$$

Thus, combining with (D.54), we have

$$\max_{r=-d,\dots,K} |\tilde{T}_r| = o(nd^{d+2}) + O_p(\sqrt{n\delta^{2d+3}\log K})\,.$$

Then $(n\delta)^{-1}\max_{r=-d,\dots,K}|\tilde{T}_r| = o(\delta^{d+1}) + O_p\{\delta^{d+1}\sqrt{(n\delta)^{-1}\log K}\}$.

Note that under Assumption (A7), $\delta^{-1} \asymp K$. Then using Assumption (A6), we have $O_p\{\delta^{d+1}\sqrt{(n\delta)^{-1}\log K}\} = o_p(\delta^{d+1})$. Thus, $(n\delta)^{-1}\max_{r=-d,\dots,K}|\tilde{T}_r| = o_p(\delta^{d+1})$. Observing (D.52) and using Corollary D.4, the result follows. □

Finally, we need to show the asymptotic property of the last term of (C.15), $\lambda \boldsymbol{N}^\top(x)\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}^{-1}\nabla_\ell^\top \boldsymbol{R}\nabla_\ell\boldsymbol{\eta}/(n\delta)$. In Claeskens et al. (2009), under assumption that the covariate variable is non-random and without any assumption on λ, the authors evaluate a term similar to $\lambda \boldsymbol{N}^\top(x)\widehat{\boldsymbol{H}}_{n,\lambda}^{-1}\nabla_\ell^\top \boldsymbol{R}\nabla_\ell\boldsymbol{\eta}/(n\delta) = b_\lambda(x) + o(\lambda n^{-1}\delta^{-\ell})$, where $\widehat{\boldsymbol{H}}_{n,\lambda}^{-1}$ is equal to $\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}^{-1}$ without the φ_j's and the $\Delta_{i,j}$'s and where $b_\lambda(x)$ is a deterministic term whose order was not studied by those authors.

In our case too, we find that it is difficult to calculate the exact order of our term similar to $b_\lambda(x)$. What we can show though is that, under Assumption (A11), this term is negligible compared to $V_{n,\delta}(x)$. Under other assumptions on λ, the local asymptotic properties of the penalised spline estimator are not clear, even in the context of complete i.i.d. data.

To show that $\lambda \boldsymbol{N}^\top(x)\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}^{-1}\nabla_\ell^\top \boldsymbol{R}\nabla_\ell\boldsymbol{\eta}/(n\delta) = o_p\{\sqrt{V_{n,\delta}(x)}\}$ under Assumption (A11), we first establish the following lemma.

Lemma D.25. *Under Assumptions (A1) to (A8), for $j = -d,\dots,K$, we have*

$$|(\nabla_\ell^\top \boldsymbol{R}\nabla_\ell\boldsymbol{\eta})_j| = O(\delta^{-\ell+1})\,.$$

Proof of Lemma D.25. For $j = -d,\dots,K$, we have

$$|(\nabla_\ell^\top \boldsymbol{R}\nabla_\ell\boldsymbol{\eta})_j| \le \max_{j=-d,\dots,K}|(\nabla_\ell^\top \boldsymbol{R}\nabla_\ell\boldsymbol{\eta})_j| = \|\nabla_\ell^\top \boldsymbol{R}\nabla_\ell\boldsymbol{\eta}\|_\infty \le \|\nabla_\ell^\top\|_\infty\|\boldsymbol{R}\|_\infty\|\nabla_\ell\boldsymbol{\eta}\|_\infty\,. \tag{D.60}$$

We study the behaviour of those three terms separately.

We start by studying $\|\nabla_\ell\|_\infty = \|\nabla_\ell^\top\|_\infty$. Recall that in the proof of Lemma D.12, we have shown (see page 48) that $\nabla_\ell = \nabla^\ell\nabla^{\ell-1}\dots\nabla^1$ where from (D.17), $\nabla^k = \delta^{-1}\boldsymbol{T}_k\boldsymbol{W}_k$ for $k = 1,\dots,\ell$, $\boldsymbol{W}_k$ is a $(K+d+1-k)\times(K+d+2-k)$ matrix defined at (D.16) and $\boldsymbol{T}_k$ is a $(K+d+1-k)\times(K+d+1-k)$ diagonal matrix defined at (D.18). It is easy to see from (D.16) that $\|\boldsymbol{W}_k\|_\infty \le 2$ and $\|\boldsymbol{T}_k\|_\infty = O(1)$. Thus, for $k = 1,\dots,\ell$, we have

$$\|\nabla^k\|_\infty \le \delta^{-1}\|\boldsymbol{T}_k\|_\infty\|\boldsymbol{W}_k\|_\infty = O(\delta^{-1})\,.$$

We deduce that

$$\|\nabla_\ell^\top\|_\infty = \|\nabla_\ell\|_\infty \le \prod_{k=1}^{\ell}\|\nabla^k\|_\infty = O(\delta^{-\ell})\,. \tag{D.61}$$

Next we calculate $\|\boldsymbol{R}\|_\infty$. Using the fact that for any symmetric real matrix $\boldsymbol{A} \in \mathbb{R}^{b\times b}$ where b is a positive integer, $\max_{i,j=1,\dots,b}|(\boldsymbol{A})_{ij}| \le e_{\boldsymbol{A},m}$, where $e_{\boldsymbol{A},m} = \max\{|e_{\boldsymbol{A}}|$ s.t. $e_{\boldsymbol{A}}$ is an eigenvalue of $\boldsymbol{A}\}$ (see Horn and Johnson, 1985, page 315, equation (6,2)), and combining with (D.14), we deduce that $\max_{i,j=-d+\ell,\dots,K}|(\boldsymbol{R})_{ij}| \le \delta(d+1-\ell)^{-1}$.

Moreover, $\boldsymbol{R}$ is a square matrix with elements $(\boldsymbol{R})_{ij} = \int_a^b N_{i,d-\ell+1}(x)N_{j,d-\ell+1}(x)\,dx$ for $i,j = -d+\ell,\dots,K$. Replacing d by $d-\ell$ in B-spline Property 1, $N_{i,d-\ell+1} = 0$ if $x \notin (t_i, t_{i+d-\ell+1})$, for all $i = -d+\ell,\dots,K$. Thereby, $N_{i,d-\ell+1}(x)N_{j,d-\ell+1}(x) = 0$ if $|i-j| > d+1-\ell$, for all $x \in [a,b]$. We have, for all $i,j = -d+\ell,\dots,K$, $(\boldsymbol{R})_{ij} = 0$ if $|i-j| > d-\ell+1$. Then we have

$$\|\boldsymbol{R}\|_\infty = \max_{-d+\ell\le i\le K}\sum_{j=-d+\ell}^{K}|(\boldsymbol{R})_{ij}| \le 2(d-\ell+1)\times\delta(d+1-\ell)^{-1} = 2\delta\,. \tag{D.62}$$

Regarding $\|\nabla_\ell\boldsymbol{\eta}\|_\infty$, in the proof of Theorem 6.25 of Schumaker (1981), it is shown (see inequality (6.50) at page 231) that for every $f \in C^{d+1}[a,b]$ and $\|f^{(d+1)}\|_\infty$, there exist a spline function $s(x) \in S_d(\boldsymbol{t})$ and a constant c independent of δ such that

$$\|f^{(r)} - s^{(r)}\|_{L_\infty[a,b]} \le c\delta^{d+1-r}\|f^{(d+1)}\|_{L_\infty[a,b]}\,, \quad \text{for} \quad r = 0,\dots,d\,.$$

Now, under Assumptions (A1) and (A2) and by (3.7), $g \in C^{d+1}[a,b]$ and $\|g^{(d+1)}\|_{L_\infty[a,b]} < \infty$. Therefore, using the above result, we can find a spline function $\tilde{s}(x) \in S_d(\boldsymbol{t})$ and a constant $\tilde{c}$ independent of δ such that

$$\|g^{(r)} - \tilde{s}^{(r)}\|_{L_\infty[a,b]} \le \tilde{c}\delta^{d+1-r}\|g^{(d+1)}\|_{L_\infty[a,b]} = O(\delta^{d+1-r}), \quad \text{for} \quad r = 0, \ldots, d. \tag{D.63}$$

Taking $r = \ell$ in (D.63), we deduce that $\|\tilde{s}^{(\ell)}\|_{L_\infty[a,b]} = \|g^{(\ell)}\|_{L_\infty[a,b]} + O(\delta^{d+1-\ell})$.

Let $\tilde{\boldsymbol{\eta}}$ be a vector of coefficients such that $\tilde{s}(x) = N^\top(x)\tilde{\boldsymbol{\eta}}$. By B-spline Property 3, under Assumption (A3), $\tilde{s}^{(\ell)}(x) = \boldsymbol{N}_{d-\ell}^\top(x)\nabla_\ell\tilde{\boldsymbol{\eta}} \in S_{d-\ell}(\boldsymbol{t})$. Applying Lemma E.1 with $q = \infty$ and replacing d there by $d - \ell$, we have $\boldsymbol{\Gamma}\nabla_\ell\tilde{\boldsymbol{\eta}}\nabla_\ell\tilde{\boldsymbol{\eta}}$ and

$$\|\tilde{s}^{(\ell)}\|_{L_\infty[a,b]} \le \|\nabla_\ell\tilde{\boldsymbol{\eta}}\|_\infty \le 9^{d-\ell}(2d - 2\ell + 3)\|\tilde{s}^{(\ell)}\|_{L_\infty[a,b]}.$$

Thereby, using Assumption (A3) that $d > \ell$ and that $\delta = o(1)$, we have

$$\|\nabla_\ell\tilde{\boldsymbol{\eta}}\|_\infty = O\{c_s\|\tilde{s}^{(\ell)}\|_{L_\infty[a,b]}\} = O\{c_s\|g^{(\ell)}\|_{L_\infty[a,b]} + O(\delta^{d+1-\ell})\} = O(1). \tag{D.64}$$

Now we deduce that $\|\nabla_\ell\boldsymbol{\eta}\|_\infty = O(1)$ from (D.64). Note that

$$\|\nabla_\ell(\boldsymbol{\eta} - \tilde{\boldsymbol{\eta}})\|_\infty \le \|\nabla_\ell\|_\infty\|\boldsymbol{\eta} - \tilde{\boldsymbol{\eta}}\|_\infty,$$

and that $s_g(x) - \tilde{s}(x) = \boldsymbol{N}^\top(x)(\boldsymbol{\eta} - \tilde{\boldsymbol{\eta}}) \in S_d(\boldsymbol{t})$. Applying Lemma E.1 to $s_g(x) - \tilde{s}(x)$, taking $q = \infty$, we have $\boldsymbol{\Gamma}(\boldsymbol{\eta} - \tilde{\boldsymbol{\eta}}) = \boldsymbol{\eta} - \tilde{\boldsymbol{\eta}}$ and

$$\|\boldsymbol{\eta} - \tilde{\boldsymbol{\eta}}\|_\infty \le C_d^{-1}\|s_g - \tilde{s}\|_{L_\infty[a,b]}.$$

Taking $r = 0$ in (D.63) and using (D.53), we have

$$\|s_g - \tilde{s}\|_{L_\infty[a,b]} \le \|(s_g - g)\|_{L_\infty[a,b]} + \|(g - \tilde{s})\|_{L_\infty[a,b]} = O(\delta^{d+1}).$$

Therefore, using (D.61),

$$\|\nabla_\ell(\boldsymbol{\eta} - \tilde{\boldsymbol{\eta}})\|_\infty \le \|\nabla_\ell\|_\infty\|\boldsymbol{\eta} - \tilde{\boldsymbol{\eta}}\|_\infty = O(\delta^{d+1-\ell}).$$

Combing with (D.64), under Assumption (A3) that $d \geq \ell$ and that $\delta = o(1)$, we have

$$\|\nabla_\ell \boldsymbol{\eta}\|_\infty \leq \|\nabla_\ell \tilde{\boldsymbol{\eta}}\|_\infty + O(\delta^{d+1-\ell}) = O(1)\,.$$

Combining this result with (D.60), (D.61) and (D.62) the result follows. □

Finally, we study the last term in (C.15).

Lemma D.26. *Under Assumption* **A***, for all* $x \in [a,b]$*, we have*

$$\frac{\lambda}{n\delta}\boldsymbol{N}^\top(x)\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}^{-1}\nabla_\ell^\top \boldsymbol{R}\nabla_\ell \boldsymbol{\eta} = o_p\{\sqrt{(n\delta)^{-1}}\}\,,$$

where $K_{n,\lambda}$ *is defined at* (D.19).

Proof of Lemma D.26. Using arguments similar to those we used at page 69, by B-spline Property 1 and letting $0 \leq i_x \leq K$ be the index such that $x \in [t_{i_x}, t_{i_x+1})$ for each $x \in [a,b]$, we have $0 < N_{i,d+1}(x) \leq 1$ for all $i \in [i_x - d, i_x]$ and $N_{i,d+1}(x) = 0$, otherwise. Thus, using B-spline Property 1, if $\mathcal{E}_n$ holds, then for all $x \in [a,b]$,

$$\begin{aligned}\left|\frac{\lambda}{n\delta}\boldsymbol{N}^\top(x)\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}^{-1}\nabla_\ell^\top \boldsymbol{R}\nabla_\ell \boldsymbol{\eta}\right| &= \left|\frac{\lambda}{n\delta}\sum_{j=-d}^{K}\sum_{i=i_x-d}^{i_x} N_{i,d+1}(x)(\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}^{-1})_{ij}(\nabla_\ell^\top \boldsymbol{R}\nabla_\ell \boldsymbol{\eta})_j\right| \\ &\leq U_n\,,\end{aligned}$$

where, using Lemma D.18,

$$\begin{aligned}U_n &= \frac{\lambda}{n\delta}\max_{j=-d,\ldots,K}|(\nabla_\ell^\top \boldsymbol{R}\nabla_\ell \boldsymbol{\eta})_j|\frac{4}{\tilde{c}_1}\sum_{j=-d,\ldots,K}^{K}\sum_{i=i_x-d}^{i_x}\nu^{|i-j|} \\ &= O\{\lambda n^{-1}\delta^{-\ell}\} = O\{\lambda n^{-1/2}\delta^{-\ell+1/2}\sqrt{(n\delta)^{-1}}\}\,,\end{aligned}$$

where the second last equality comes from the convergence of infinite geometric series for $\sup_{n\in\mathbb{N}}|\nu| < 1$.

Using (D.26) and Lemma D.1, we have

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{\lambda}{n\delta}\boldsymbol{N}^\top(x)\widehat{\boldsymbol{H}}_{n,\lambda,\varphi}^{-1}\nabla_\ell^\top \boldsymbol{R}\nabla_\ell \boldsymbol{\eta}\right| \leq U_n\right) = 1\,.$$

Using Assumption (A7), we have $\delta^{-1} \asymp K$. Then under Assumption (A11), $\lambda n^{-1/2}\delta^{-\ell+1/2} \asymp (\lambda K^{-1})^{1/2}(\lambda K^{2\ell}/n)^{1/2} = o(1)$. Thus $U_n = o\{\sqrt{(n\delta)^{-1}}\}$ and using Corollary D.4, the result follows. □

E Known facts about splines

E.1 Definitions

For any integer $d \geq 1$, given an interval $[a, b]$ on the real line and a sequence $\boldsymbol{t} = \{t_{-d}, t_{-d+1}, \ldots, t_{K+d}, t_{K+d+1}\}$ of knots such that $t_{-d} = \ldots = t_0 = a < t_1 < t_2 < \ldots < t_K < b = t_{K+1} = \ldots = t_{K+d+1}$, with K a positive integer, a spline function $s(x)$ of degree d (equivalently, of order $d+1$) with knots $\boldsymbol{t}$ is defined as a function that satisfies the following two conditions:

1. $s(x)$ is a $(d-1)$-times continuously differentiable function on the whole interval $[a, b]$, i.e. $s(x) \in C^{d-1}[a, b]$;

2. on each subinterval $[t_i, t_{i+1}]$, $i = 0, 1, \ldots, K$, s is equal to a polynomial of degree d.

We denote the class of such spline functions by $S_d(\boldsymbol{t})$.

In order to study the properties of our penalised spline estimator, we use a particular basis function for the spline space, called the B-spline basis. Besides $\boldsymbol{t}$, let $a = t_{-d} = t_{-d-1} = \ldots$ and $b = t_{K+d+1} = t_{K+d+2} = \ldots$ be sequences of real numbers, which may be finite or infinite sequences. Given an integer $d \geq 1$ and $i \in \mathbb{Z}$, let the functions $N_{i,1}, N_{i,2}, \ldots$

be defined recursively as follows:

$$N_{i,1}(x) = \begin{cases} 1\,, & \text{if } t_i < x < t_{i+1}\,, \\ 0\,, & \text{otherwise}\,. \end{cases}$$

$$N_{i,d+1}(x) = \frac{x - t_i}{t_{i+d} - t_i} N_{i,d}(x) + \frac{t_{i+d+1} - x}{t_{i+d+1} - t_{i+1}} N_{i+1,d}(x)\,, \tag{E.1}$$

for all $x \in \mathbb{R}$, and where we use the convention $0/0 = 0$. In particular, for all $j \geq 1$, $N_{i,j}(x) = 0$ for $i > K$ and $i < 0$. $N_{i,d+1}$ is called the B-spline of degree d (equivalently, of order $d + 1$) associated with the knots $t_i, \ldots, t_{i+d+1}$, and the sequence $\{N_{i,d+1}\}_{i=-d}^{K}$ forms a basis of $S_d(\boldsymbol{t})$. This recursive definition of B-splines is equivalent to the one using the divided difference operator at page 116 of Schumaker (1981). The proof can be found at page 90 of De Boor (2001). We summarise some useful properties of B-splines from these two books in the following subsection.

E.2 Preliminaries on Splines

One useful property of a B-spline is that it is always non-negative on its support and the set of B-spline basis functions forms a partition of unity:

B-spline Property 1 (Positivity and partition of unity)**.** *The B-splines $N_{i,d+1}(x)$, $i = -d, \ldots, K$ associated with $S_d(\boldsymbol{t})$ are continuous functions on $[a, b]$ and*

$$\begin{cases} 0 < N_{i,d+1}(x) \leq 1\,, & \textit{if } t_i < x < t_{i+d+1}\,, \\ N_{i,d+1}(x) = 0\,, & \textit{otherwise}\,, \end{cases} \tag{E.2}$$

and

$$\sum_{i=-d}^{K} N_{i,d+1}(x) = 1\,, \quad \textit{for all} \;\; a \leq x \leq b\,. \tag{E.3}$$

Proof. See Schumaker (1981), page 116. □

As the penalty term in (3.13) involves a derivative and an integral of the spline, we give the following related properties:

B-spline Property 2 (Differentiation)**.** *Let* $N_{i,d+1}^{(1)}(x)$ *denote the first derivative of* $N_{i,d+1}(x)$. *Then for* $i = -d, \ldots, K$,

$$N_{i,d+1}^{(1)}(x) = d\left\{\frac{N_{i,d}(x)}{t_{i+d} - t_i} - \frac{N_{i+1,d}(x)}{t_{i+d+1} - t_{i+1}}\right\}. \tag{E.4}$$

Proof. See De Boor (2001), page 115. □

This shows that the derivatives of the B-splines also have a recursive structure, which leads to the following property about the derivatives of splines:

B-spline Property 3 (Derivatives of splines)**.** *Consider integers* $d > \ell \geq 0$ *and a spline function* $s(x) = \sum_{i=-d}^{K} \beta_i N_{i,d+1}(x) \in S_d(\boldsymbol{t})$ *with coefficients* β_i, $i = -d, \ldots, K$. *Then we have*

$$s^{(\ell)}(x) = \sum_{i=-d}^{K} \beta_i N_{i,d+1}^{(\ell)}(x) = \sum_{i=-d+\ell}^{K} \beta_i^{(\ell)} N_{i,d+1-\ell}(x)\,, \tag{E.5}$$

where the $\beta_i^{(\ell)}$*'s are defined recursively by:*

$$\begin{aligned} &\beta_i^{(j)} = \frac{(d+1-j)\{\beta_i^{(j-1)} - \beta_{i-1}^{(j-1)}\}}{t_{i+d+1-j} - t_i}\,, \quad \text{for } i = -d+j, \ldots, K \text{ and } j = 1, \ldots, \ell \\ &\textit{and} \;\; \beta_i^{(0)} = \beta_i, \;\; \textit{for} \;\; i = -d, \ldots, K\,. \end{aligned} \tag{E.6}$$

Therefore,

$$s^{(\ell)}(x) \in S_{d-\ell}(\boldsymbol{t})\,. \tag{E.7}$$

Proof. See De Boor (2001), page 117. □

From this property, if we let $\boldsymbol{\beta} = (\beta_{-d}, \ldots, \beta_K)^\top$ denote the vector of coefficients of a spline in the space $S_d(\boldsymbol{t})$ and $\boldsymbol{\beta}^{(\ell)} = (\beta_{-d+l}^{(\ell)}, \ldots, \beta_K^{(\ell)})^\top$, then using (E.6), we can find a matrix ∇_ℓ such that $\nabla_\ell \boldsymbol{\beta} = \boldsymbol{\beta}^{(\ell)}$. Thus, if we let $\boldsymbol{R}$ denote the matrix with entries $(\boldsymbol{R})_{ij} = \int_a^b N_{i,d+1-\ell}(x) N_{j,d+1-\ell}(x)\,dx$, $i,j = -d+l, \ldots, K$, then $\int_a^b \left\{\sum_{i=-d}^{K} \beta_i N_{i,d+\ell}^{(\ell)}(x)\right\}^2 dx = \boldsymbol{\beta}^\top \nabla_\ell^\top \boldsymbol{R} \nabla_\ell \boldsymbol{\beta}$.

Recall from page 11 in Delaigle et al. (2017) that $\boldsymbol{D}_\ell$ is a matrix with entries $(\boldsymbol{D}_\ell)_{ij} = \int_a^b N_{i,d+1}^{(\ell)}(x) N_{j,d+1}^{(\ell)}(x)\,dx$, for $i,j = -d,\ldots,K$. Then the penalty term is such that $\int_a^b \left\{ \sum_{i=-d}^K \beta_i N_{i,d+1}^{(\ell)}(x) \right\}^2 dx = \boldsymbol{\beta}^\top \boldsymbol{D}_\ell \boldsymbol{\beta}$. That is, for any $\boldsymbol{\beta} \in \mathbb{R}^{K+d+1}$, $\boldsymbol{\beta}^\top \nabla_\ell^\top \boldsymbol{R} \nabla_\ell \boldsymbol{\beta} = \boldsymbol{\beta}^\top \boldsymbol{D}_\ell \boldsymbol{\beta}$. Thus, combining with the fact that $\nabla_\ell^\top \boldsymbol{R} \nabla_\ell$ and $\boldsymbol{D}_\ell$ are symmetric matrices, we have

$$\boldsymbol{D}_\ell = \nabla_\ell^\top \boldsymbol{R} \nabla_\ell\,, \tag{E.8}$$

which is a useful decomposition of $\boldsymbol{D}_\ell$, and has been used in Claeskens et al. (2009).

The following property concerns the integral of the B-spline:

B-spline Property 4 (Integration)**.** *For the B-splines associated with $S_d(\boldsymbol{t})$, $i = -d,\ldots,K$,*

$$\int_a^b N_{i,d+1}(x)\,dx = \frac{t_{i+d+1} - t_i}{d+1}\,. \tag{E.9}$$

Proof. See Schumaker (1981), page 128. □

The following lemma concerns the L_q norm, $1 \le q \le \infty$, of the splines, which has been used in Zhou et al. (1998). However, neither Zhou et al. (1998) nor the reference they used proved this result, and so we provide a proof here.

Lemma E.1. *Let $\boldsymbol{\Gamma} = \mathrm{diag}\{(t_1 - t_{-d})^{1/q}, \ldots, (t_{K+d+1} - t_K)^{1/q}\}$ be a diagonal matrix of dimension $(K+d+1)\times(K+d+1)$. For any spline function $s(x) = \sum_{i=-d}^K \beta_i N_{i,d+1}(x) \in S_d(\boldsymbol{t})$ with coefficients β_i, $i = -d,\ldots,K$, and for all $1 \le q \le \infty$, let $C_d = 1/\{9^d(2d+3)\}$, using the convention that $1/\infty = 0$, we have*

$$(d+1)^{-1/q} C_d \|\boldsymbol{\Gamma}\boldsymbol{\beta}\|_q \le \|s\|_{L_q[a,b]} \le (d+1)^{-1/q} \|\boldsymbol{\Gamma}\boldsymbol{\beta}\|_q\,. \tag{E.10}$$

Proof. **Lower bound**. According to (4.79) at page 143 of Schumaker (1981) and Theorem 4.41 at page 145 of Schumaker (1981), for any spline function $s(x) = \sum_{i=-d}^K \beta_i N_{i,d+1}(x) \in$

$S_d(\boldsymbol{t})$ with coefficients β_i, $i = -d, \ldots, K$, and for all $1 \le q \le \infty$, using the convention that $1/\infty = 0$, we have

$$\frac{1}{9^d(2d+3)}(t_{i+d+1} - t_i)^{1/q}|\beta_i| \le \|s\|_{L_q[t_i, t_{i+d+1}]} \quad \text{for all} \quad -d \le i \le K\,. \tag{E.11}$$

Thus, when $q = \infty$, $[1/\{9^d(2d+3)\}]\|\boldsymbol{\Gamma\beta}\|_q = [1/\{9^d(2d+3)\}]\max_{-d\le i\le K}|\beta_i|$ and (E.11) implies that

$$\frac{1}{9^d(2d+3)}\max_{-d\le i\le K}|\beta_i| \le \max_{-d\le i\le K}\|s\|_{L_\infty[t_i,t_{i+d+1})} \le \|s\|_{L_\infty[a,b]}\,.$$

For $1 \le q < \infty$, using (E.11), we have

$$\left\{\frac{1}{9^d(2d+3)}(t_{i+d+1} - t_i)^{1/q}|\beta_i|\right\}^q \le \int_{t_i}^{t_{i+d+1}} |s(x)|^q\,dx \quad \text{for all} \quad -d \le i \le K\,,$$

which leads to

$$\begin{aligned}
&\sum_{i=-d}^{K}\left\{\frac{1}{9^d(2d+3)}(t_{i+d+1} - t_i)^{1/q}|\beta_i|\right\}^q \\
&\le \sum_{i=-d}^{K}\int_{t_i}^{t_{i+d+1}} |s(x)|^q\ dx \\
&= \int_{t_{-d}}^{t_{-d+1}} |s(x)|^q\,dx + \ldots + d\int_{t_{-1}}^{t_0} |s(x)|^q\,dx + (d+1)\sum_{i=0}^{K}\int_{t_i}^{t_{i+1}} |s(x)|^q\ dx \\
&+d\int_{t_{K+1}}^{t_{K+2}} |s(x)|^q\,dx + \ldots + \int_{t_{K+d}}^{t_{K+d+1}} |s(x)|^q\ dx \\
&= (d+1)\int_a^b |s(x)|^q\,dx\,,
\end{aligned}$$

since $t_{-d} = \ldots = t_0 = a < t_1 < \ldots < t_{K+1} = b = t_{k+2} = \ldots = t_{K+d+1}$. Thus, we have

$$\left\{\frac{1}{9^d(2d+3)}\right\}^q \sum_{i=-d}^{K}\{(t_{i+d+1} - t_i)^{1/q}|\beta_i|\}^q \le (d+1)\int_a^b |s(x)|^q\,dx\,,$$

that is

$$\frac{1}{9^d(2d+3)(d+1)^{1/q}}\left[\sum_{i=-d}^{K}\left\{(t_{i+d+1}-t_i)^{1/q}|\beta_i|\right\}^q\right]^{1/q} \leq \left\{\int_a^b |s(x)|^q\,dx\right\}^{1/q}.$$

Using definitions of norms of vectors and functions, the result follows.

Upper bound. When $q=\infty$, $(d+1)^{-1/q}\|\mathbf{\Gamma}\boldsymbol{\beta}\|_q = \|\boldsymbol{\beta}\|_\infty$. By B-spline Property 1 which states that $N_{i,d+1}(x)\geq 0$ for all $i=-d,\ldots,K$ and $\sum_{i=-d}^{K} N_{i,d+1}(x)=1$ for all $x\in[a,b]$, we have

$$\begin{aligned}\|s\|_{L_\infty[a,b]} &= \sup_{x\in[a,b]}\left|\sum_{i=-d}^{K}\beta_i N_{i,d+1}(x)\right| \leq \sup_{x\in[a,b]}\left\{\max_{-d\leq i\leq K}|\beta_i|\sum_{i=-d}^{K}N_{i,d+1}(x)\right\}\\ &= \max_{-d\leq i\leq K}|\beta_i| = \|\beta\|_\infty\,.\end{aligned}$$

For $1\leq q<\infty$, by definition,

$$\left(\|s\|_{L_q[a,b]}\right)^q = \int_a^b\left|\sum_{i=-d}^{K}\beta_i N_{i,d+1}(x)\right|^q dx = \int_a^b\left|\sum_{i=-d}^{K}\beta_i N_{i,d+1}^{1/q}(x)N_{i,d+1}^{1-1/q}(x)\right|^q dx\,.$$

Using Hölder's inequality on the summation, we have

$$\begin{aligned}&\int_a^b\left|\sum_{i=-d}^{K}\beta_i N_{i,d+1}^{1/q}(x)N_{i,d+1}^{1-1/q}(x)\right|^q dx \leq \int_a^b\left(\sum_{i=-d}^{K}\left|\beta_i N_{i,d+1}^{1/q}(x)N_{i,d+1}^{1-1/q}(x)\right|\right)^q dx\\ &\leq \int_a^b\left(\left\{\sum_{i=-d}^{K}\left|\beta_i N_{i,d+1}^{1/q}(x)\right|^q\right\}^{1/q}\left[\sum_{i=-d}^{K}\left\{N_{i,d+1}^{1-1/q}(x)\right\}^{\frac{1}{1-1/q}}\right]^{1-1/q}\right)^q dx\\ &= \int_a^b\left[\left\{\sum_{i=-d}^{K}|\beta_i|^q N_{i,d+1}(x)\right\}^{1/q}\left\{\sum_{i=-d}^{K}N_{i,d+1}(x)\right\}^{1-1/q}\right]^q dx\,.\end{aligned}$$

By (E.3), $\left\{\sum_{i=-d}^{K} N_{i,d+1}(x)\right\}^{1-1/q}=1$ for all $x\in[a,b]$, so that

$$\int_a^b\left[\left\{\sum_{i=-d}^{K}|\beta_i|^q N_{i,d+1}(x)\right\}^{1/q}\left\{\sum_{i=-d}^{K}N_{i,d+1}(x)\right\}^{1-1/q}\right]^q dx$$

$$= \int_a^b \sum_{i=-d}^{K} |\beta_i|^q N_{i,d+1}(x)\, dx = \sum_{i=-d}^{K} |\beta_i|^q \int_a^b N_{i,d+1}(x)\, dx$$
$$= \sum_{i=-d}^{K} |\beta_i|^q \left(\frac{t_{i+d+1} - t_i}{d+1} \right) = (d+1)^{-1} \sum_{i=-d}^{K} \left\{ |\beta_i| (t_{i+d+1} - t_i)^{1/q} \right\}^q ,$$

where the before last equation follows from B-spline Property 4. We deduce that

$$\|s\|_{L_q[a,b]} \le (d+1)^{-1/q} \left[\sum_{i=-d}^{K} \left\{ |\beta_i| (t_{i+d+1} - t_i)^{1/q} \right\}^q \right]^{1/q} = (d+1)^{-1/q} \|\mathbf{\Gamma}\boldsymbol{\beta}\|_q .$$

□

F Technical details for Theorem 2

F.1 Consistency of parametric estimator

To prove Theorem 2, we first establish consistency of $\widehat{\boldsymbol{\theta}}_n$ in the following lemma.

Lemma F.1. *Let $\ell_0(\boldsymbol{\theta}) = \mathbb{E}\{\ell(\boldsymbol{\theta})\}$, where $\ell(\boldsymbol{\theta})$ is defined at page 19 in Delaigle et al. (2017). Under Assumptions (P1) to (P10), if $\boldsymbol{\theta}_0$ uniquely maximises $\ell_0(\boldsymbol{\theta})$ subject to $\boldsymbol{\theta} \in \Theta$, then $\widehat{\boldsymbol{\theta}}_n \xrightarrow{P} \boldsymbol{\theta}_0$ as $n \to \infty$.*

Proof of Lemma F.1. We prove Lemma F.1 by verifying the conditions of Lemma F.2 in section F.3. Condition (i) in Lemma F.2 is one of the assumptions of Lemma F.1 and Condition (ii) is satisfied by Assumption (P2). Condition (iii) is also satisfied because, observing (5.4) and using Assumptions (P1), (P3), (P4) and (P10), we have that $\ell_0(\boldsymbol{\theta})$ is continuous. it remains to verify Condition (iv) of Lemma F.2, which is what we do below.

We prove Condition (iv) in two parts. First we prove that

$$\sup_{\boldsymbol{\theta} \in \Theta} \left| \widehat{\ell}_n(\boldsymbol{\theta}) - \ell(\boldsymbol{\theta}) \right| \xrightarrow{P} 0 \text{ as } n \to \infty \tag{F.1}$$

and then we show that

$$\sup_{\boldsymbol{\theta}\in\Theta}\left|\ell(\boldsymbol{\theta})-\ell_0(\boldsymbol{\theta})\right| \xrightarrow{P} 0 \text{ as } n\to\infty\,. \tag{F.2}$$

Proof of (F.1): Recalling from page 19 of Delaigle et al. (2017) that $\ell(\boldsymbol{\theta}) = \log\{\mathcal{L}(\boldsymbol{\theta}|\mathcal{Y},\Delta,\mathcal{X}_{\text{obs}})\}/J$ with $\mathcal{L}$ defined at (5.4) and $\widehat{\ell}_n(\boldsymbol{\theta})$ defined at (5.8), we see that to prove (F.1), it suffices to show that

$$\sup_{\boldsymbol{\theta}\in\Theta}\left|\widehat{f}_{n,\boldsymbol{\theta}}(Y_j^*,\mathcal{X}_{j,\text{obs}},\Delta_j)-f_{\boldsymbol{\theta}}(Y_j^*,\mathcal{X}_{j,\text{obs}},\Delta_j)\right| \xrightarrow{P} 0 \text{ as } n\to\infty \text{ for all } j=1,\ldots,J\,. \tag{F.3}$$

By definition of $f_{\boldsymbol{\theta}}(Y_j^*,\mathcal{X}_{j,\text{obs}},\Delta_j)$ below (5.4) together with (5.5) and (5.6), and by definition of $\widehat{f}_{n,\boldsymbol{\theta}}(Y_j^*,\mathcal{X}_{j,\text{obs}},\Delta_j)$ below (5.7), we see that to prove (F.3), it suffices to show that for all $i=1,\ldots,n_j$, $j=1,\ldots,J$,

$$\widehat{p}_0\{1-p_{\boldsymbol{\theta}}(X_{i,j})\}\widehat{f}_X(X_{i,j}) \xrightarrow{P} p_0\{1-p_{\boldsymbol{\theta}}(X_{i,j})\}f_X(X_{i,j})\,, \tag{F.4}$$

$$\widehat{p}_1 p_{\boldsymbol{\theta}}(X_{i,j})\widehat{f}_X(X_{i,j}) \xrightarrow{P} p_1 p_{\boldsymbol{\theta}}(X_{i,j})f_X(X_{i,j})\,, \tag{F.5}$$

$$(1-\widehat{p}_0)\int_{\mathbb{R}}\{1-p_{\boldsymbol{\theta}}(x)\}\widehat{f}_X(x)\,dx \xrightarrow{P} (1-p_0)\int_{\mathbb{R}}\{1-p_{\boldsymbol{\theta}}(x)\}f_X(x)\,dx\,, \tag{F.6}$$

$$(1-\widehat{p}_1)\int_{\mathbb{R}}p_{\boldsymbol{\theta}}(x)\widehat{f}_X(x)\,dx \xrightarrow{P} (1-p_1)\int_{\mathbb{R}}p_{\boldsymbol{\theta}}(x)f_X(x)\,dx \tag{F.7}$$

hold uniformly in $\boldsymbol{\theta}\in\Theta$ as $n\to\infty$. Now, using Lemmas C.3 and C.4, since $c_0<c_p$ (see Assumption (P10)), we have that $\widehat{p}_1\xrightarrow{P}p_1$ and $\widehat{p}_0\xrightarrow{P}p_0$ as $n\to\infty$. Therefore, using Assumption (P1), to prove (F.4)–(F.7), it remains to show that as $n\to\infty$,

$$\widehat{f}_X(X_{i,j}) \xrightarrow{P} f_X(X_{i,j}) \text{ for all } i=1,\ldots,n_j,\ j=1,\ldots,J \tag{F.8}$$

and

$$\int_{\mathbb{R}}|\widehat{f}_X(x)-f_X(x)|\,dx = o_p(1)\,. \tag{F.9}$$

To prove these two results, we study asymptotic properties of $\widehat{f}_X(x)=\widehat{f}_{X,\text{obs}}(x)/\widehat{\pi}(x)$, where $\widehat{\pi}(x)=\widehat{p}_1\widehat{p}_{\text{NW}}(x)+\widehat{p}_0\{1-\widehat{p}_{\text{NW}}(x)\}$ (see (5.7)). Recalling from page 17 of Delaigle et al. (2017) that $\widehat{p}_{\mathrm{NW}}(x)=\{1-\widehat{g}_{\mathrm{NW}}(x)\}/\{1+(\widehat{p}_1\widehat{p}_0^{-1}-1)\widehat{g}_{\mathrm{NW}}(x)\}$, we have $\widehat{\pi}(x)=\widehat{p}_1/\{1+(\widehat{p}_1\widehat{p}_0^{-1}-1)\widehat{g}_{\mathrm{NW}}(x)\}$, so that $\widehat{f}_X(x)-f_X(x)=\widehat{f}_{X,\mathrm{obs}}(x)\{1+(\widehat{p}_1\widehat{p}_0^{-1}-1)\widehat{g}_{\mathrm{NW}}(x)\}/\widehat{p}_1-f_X(x)$. Using (3.6) and (3.7) we also have $\pi(x)=p_1/\{1+(p_1p_0^{-1}-1)g(x)\}$ so that

$$\begin{aligned}\widehat{f}_X(x)-f_X(x)=&\widehat{f}_{X,\mathrm{obs}}(x)/\widehat{p}_1-f_X(x)\pi(x)/p_1+\{(\widehat{p}_1\widehat{p}_0^{-1}-1)\widehat{g}_{\mathrm{NW}}(x)\}\widehat{f}_{X,\mathrm{obs}}(x)/\widehat{p}_1\\&-\{(p_1p_0^{-1}-1)g(x)\}f_X(x)\pi(x)/p_1\,. \qquad \text{(F.10)}\end{aligned}$$

Using Lemmas C.3 and C.4, and the fact that $c_0<c_p$ (see Assumption (P10)), we have that $\widehat{p}_1^{-1}=p_1^{-1}\{1+o_p(1)\}$ and $(\widehat{p}_1\widehat{p}_0^{-1}-1)/\widehat{p}_1=\{1+o_p(1)\}(p_1p_0^{-1}-1)/p_1$. Recalling from (5.1) that $\widehat{g}_{\mathrm{NW}}(x)=\widetilde{G}_n(x)/\widehat{f}_{X,\mathrm{obs}}(x)$, where $\widetilde{G}_n(x)=n^{-1}\sum_{j=1}^J\sum_{i=1}^{n_j}\widehat{q}_0^{1-n_j}Z_j^*\Delta_{i,j}L_h(x-X_{i,j})$, we deduce that

$$\begin{aligned}\widehat{f}_X(x)-f_X(x)=&\{\widehat{f}_{X,\mathrm{obs}}(x)-f_X(x)\pi(x)\}/p_1\\&+\{\widetilde{G}_n(x)-g(x)f_X(x)\pi(x)\}(p_1p_0^{-1}-1)/p_1\\&+\{\widehat{f}_{X,\mathrm{obs}}(x)+\widetilde{G}_n(x)\}o_p(1) \qquad \text{(F.11)}\end{aligned}$$

holds uniformly in $x\in\mathbb{R}$.

From there, to prove (F.8) and (F.9), using Assumption (P7), note that $\widehat{f}_{X,\mathrm{obs}}(x)$ and $\widetilde{G}_n(x)$ are uniformly bounded almost surely, and that for $i=1,\ldots,n_j$ and $j=1,\ldots,J$, $\inf_{x\in\mathbb{R}}L\{(x-X_{i,j})/h\}\geq 0$ and $\int_{\mathbb{R}}L\{(x-X_{i,j})/h\}/h\,dx=1$ almost surely. Therefore there exists a constant $c_1>0$ such that $\int_{\mathbb{R}}\{\widehat{f}_{X,\mathrm{obs}}(x)+\widetilde{G}_n(x)\}\,dx\leq c_1\sum_{j=1}^J\sum_{i=1}^{n_j}n^{-1}\int_{\mathbb{R}}L\{(x-X_{i,j})/h\}/h\,dx=O_p(1)$, and using (F.11), we have that for $i=1,\ldots,n_j$ and $j=1,\ldots,J$,

$$\begin{aligned}\widehat{f}_X(X_{i,j})-f_X(X_{i,j}) \;=\;& \{\widehat{f}_{X,\mathrm{obs}}(X_{i,j})-f_X(X_{i,j})\pi(X_{i,j})\}/p_1\\&+\{\widetilde{G}_n(X_{i,j})-g(X_{i,j})f_X(X_{i,j})\pi(X_{i,j})\}(p_1p_0^{-1}-1)/p_1+o_p(1)\end{aligned}$$

and

$$\int_{\mathbb{R}}|\widehat{f}_X(x)-f_X(x)|\,dx \;\leq\; \int_{\mathbb{R}}|\widehat{f}_{X,\mathrm{obs}}(x)-f_X(x)\pi(x)|\,dx/p_1$$

$$+|p_1 p_0^{-1} - 1|/p_1 \int_{\mathbb{R}} |\widetilde{G}_n(x) - g(x) f_X(x)\pi(x)|\, dx + o_p(1)\,.$$

To prove (F.8) and (F.9), it remains to show that

$$\text{for } i = 1, \ldots, n_j,\ j = 1, \ldots, J,\ \widehat{f}_{X,\mathrm{obs}}(X_{i,j}) \xrightarrow{P} f_X(X_{i,j})\pi(X_{i,j}) \text{ as } n \to \infty\,, \tag{F.12}$$

$$\int_{\mathbb{R}} |\widehat{f}_{X,\mathrm{obs}}(x) - f_X(x)\pi(x)|\, dx \xrightarrow{P} 0 \text{ as } n \to \infty\,, \tag{F.13}$$

and that for $i = 1, \ldots, n_j$, $j = 1, \ldots, J$, $\widetilde{G}_n(X_{i,j}) - g(X_{i,j}) f_X(X_{i,j})\pi(X_{i,j}) \xrightarrow{P} 0$ as $n \to \infty$, and $\int_{\mathbb{R}} |\widetilde{G}_n(x) - g(x) f_X(x)\pi(x)|\, dx \xrightarrow{P} 0$ as $n \to \infty$. Next we prove (F.12) and (F.13). The other two results can be proved using the same arguments.

We start by proving (F.12). Let $\widetilde{X}_i = X_{k,\ell}$ and $\widetilde{\Delta}_i = \Delta_{k,\ell}$, for $i = (\ell - 1)k + k$, $k = 1, \ldots, n_\ell$ and $\ell = 1, \ldots, J$. For $i = 1, \ldots, n$, using Chebyshev's inequality, we prove (F.12) by showing that $\mathbb{E}[\{\widehat{f}_{X,\mathrm{obs}}(\widetilde{X}_i) - \pi(\widetilde{X}_i) f_X(\widetilde{X}_i)\}^2] \to 0$ as $n \to \infty$. We have

$$\begin{aligned} &\mathbb{E}[\{\widehat{f}_{X,\mathrm{obs}}(\widetilde{X}_i) - \pi(\widetilde{X}_i) f_X(\widetilde{X}_i)\}^2] \\ &= \mathbb{E}\big[\mathbb{E}\big\{\widehat{f}^2_{X,\mathrm{obs}}(\widetilde{X}_i) - 2\widehat{f}_{X,\mathrm{obs}}(\widetilde{X}_i)\pi(\widetilde{X}_i) f_X(\widetilde{X}_i) + \pi^2(\widetilde{X}_i) f_X^2(\widetilde{X}_i)\big|\widetilde{X}_i\big\}\big]\,. \end{aligned} \tag{F.14}$$

For the second term of (F.14), we have

$$\begin{aligned} \mathbb{E}\{\widehat{f}_{X,\mathrm{obs}}(\widetilde{X}_i)|\widetilde{X}_i\} =& \frac{1}{n}\sum_{k=1}^{n} \mathbb{E}\left[\frac{1}{h} L\{(\widetilde{X}_i - \widetilde{X}_k)/h\}\widetilde{\Delta}_k \middle| \widetilde{X}_i\right] \\ =& \frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{k=1\\k\neq i}}^{n} \mathbb{E}\left[\frac{1}{h} L\{(\widetilde{X}_i - \widetilde{X}_k)/h\}\pi(\widetilde{X}_k) \middle| \widetilde{X}_i\right] + \frac{1}{nh} L(0)\pi(\widetilde{X}_i) \\ =& \frac{n-1}{n}\int_{\mathbb{R}} L(z)\pi(\widetilde{X}_i - hz) f_X(\widetilde{X}_i - hz)\, dz + \frac{1}{nh} L(0)\pi(\widetilde{X}_i)\,. \end{aligned}$$

Using Assumptions (P3) and (P5), and a first order Taylor expansion, there exists a real number $\xi_{z,\widetilde{X}_i,n}$ between $\widetilde{X}_i - hz$ and $\widetilde{X}_i$ such that

$$\mathbb{E}\{\widehat{f}_{X,\mathrm{obs}}(\widetilde{X}_i)|\widetilde{X}_i\} = \frac{n-1}{n}\int_{\mathbb{R}} L(z)\{\pi(\widetilde{X}_i) f_X(\widetilde{X}_i) - hz(\pi f_X)'(\xi_{z,\widetilde{X}_i,n})\}\, dz + \frac{1}{nh} L(0)\pi(\widetilde{X}_i)\,.$$

Therefore, under Assumptions (P3), (P5), (P7) and (P8), we have almost surely that

$$\mathbb{E}\{\widehat{f}_{X,\mathrm{obs}}(\widetilde{X}_i)|\widetilde{X}_i\} = \pi(\widetilde{X}_i)f_X(\widetilde{X}_i) + R_n\,, \tag{F.15}$$

where $|R_n| \le n^{-1}\|f_X\|_{L_\infty} + c_2 h + (nh)^{-1}L(0)$ with $c_2 = \|(\pi f_X)'\|_{L_\infty}\int_{\mathbb{R}}|z|L(z)\,dz$, and thus $\mathbb{E}(R_n) = o(1)$. Therefore, the left hand side of (F.14) satisfies

$$\mathbb{E}[\{\widehat{f}_{X,\mathrm{obs}}(\widetilde{X}_i) - \pi(\widetilde{X}_i)f_X(\widetilde{X}_i)\}^2] = \mathbb{E}\big[\mathbb{E}\{\widehat{f}^2_{X,\mathrm{obs}}(\widetilde{X}_i) - \pi^2(\widetilde{X}_i)f^2_X(\widetilde{X}_i)|\widetilde{X}_i\}\big] + o(1)\,. \tag{F.16}$$

Next, to prove that (F.16) tends to zero, we study the first term on the right hand side of (F.16). We have

$$\begin{aligned}
\mathbb{E}\{\widehat{f}^2_{X,\mathrm{obs}}(\widetilde{X}_i)|\widetilde{X}_i\} \;&=\; \frac{1}{n^2h^2}\sum_{\substack{k=1\\k\neq i}}^{n}\mathbb{E}\left\{L^2\left(\frac{\widetilde{X}_i-\widetilde{X}_k}{h}\right)\widetilde{\Delta}_k^2\,\middle|\,\widetilde{X}_i\right\}\\
&\quad+\frac{1}{n^2h^2}\sum_{\substack{\ell=1\\ \ell\neq i}}^{n}\sum_{\substack{k=1\\k\neq i,\ell}}^{n}\mathbb{E}\left\{L\left(\frac{\widetilde{X}_i-\widetilde{X}_k}{h}\right)L\left(\frac{\widetilde{X}_i-\widetilde{X}_\ell}{h}\right)\widetilde{\Delta}_k\widetilde{\Delta}_\ell\,\middle|\,\widetilde{X}_i\right\}\\
&\quad+\frac{2}{n^2h^2}\sum_{\substack{k=1\\k\neq i}}^{n}L(0)\pi(\widetilde{X}_i)\mathbb{E}\left\{L\left(\frac{\widetilde{X}_i-\widetilde{X}_k}{h}\right)\widetilde{\Delta}_k\,\middle|\,\widetilde{X}_i\right\}\\
&\quad+\frac{1}{n^2h^2}L^2(0)\pi(\widetilde{X}_i)\\
&\equiv\; I_1+I_2+I_3+I_4\,,
\end{aligned}$$

where by Assumption (P5), $I_4 \le (nh)^{-2}L(0)$ almost surely. Using arguments similar to those used to obtain (F.15), we also have, almost surely, $|I_3| \le 2c_2 n^{-1}L(0) + 2(nh)^{-1}L(0)\|f_X\|_{L_\infty}$ and

$$\begin{aligned}
I_1 &= \frac{n-1}{n^2h}\int_{\mathbb{R}}L^2(z)\pi(\widetilde{X}_i-hz)f_X(\widetilde{X}_i-hz)\,dz \le (nh)^{-1}\|f_XL\|_{L_\infty}\,,\\
I_2 &= \frac{(n-1)(n-2)}{n^2}\left\{\int_{\mathbb{R}}L(z)\pi(\widetilde{X}_i-hz)f_X(\widetilde{X}_i-hz)\,dz\right\}^2 = \pi^2(\widetilde{X}_i)f^2_X(\widetilde{X}_i)+\widetilde{R}_n\,,
\end{aligned}$$

where $|\widetilde{R}_n| \le c_3 n^{-1}\|f_X\|^2_{L_\infty} + c_2 h\|f_X\|_{L_\infty} + c_2^2 h^2$ with c_2 defined below (F.15) and for some positive constant c_3. Using Assumptions (P3), (P7) and (P8), we conclude from (F.16) that

$$\mathbb{E}[\{\widehat{f}_{X,\mathrm{obs}}(\widetilde{X}_i) - \pi(\widetilde{X}_i) f_X(\widetilde{X}_i)\}^2] = o(1).$$

This proves (F.12).

Next we prove (F.13). For this we write

$$\begin{aligned}\int_{\mathbb{R}} |\widehat{f}_{X,\mathrm{obs}}(x) - \pi(x) f_X(x)|\, dx \quad &\le \quad \int_{\mathbb{R}} |\widehat{f}_{X,\mathrm{obs}}(x) - \mathbb{E}\{\widehat{f}_{X,\mathrm{obs}}(x)\}|\, dx \\ &\qquad + \int_{\mathbb{R}} |\mathbb{E}\{\widehat{f}_{X,\mathrm{obs}}(x)\} - \pi(x) f_X(x)|\, dx\,. \qquad \text{(F.17)}\end{aligned}$$

Now we have

$$\begin{aligned}\mathbb{E}\{\widehat{f}_{X,\mathrm{obs}}(x)\} &= \int_{\mathbb{R}} L(z)\pi(x-hz) f_X(x-hz)\, dz \\ &= \pi(x) f_X(x) - h\int_{\mathbb{R}}\int_0^1 zL(z)(\pi f_X)'(x-hz\xi)\, d\xi\, dz\,,\end{aligned}$$

where we have used a Taylor expansion with integral form of the remainder. Using Assumptions (P7) to (P9), we deduce that $\int_{\mathbb{R}} |\mathbb{E}\{\widehat{f}_{X,\mathrm{obs}}(x)\} - \pi(x) f_X(x)|\, dx = o(1)$.

Next we prove that the first term on the right hand side of (F.17) satisfies $\int_{\mathbb{R}} |\widehat{f}_{X,\mathrm{obs}}(x) - \mathbb{E}\{\widehat{f}_{X,\mathrm{obs}}(x)\}|\, dx = o_p(1)$. Letting $L_h(x) = L(x/h)/h$ and $\mu_n(x) = n^{-1}\sum_{i=1}^n \gamma(x-\widetilde{X}_i)\widetilde{\Delta}_i$, where γ is the Dirac delta function, we have

$$\begin{aligned}\int_{\mathbb{R}} \big|\widehat{f}_{X,\mathrm{obs}}(x) - \mathbb{E}\{\widehat{f}_{X,\mathrm{obs}}(x)\}\big|\, dx \quad &= \quad \int_{\mathbb{R}} \big|(\mu_n * L_h)(x) - \{(\pi f_X) * L_h\}(x)\}\big|\, dx \\ &\le \quad \int_{\mathbb{R}} \big|(\mu_n * \widetilde{L}_h)(x) - \{(\pi f_X) * \widetilde{L}_h\}(x)\}\big|\, dx \\ &\qquad + \int_{\mathbb{R}} (\mu_n * |L_h - \widetilde{L}_h|)(x)\, dx \\ &\qquad + \int_{\mathbb{R}} \{(\pi f_X) * |L_h - \widetilde{L}_h|\}(x)\, dx\,, \qquad \text{(F.18)}\end{aligned}$$

where $\widetilde{L}(x) = L(x) \cdot 1\{x \in (-h^{-1}, h^{-1})\}$.

For the second and third terms of (F.18), using Young's inequality for convolutions and Assumptions (P7) and (P8), we have

$$\begin{aligned}
\int_{\mathbb{R}} (\mu_n * |L_h - \widetilde{L}_h|)(x)\, dx + \int_{\mathbb{R}} \{(\pi f_X) * |L_h - \widetilde{L}_h|\}(x)\, dx \\
&\le \left\{ 1 + \int_{\mathbb{R}} (\pi f_X)(x)\, dx \right\} \int_{\mathbb{R}} |L_h(x) - \widetilde{L}_h(x)|\, dx \\
&\le 4 \int_{h^{-1}}^{\infty} L(x)\, dx \\
&= o(1)\,. \qquad \text{(F.19)}
\end{aligned}$$

Next we prove that the first term of (F.18) satisfies

$$\int_{\mathbb{R}} \big| (\mu_n * \widetilde{L}_h)(x) - \{(\pi f_X) * \widetilde{L}_h\}(x)\} \big|\, dx = o_p(1)\,. \qquad \text{(F.20)}$$

Letting $\widehat{h}(x) = (\mu_n * \widetilde{L}_h)(x) - \{(\pi f_X) * \widetilde{L}_h\}(x)$ and using Chebyshev's inequality, to prove this we show that $\mathbb{E}\big[\{ \int_{\mathbb{R}} |\widehat{h}(x)|\, dx\}^2\big] = o(1)$. By the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
\mathbb{E}\left[\left\{ \int_{\mathbb{R}} |\widehat{h}(x)|\, dx \right\}^2\right] &= \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{E}|\widehat{h}(x)\widehat{h}(y)|\, dx\, dy \\
&\le \int_{\mathbb{R}} \int_{\mathbb{R}} \sqrt{\mathbb{E}\{\widehat{h}^2(x)\}\mathbb{E}\{\widehat{h}^2(y)\}}\, dx\, dy \\
&= \left\{ \int_{\mathbb{R}} \sqrt{\mathbb{E}\{\widehat{h}^2(x)\}}\, dx \right\}^2. \qquad \text{(F.21)}
\end{aligned}$$

Therefore, to prove (F.20) it suffices to show that $\int_{\mathbb{R}} \sqrt{\mathbb{E}\{\widehat{h}^2(x)\}}\, dx = \int_{\mathbb{R}} \big[\mathrm{var}\{(\mu_n * \widetilde{L}_h)(x)\}\big]^{1/2}\, dx = o(1)$. We have

$$\begin{aligned}
\mathrm{var}\{(\mu_n * \widetilde{L}_h)(x)\} &= \frac{1}{n^2 h^2} \sum_{i=1}^{n} \mathrm{var}\left\{ \widetilde{L}\left(\frac{x - \widetilde{X}_i}{h} \right) \widetilde{\Delta}_i \right\} \\
&\le \frac{1}{n^2 h^2} \sum_{i=1}^{n} \mathbb{E}\left\{ \widetilde{L}^2\left(\frac{x - \widetilde{X}_i}{h} \right) \widetilde{\Delta}_i^2 \right\}
\end{aligned}$$

$$
\begin{aligned}
&= (nh)^{-1} \int_{\mathbb{R}} \widetilde{L}^2(z)\pi(x-zh)f_X(x-zh)\,dz \\
&= (nh)^{-1} \int_{-h^{-1}}^{h^{-1}} L^2(z)\pi(x-zh)f_X(x-zh)\,dz\,.
\end{aligned}
$$

Using the mean value theorem for definite integrals and Assumptions (P3) and (P5), we have that $\forall x \in \mathbb{R}$,

$$
\int_{-h^{-1}}^{h^{-1}} L^2(z)\pi(x-zh)f_X(x-zh)\,dz = \pi(x-c_x)f_X(x-c_x)\int_{-h^{-1}}^{h^{-1}} L^2(z)\,dz\,, \tag{F.22}
$$

for some $c_x \in [-1,1]$. Therefore,

$$
\int_{\mathbb{R}} \big[\mathrm{var}\{(\mu_n * \widetilde{L}_h)(x)\}\big]^{1/2}\,dx = (nh)^{-1/2}\int_{\mathbb{R}} \sqrt{\pi(x-c_x)f_X(x-c_x)\int_{-h^{-1}}^{h^{-1}} L^2(z)\,dz}\,dx = o(1)\,,
$$

where, for all $x \in \mathbb{R}$, $c_x \in [-1,1]$ and where we used Assumptions (P7) to (P9). This completes the proof of (F.20).

Combining (F.18), (F.19) and (F.20) we deduce that $\int_{\mathbb{R}} |\widehat{f}_{X,\mathrm{obs}}(x) - \mathbb{E}\{\widehat{f}_{X,\mathrm{obs}}(x)\}|\,dx = o_p(1)$, which, using (F.17) and the calculations below that equation, proves (F.13). This completes the proof of (F.1).

Proof of (F.2): Recalling the definition of ℓ at page 19 in Delaigle et al. (2017), and noting that the $(Y_j^*, \mathcal{X}_{j,\mathrm{obs}}, \Delta_j)$'s are independent, we have

$$
\begin{aligned}
\mathrm{var}\{\ell(\boldsymbol{\theta})\} &= \frac{1}{J^2}\sum_{j=1}^{J} \mathrm{var}\big\{\log f_{\boldsymbol{\theta}}(Y_j^*, \mathcal{X}_{j,\mathrm{obs}}, \Delta_j)\big\} \\
&\le \frac{1}{J^2}\sum_{j=1}^{J} \mathbb{E}\Big[\big\{\log f_{\boldsymbol{\theta}}(Y_j^*, \mathcal{X}_{j,\mathrm{obs}}, \Delta_j)\big\}^2\Big] \\
&= O(J^{-1})\,,
\end{aligned} \tag{F.23}
$$

where the last equality comes from the fact under Assumptions (P1), (P3) and (P10), $f_{\boldsymbol{\theta}}(Y_j^*, \mathcal{X}_{j,\mathrm{obs}}, \Delta_j)$ is positive and bounded almost surely for $j = 1, \ldots, J$ and all $\boldsymbol{\theta} \in \Theta$.

Recalling that by definition, $\ell_0(\boldsymbol{\theta}) = \mathbb{E}\{\ell(\boldsymbol{\theta})\}$, this implies that for all $\boldsymbol{\theta} \in \Theta$, we have $\ell(\boldsymbol{\theta}) - \ell_0(\boldsymbol{\theta}) \xrightarrow{P} 0$ as $n \to \infty$.

Now under Assumptions (P1) and (P10), we have

$$\mathbb{P}\{\inf_{\boldsymbol{\theta}\in\Theta} f_{\boldsymbol{\theta}}(Y_j^*, \mathcal{X}_{j,\text{obs}}, \mathit{\Delta}_j) > 0\} = 1\,, \tag{F.24}$$

for $j = 1, \ldots, J$, so that, using the Lipschitz continuity of the log function, we have that for $j = 1, \ldots, J$ and for all $\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}} \in \Theta$, there exists a constant $c_4 > 0$ such that almost surely $|\log\{f_{\boldsymbol{\theta}}(Y_j^*, \mathcal{X}_{j,\text{obs}}, \mathit{\Delta}_j)\} - \log\{f_{\widetilde{\boldsymbol{\theta}}}(Y_j^*, \mathcal{X}_{j,\text{obs}}, \mathit{\Delta}_j)\}| \le c_4|f_{\boldsymbol{\theta}}(Y_j^*, \mathcal{X}_{j,\text{obs}}, \mathit{\Delta}_j) - f_{\widetilde{\boldsymbol{\theta}}}(Y_j^*, \mathcal{X}_{j,\text{obs}}, \mathit{\Delta}_j)|$

Using Assumption (P4), and noting that sums and products of bounded Hölder continuous functions are also Hölder continuous, we have that the $f_{\boldsymbol{\theta},j}$'s are Hölder continuous in $\boldsymbol{\theta} \in \Theta$ almost surely. That is, for $j = 1, \ldots, J$ and for all $\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}} \in \Theta$, there exist constants $\alpha > 0$ and $c_5 > 0$ such that almost surely $|f_{\boldsymbol{\theta}}(Y_j^*, \mathcal{X}_{j,\text{obs}}, \mathit{\Delta}_j) - f_{\widetilde{\boldsymbol{\theta}}}(Y_j^*, \mathcal{X}_{j,\text{obs}}, \mathit{\Delta}_j)| \le c_5\|\boldsymbol{\theta} - \widetilde{\boldsymbol{\theta}}\|^\alpha$. Using Assumption (P2) and applying Lemma F.3, this proves (F.2). □

F.2 Proof of Theorem 2

Using Assumption (P1) and recalling the definition of $\widehat{\ell}_n$ at (5.8), we have that $\widehat{\ell}_n(\boldsymbol{\theta})$ is almost surely twice continuously differentiable on Θ. Since, under Assumption (P2), $\boldsymbol{\theta}_0 \in \text{interior}(\Theta)$, and by definition, $\widehat{\boldsymbol{\theta}}$ maximises $\widehat{\ell}_n(\boldsymbol{\theta})$, then applying a Taylor expansion of order zero to $\nabla_{\boldsymbol{\theta}}\widehat{\ell}_n(\widehat{\boldsymbol{\theta}})$ around $\boldsymbol{\theta}_0$, we have

$$0 = \nabla_{\boldsymbol{\theta}}\widehat{\ell}_n(\widehat{\boldsymbol{\theta}}) = \nabla_{\boldsymbol{\theta}}\widehat{\ell}_n(\boldsymbol{\theta}_0) + (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^\top \nabla_{\boldsymbol{\theta}\boldsymbol{\theta}}\widehat{\ell}_n(\bar{\boldsymbol{\theta}})\,, \tag{F.25}$$

for some $\bar{\boldsymbol{\theta}}$ between $\boldsymbol{\theta}_0$ and $\widehat{\boldsymbol{\theta}}$.

Shortly we will prove that

$$\lim_{n\to\infty} \mathbb{P}\{\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}}\widehat{\ell}_n(\bar{\boldsymbol{\theta}}) \text{ is nonsingular}\} = 1\,, \tag{F.26}$$

which, with (F.25) and Lemma D.1, leads to

$$\lim_{n\to\infty} \mathbb{P}\big[\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = -\{\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}}\widehat{\ell}_n(\bar{\boldsymbol{\theta}})\}^{-1}\nabla_{\boldsymbol{\theta}}\widehat{\ell}_n(\boldsymbol{\theta}_0)\big] = 1\,.$$

This implies that $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = -\{\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}}\widehat{\ell}_n(\bar{\boldsymbol{\theta}})\}^{-1}\nabla_{\boldsymbol{\theta}}\widehat{\ell}_n(\boldsymbol{\theta}_0) + o_p(1)$, so that to prove the theorem, it suffices to show that

$$\sqrt{J}\Sigma_n^{1/2}\{\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}}\widehat{\ell}_n(\bar{\boldsymbol{\theta}})\}^{-1}\nabla_{\boldsymbol{\theta}}\widehat{\ell}_n(\boldsymbol{\theta}_0) \xrightarrow{D} N(0, I_{d_p}) \tag{F.27}$$

as $n \to \infty$.

Now we prove (F.26). Recalling the definition of Σ_n in the statement of the theorem, using Assumption (P11), Lemma D.9 and the fact that, for $j = 1, \dots, J$, $\mathbb{E}[\nabla_{\boldsymbol{\theta}} \log\{f_j(\boldsymbol{\theta}_0)\}\nabla_{\boldsymbol{\theta}} \log\{f_j(\boldsymbol{\theta}_0)\}^\top]$ does not depend on n, we have that

$$\Sigma_n \;\text{ is nonsingular }\; \text{for all } n \text{ and as } n \to \infty\,. \tag{F.28}$$

Moreover, using Assumptions (P1) and properties of Fisher information matrices, we have $\mathbb{E}[\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}} \log\{f_j(\boldsymbol{\theta}_0)\}] = \mathbb{E}[\nabla_{\boldsymbol{\theta}} \log\{f_j(\boldsymbol{\theta}_0)\}\nabla_{\boldsymbol{\theta}} \log\{f_j(\boldsymbol{\theta}_0)\}^\top]$ for $j = 1, \dots, J$, which implies that $\mathbb{E}\{\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}}\ell(\boldsymbol{\theta}_0)\} = \Sigma_n$. Therefore, since, by Lemma F.1, $\widehat{\boldsymbol{\theta}} \xrightarrow{P} \boldsymbol{\theta}_0$ as $n \to \infty$ and $\bar{\boldsymbol{\theta}}$ is between $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$, if the following two equations hold as $n \to \infty$:

$$\sup_{\boldsymbol{\theta}\in\Theta} \|\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}}\widehat{\ell}_n(\boldsymbol{\theta}) - \nabla_{\boldsymbol{\theta}\boldsymbol{\theta}}\ell(\boldsymbol{\theta})\|_2 \xrightarrow{P} 0\,, \tag{F.29}$$

$$\sup_{\boldsymbol{\theta}\in\Theta} \|\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}}\ell(\boldsymbol{\theta}) - \mathbb{E}\{\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}}\ell(\boldsymbol{\theta})\}\|_2 \xrightarrow{P} 0\,, \tag{F.30}$$

then we have that

$$\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}}\widehat{\ell}_n(\bar{\boldsymbol{\theta}}) - \Sigma_n \xrightarrow{P} 0 \tag{F.31}$$

as $n \to \infty$, so that using (F.28) we can deduce that (F.26) holds.

Next we prove (F.29). Let $\widehat{f}_j(\boldsymbol{\theta}) = \widehat{f}_{n,\boldsymbol{\theta}}(Y_j^*, \mathcal{X}_{j,\mathrm{obs}}, \Delta_j)$ defined below (5.7). Recalling the definition of $f_j(\boldsymbol{\theta})$ in the statement of the theorem, the definition of $\widehat{\ell}_n$ at (5.8) and of

ℓ at page 19 in Delaigle et al. (2017), we have

$$\begin{aligned}\sup_{\boldsymbol\theta\in\Theta}\|\nabla_{\boldsymbol{\theta\theta}}\widehat{\ell}_n(\boldsymbol\theta)-\nabla_{\boldsymbol{\theta\theta}}\ell(\boldsymbol\theta)\|_2 \;\leq\; & \frac{1}{J}\sum_{j=1}^{J}\sup_{\boldsymbol\theta\in\Theta}\frac{\|\nabla_{\boldsymbol{\theta\theta}}\widehat{f}_j(\boldsymbol\theta)-\nabla_{\boldsymbol{\theta\theta}}f_j(\boldsymbol\theta)\|_2}{\widehat{f}_j(\boldsymbol\theta)}\\
&+\sup_{\boldsymbol\theta\in\Theta}\|\nabla_{\boldsymbol{\theta\theta}}f_j(\boldsymbol\theta)\|_2\left\|\frac{1}{\widehat{f}_j(\boldsymbol\theta)}-\frac{1}{f_j(\boldsymbol\theta)}\right\|_2\\
&+\sup_{\boldsymbol\theta\in\Theta}\frac{\|\nabla_{\boldsymbol\theta}\widehat{f}_j(\boldsymbol\theta)+\nabla_{\boldsymbol\theta}f_j(\boldsymbol\theta)\|_2\|\nabla_{\boldsymbol\theta}\widehat{f}_j(\boldsymbol\theta)-\nabla_{\boldsymbol\theta}f_j(\boldsymbol\theta)\|_2}{\widehat{f}_j^2(\boldsymbol\theta)}\\
&+\sup_{\boldsymbol\theta\in\Theta}\|\nabla_{\boldsymbol\theta}f_j(\boldsymbol\theta)\|_2^2\left\|\frac{1}{\widehat{f}_j^2(\boldsymbol\theta)}-\frac{1}{f_j^2(\boldsymbol\theta)}\right\|_2.\end{aligned}$$

Since, by definition, $\widehat{f}_j(\boldsymbol\theta)=\widehat{f}_{n,\boldsymbol\theta}(Y_j^*,\mathcal{X}_{j,\text{obs}},\Delta_j)$ and $f_j(\boldsymbol\theta)=f_{\boldsymbol\theta}(Y_j^*,\mathcal{X}_{j,\text{obs}},\Delta_j)$ for $j=1,\ldots,J$, where $f_{\boldsymbol\theta}$ is defined below (5.4) and $\widehat{f}_{n,\boldsymbol\theta}$ below (5.7), then using Assumptions (P1) and (P10), and equation (F.24), we have almost surely that, for $j=1,\ldots,J$,

$$\sup_{\boldsymbol\theta\in\Theta}f_j^{-1}(\boldsymbol\theta),\ \sup_{\boldsymbol\theta\in\Theta}\widehat{f}_j^{-1}(\boldsymbol\theta),\ \sup_{\boldsymbol\theta\in\Theta}\big\{\|\nabla_{\boldsymbol\theta}\widehat{f}_j(\boldsymbol\theta)\|_2+\|\nabla_{\boldsymbol\theta}f_j(\boldsymbol\theta)\|_2+\|\nabla_{\boldsymbol{\theta\theta}}f_j(\boldsymbol\theta)\|_2\big\}\text{ are finite}. \tag{F.32}$$

Therefore, to show (F.29) it suffices to show that, for $j=1,\ldots,J$,

$$\sup_{\boldsymbol\theta\in\Theta}\left|\frac{1}{\widehat{f}_j(\boldsymbol\theta)}-\frac{1}{f_j(\boldsymbol\theta)}\right|\xrightarrow{P}0,\qquad \sup_{\boldsymbol\theta\in\Theta}\left|\frac{1}{\widehat{f}_j^2(\boldsymbol\theta)}-\frac{1}{f_j^2(\boldsymbol\theta)}\right|\xrightarrow{P}0, \tag{F.33}$$

$$\sup_{\boldsymbol\theta\in\Theta}\|\nabla_{\boldsymbol\theta}\widehat{f}_j(\boldsymbol\theta)-\nabla_{\boldsymbol\theta}f_j(\boldsymbol\theta)\|_2\xrightarrow{P}0\quad\text{and}\quad\sup_{\boldsymbol\theta\in\Theta}\|\nabla_{\boldsymbol{\theta\theta}}\widehat{f}_j(\boldsymbol\theta)-\nabla_{\boldsymbol{\theta\theta}}f_j(\boldsymbol\theta)\|_2\xrightarrow{P}0 \tag{F.34}$$

as $n\to\infty$. Using (F.3) and (F.32), we have that (F.33) holds. Moreover, using Assumption (P1) and arguments similar to the ones that lead to (F.3), but applied to $\nabla_{\boldsymbol\theta}f_j(\boldsymbol\theta)$ and $\nabla_{\boldsymbol{\theta\theta}}f_j(\boldsymbol\theta)$ instead of $f_j(\boldsymbol\theta)$, we have that (F.34) holds. This concludes the proof of (F.29).

Next we prove (F.30) using Lemma F.3. Recalling the definition of ℓ at page 19 in Delaigle et al. (2017) and using Assumption (P1), we have that ℓ is twice continuously differentiable. Thus $\nabla_{\boldsymbol{\theta\theta}}\ell(\boldsymbol\theta)$ is continuous and so is $\mathbb{E}\{\nabla_{\boldsymbol{\theta\theta}}\ell(\boldsymbol\theta)\}$. Now $\nabla_{\boldsymbol{\theta\theta}}\ell(\boldsymbol\theta)=J^{-1}\sum_{j=1}^{J}[f_j(\boldsymbol\theta)\nabla_{\boldsymbol{\theta\theta}}f_j(\boldsymbol\theta)-\{\nabla_{\boldsymbol\theta}f_j(\boldsymbol\theta)\}\{\nabla_{\boldsymbol\theta}f_j(\boldsymbol\theta)\}^\top]/f_j^2(\boldsymbol\theta)$. Recall that $\boldsymbol\theta\in\Theta\subset\mathbb{R}^{d_p}$, with d_p

a positive finite integer, and let $\{\nabla_{\boldsymbol{\theta\theta}}\ell\}_k(\boldsymbol{\theta})$ be the kth element of $\nabla_{\boldsymbol{\theta\theta}}\ell(\boldsymbol{\theta})$ for $k = 1, \ldots, d_p$. Under Assumptions (P1), (P3) and (P10), using the same argument as the one leads to (F.23), we have that for all $\boldsymbol{\theta} \in \Theta$, $\mathrm{var}[\{\nabla_{\boldsymbol{\theta\theta}}\ell\}_k(\boldsymbol{\theta})] = O(J^{-1})$, for $k = 1, \ldots, d_p$. Therefore, we have $\nabla_{\boldsymbol{\theta\theta}}\ell(\boldsymbol{\theta}) - \mathbb{E}\{\nabla_{\boldsymbol{\theta\theta}}\ell(\boldsymbol{\theta})\} \overset{P}{\to} 0$ as $n \to \infty$ for all $\boldsymbol{\theta} \in \Theta$. Moreover, using (F.24) and Assumption (P4), we have that there exist constants $\alpha > 0$ and $c > 0$ such that for all $\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}} \in \Theta$, $\|\nabla_{\boldsymbol{\theta\theta}}\ell(\boldsymbol{\theta}) - \nabla_{\boldsymbol{\theta\theta}}\ell(\widetilde{\boldsymbol{\theta}})\|_2 \leq c\|\boldsymbol{\theta} - \widetilde{\boldsymbol{\theta}}\|_2^\alpha$ almost surely. Using Assumption (P2) and applying Lemma F.3, this proves that (F.30) holds.

Now that we have proved (F.29) and (F.30), we conclude that (F.31) holds, which, as discussed under (F.31), proves (F.26). To prove the theorem, it remains to prove (F.27). To do this, we show that

$$\Sigma_n\{\nabla_{\boldsymbol{\theta\theta}}\widehat{\ell}_n(\bar{\boldsymbol{\theta}})\}^{-1} \quad \overset{P}{\to} \quad I_{d_p} \tag{F.35}$$

and

$$\sqrt{J}\Sigma_n^{-1/2}\nabla_{\boldsymbol{\theta}}\widehat{\ell}_n(\boldsymbol{\theta}_0) \quad \overset{D}{\to} \quad N(0, I_{d_p}) \tag{F.36}$$

as $n \to \infty$.

To prove (F.35), recalling the definition of Σ_n in the statement of the theorem and (F.32), we have that $\|\Sigma_n\|_2 = O(1)$. Moreover, it follows from (F.26), (F.28), (F.31), the continuity of the inverse transformation of matrices and the continuous mapping theorem that $\{\nabla_{\boldsymbol{\theta\theta}}\widehat{\ell}_n(\bar{\boldsymbol{\theta}})\}^{-1} - \Sigma_n^{-1} \overset{P}{\to} 0$. Therefore, (F.35) holds.

To prove (F.36), using the same argument as the one leading to (F.29), we have that $\|\nabla_{\boldsymbol{\theta}}\widehat{\ell}_n(\boldsymbol{\theta}_0) - \nabla_{\boldsymbol{\theta}}\ell(\boldsymbol{\theta}_0)\|_2 \overset{P}{\to} 0$, as $n \to \infty$. Therefore, (F.36) is proved if we show that

$$\sqrt{J}\Sigma_n^{-1/2}\nabla_{\boldsymbol{\theta}}\ell(\boldsymbol{\theta}_0) \quad \overset{D}{\to} \quad N(0, I_{d_p}) \tag{F.37}$$

as $n \to \infty$. We prove (F.37) using a Lyapunov-type multidimensional central limit theorem (Bentkus, 2005) as follows. Let $S = \sum_{i=1}^J \boldsymbol{X}_j$ with $\{\boldsymbol{X}_j, j = 1, \ldots, J\}$ a sequence of independent random vectors s.t. $\mathbb{E}(S) = 0$ and $C \equiv \mathrm{cov}(S)$ is invertible.

If $\sum_{j=1}^{J} \mathbb{E}\|C^{-1/2}\boldsymbol{X}_j\|_2^3 \xrightarrow{P} 0$ as $J \to \infty$, then $S \xrightarrow{D} N(0, C)$ as $J \to \infty$. Writing $\sqrt{J}\Sigma_n^{-1/2}\nabla_{\boldsymbol{\theta}}\ell(\boldsymbol{\theta}_0) = \sum_{j=1}^{J} J^{-1/2}\Sigma_n^{-1/2}\nabla_{\boldsymbol{\theta}} \log\{f_j(\boldsymbol{\theta}_0)\}$, we apply the central limit theorem above to show (F.37) by letting $\boldsymbol{X}_j = J^{-1/2}\Sigma_n^{-1/2}\nabla_{\boldsymbol{\theta}} \log\{f_j(\boldsymbol{\theta}_0)\}$ for $j = 1, \ldots, J$.

Since $\boldsymbol{\theta}_0$ maximises $\mathbb{E}\{\ell(\boldsymbol{\theta})\}$ and $\boldsymbol{\theta}_0 \in \text{interior}(\Theta)$, we have $\mathbb{E}\{\nabla_{\boldsymbol{\theta}}\ell(\boldsymbol{\theta}_0)\} = 0$. Moreover, recalling the definition of Σ_n in the statement of the theorem, we have $\text{cov}\big\{\sqrt{J}\Sigma_n^{-1/2}\nabla_{\boldsymbol{\theta}}\ell(\boldsymbol{\theta}_0)\big\} = I_{d_p}$. Now, to show that $\sum_{i=1}^{J} \mathbb{E}\|C^{-1/2}\boldsymbol{X}_j\|_2^3 \xrightarrow{P} 0$ as $J \to \infty$, using (F.28), we have $\big\|\Sigma_n^{-1/2}\big\|_2^3 = O(1)$. Moreover, using (F.32), we have $\mathbb{E}\|\nabla_{\boldsymbol{\theta}} \log\{f_j(\boldsymbol{\theta}_0)\}\|_2^3 < \infty$. Therefore, using Assumption (A9), $\sum_{j=1}^{J} \mathbb{E}\|J^{-1/2}\Sigma_n^{-1/2}\nabla_{\boldsymbol{\theta}} \log\{f_j(\boldsymbol{\theta}_0)\}\|_2^3 = O(n^{-1/2}) \xrightarrow{P} 0$ as $n \to \infty$. Using the central limit theorem, this proves (F.37).

F.3 Auxiliary lemmas

The following lemmas are established (Lemma F.2) or follow directly (Lemma F.3) from Newey and McFadden (1994), and will be used to establish the asymptotic properties of $\widehat{\boldsymbol{\theta}}_n$:

Lemma F.2 (Newey and McFadden, 1994, Theorem 2.1)**.** *If there is a function $\ell_0 : \Theta \to \mathbb{R}$ such that (i) $\ell_0(\boldsymbol{\theta})$ is uniquely maximized at $\boldsymbol{\theta}_0$; (ii) Θ is compact; (iii) $\ell_0(\boldsymbol{\theta})$ is continuous; (iv) $\widehat{\ell}_n(\boldsymbol{\theta})$ converges uniformly in probability to $\ell_0(\boldsymbol{\theta})$ as $n \to \infty$, then as $n \to \infty$, $\widehat{\boldsymbol{\theta}} \xrightarrow{P} \boldsymbol{\theta}_0$, where $\widehat{\boldsymbol{\theta}}$ maximizes $\widehat{\ell}_n(\boldsymbol{\theta})$.*

Lemma F.3 (Corollary of Newey and McFadden, 1994, Lemma 2.9)**.** *Suppose $\ell_0, \ell_n : \mathbb{R}^{d_p} \to \mathbb{R}^{d_p'}$ for some finite integer $d_p, d_p' \geq 1$. If Θ is compact, $\ell_0(\boldsymbol{\theta})$ is continuous, $\ell_n(\boldsymbol{\theta}) \xrightarrow{P} \ell_0(\boldsymbol{\theta})$ as $n \to \infty$ for all $\boldsymbol{\theta} \in \Theta$, and there exist $\alpha > 0$ and $\widehat{B}_n = O_p(1)$ such that for all $\tilde{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \Theta$, $\|\ell_n(\boldsymbol{\theta}) - \ell_n(\tilde{\boldsymbol{\theta}})\|_2 \leq \widehat{B}_n\|\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}\|_2^{\alpha}$ almost surely, then $\sup_{\boldsymbol{\theta}\in\Theta} \|\ell_n(\boldsymbol{\theta}) - \ell_0(\boldsymbol{\theta})\|_2 \xrightarrow{P} 0$ as $n \to \infty$.*

References

Barrow, D. and Smith, P. (1978). Asymptotic properties of best $L_2[0, 1]$ approximation by splines with variable knots. *Q. J. Mech. Appl. Math.*, **36**, 293–304.

Bentkus, V. (2005). A Lyapunov-type bound in Rd. *Theory Probab. Appl.*, **49**, 311–323.

Claeskens, G., Krivobokova, T., and Opsomer, J.D. (2009). Asymptotic properties of penalized spline estimators. *Biometrika*, **96**, 529–544.

De Boor, C. (2001). *A Practical Guide to Splines*. Springer-Verlag, Berlin.

Delaigle, A., Hall, P., and Wishart, J. (2014). New approaches to nonparametric and semiparametric regression for univariate and multivariate group testing data. *Biometrika*, **101**, 567–585.

Delaigle, A., Huang, W., and Lei, S. (2017). Estimation of conditional prevalence from group testing data with missing covariates. *Submitted manuscript corresponding to this appendix*.

Delaigle, A. and Meister, A. (2011). Nonparametric regression analysis for group testing data. *J. Amer. Statist. Assoc.*, **106**, 640–650.

Demko, S. (1977). Inverses of band matrices and local convergence of spline projections. *SIAM J. Numer. Anal.*, **14**, 616–619.

Demko, S., Moss, W.F., and Smith, P.W. (1984). Decay rates for inverses of band matrices. *Math. Comput.*, **43**, 491–499.

Eubank, R.L. (1999). *Nonparametric Regression and Spline Smoothing*. Marcel Dekker, New York.

Holland, A.D. (2017). Penalized spline estimation in the partially linear model. *J. Multivar. Anal.*, **153**, 211–235.

Horn, R.A. and Johnson, C.R. (1985). *Matrix Analysis*. Cambridge University Press, Cambridge.

Huang, J.Z. (2003). Local asymptotics for polynomial spline regression. *Ann. Statist.*, **31**, 1600–1635.

Kulkarni, D., Schmidt, D., and Tsui, S.-K. (1999). Eigenvalues of tridiagonal pseudo-Toeplitz matrices. *Linear Algebra Appl.*, **297**, 63–80.

Lehmer, D.H. (1940). On the maxima and minima of Bernoulli polynomials. *Am. Math. Mon.*, **47**, 533–538.

Newey, W.K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of Econom.*, **4**, 2011–2245.

Pollard, D. (2012). *Convergence of Stochastic Processes*. Springer-Verlag, New York.

Ruppert, D., Wand, M.P., and Carroll, R.J. (2003). *Semiparametric Regression.* Cambridge University Press, Cambridge.

Schumaker, L.L. (1981). *Spline Functions: Basic Theory*. Cambridge University Press, Cambridge.

Speckman, P. (1981). The asymptotic integrated mean square error for smoothing noisy data by splines. *Unpublished manuscript*.

Speckman, P. (1985). Spline smoothing and optimal rates of convergence in nonparametric regression models. *Ann. Statist.*, **13**, 970–983.

Tauchen, G. (1985). Diagnostic testing and evaluation of maximum likelihood models. *J. Econometrics*, **30**, 415–443.

Utreras, F. (1983). Natural spline functions, their associated eigenvalue problem. *Numer. Math.*, **42**, 107–117.

Van der Vaart, A.W. (2000). *Asymptotic Statistics*. Cambridge University Press, Cambridge.

Zhou, S., Shen, X., and Wolfe, D. (1998). Local asymptotics for regression splines and confidence regions. *Ann. Statist.*, **26**, 1760–1782.