Econometrica Supplementary Material

SUPPLEMENT TO "NONPARAMETRIC INSTRUMENTAL VARIABLE ESTIMATION OF STRUCTURAL QUANTILE EFFECTS": PROOFS AND EXAMPLE

(*Econometrica*, Vol. 80, No. 4, July 2012, 1533–1562)

BY PATRICK GAGLIARDINI AND OLIVIER SCAILLET

THIS SUPPLEMENTAL MATERIAL contains the proofs of Lemmas A.1 and A.2 in Section SM.1 (ill-posedness), and Lemma A.3 in Section SM.2 (consistency). Section SM.2 also establishes existence of the Q-TiR estimator. Section SM.3 provides the Frechet derivative of operator $\mathcal{A}$ and a characterization of its adjoints. Lemmas A.4–A.11 are proved in Section SM.4 (asymptotic distribution) and Lemmas A.12–A.15 are proved in Section SM.5 (estimation of the asymptotic variance). In Section SM.6, we characterize the asymptotic MISE. In Section SM.7, we provide an example of a NIVQR model, and derive the spectrum of $\tilde{A}A$ (Remark 1), the spectrum of A^*A (Remark 2), and the asymptotic behavior of the variance function (Remark 3). To streamline the presentation, we gather the proofs of the secondary technical Lemmas B.1–B.15 and C.1–C.10 at the end of this technical report (Section SM.8). Equation labels that do not include SM refer to the main paper. To simplify the proofs, we adopt a product kernel in the estimation of the density of (X, Y, Z) in $\mathbb{R}^d$. We use the generic notation K for both the d-dimensional product kernel and each of its components. We take C as a generic constant.

SM.1. ILL-POSEDNESS

SM.1.1. *Proof of Lemma A.1*

Take a sequence (φ_n) as in condition (ii). Since $\varphi_n \nrightarrow \varphi_0$, without loss of generality (w.l.o.g.) we can assume that $\|\varphi_n - \varphi_0\| \geq \varepsilon$ for some $0 < \varepsilon < r$ (otherwise take a subsequence). Since $\mathcal{A}$ is compact (condition (i)) and (φ_n) is bounded, the sequence $\mathcal{A}(\varphi_n)$ admits a convergent subsequence $\mathcal{A}(\varphi_{m(n)}) \to \xi$. Since the weak limit is unique, we have $\xi = \mathcal{A}(\varphi_0)$. Thus $\mathcal{A}(\varphi_{m(n)}) \to \mathcal{A}(\varphi_0)$ and $Q_\infty(\varphi_{m(n)}) \to 0$, but $\|\varphi_{m(n)} - \varphi_0\| \geq \varepsilon$, hence the stated result follows.

SM.1.2. *Proof of Lemma A.2*

We have to show that $\mathcal{A}$ maps closed sets into relatively compact sets. Let $S \subset L^2[0,1]$ be bounded. We have to prove that the closure of $\mathcal{A}(S) \subset L^2(F_Z, \tau)$ is compact. We can equivalently use $\|\cdot\|_{L^2(F_Z,\tau)}$ or $\|\cdot\|_{L^2(F_Z)}$. Proposition 2.24 in Alt (1992) states that $\mathcal{A}(S)$ is relatively compact if and

 DOI: 10.3982/ECTA7937

only if

$$\sup_{\varphi \in S} \|\mathcal{A}(\varphi)\|_{L^2(F_Z)} < \infty, \tag{SM.1}$$

$$\sup_{\varphi \in S} \|\mathcal{A}(\varphi)(\cdot + h) - \mathcal{A}(\varphi)\|_{L^2(F_Z)} \to 0 \quad \text{as} \quad |h| \to 0, \tag{SM.2}$$

and

$$\sup_{\varphi \in S} \big\| \mathcal{A}(\varphi) \cdot \chi_{\mathbb{R}^{d_Z} \setminus B_R(0)} \big\|_{L^2(F_Z)} \to 0 \quad \text{as} \quad R \nearrow \infty, \tag{SM.3}$$

where $\chi_{\mathbb{R}^{d_Z} \setminus B_R(0)}(z) := 1\{z \in \mathbb{R}^{d_Z} \setminus B_R(0)\}$ and $B_R(0)$ is a ball in $\mathbb{R}^{d_Z}$ of radius R around 0. To prove (SM.1), notice that for any z,

$$\begin{aligned} |\mathcal{A}(\varphi)(z)| &= \int_{\mathcal{X}} f_{X|Z}(x|z) F_{U|X,Z}\big(g^{-1}(x, \varphi(x))|x, z\big)\, dx \\ &\le \int_{\mathcal{X}} f_{X|Z}(x|z)\, dx = 1. \end{aligned} \tag{SM.4}$$

Thus $\|\mathcal{A}(\varphi)\|_{L^2(F_Z)} \le 1$ for any $\varphi \in L^2[0,1]$ and (SM.1) follows. To prove (SM.2), we use

$$\begin{aligned} &|\mathcal{A}(\varphi)(z+h) - \mathcal{A}(\varphi)(z)| \\ &\quad \le \int_{\mathcal{X}} \big| f_{X|Z}(x|z+h) - f_{X|Z}(x|z) \big| F_{U|X,Z}\big(g^{-1}(x, \varphi(x))|x, z+h\big)\, dx \\ &\qquad + \int_{\mathcal{X}} f_{X|Z}(x|z) \big| F_{U|X,Z}\big(g^{-1}(x, \varphi(x))|x, z+h\big) \\ &\qquad - F_{U|X,Z}\big(g^{-1}(x, \varphi(x))|x, z\big) \big|\, dx \\ &\quad \le C|h|, \end{aligned}$$

where $C := \sup_{u,x,z} |\nabla_z F_{U|X,Z}(u|x,z)| f_{X|Z}(x|z) + \sup_{x,z} |\nabla_z f_{X|Z}(x|z)| < \infty$ from Assumption A.3(ii) and (iii). Thus we get

$$\|\mathcal{A}(\varphi)(\cdot + h) - \mathcal{A}(\varphi)\|_{L^2(F_Z)} \le C|h| \to 0 \quad \text{as} \quad h \to 0, \tag{SM.5}$$

uniformly in $\varphi \in L^2[0,1]$. Thus, (SM.2) is proved. Finally, from (SM.4), we get that for $\varphi \in L^2[0,1]$,

$$\big\| \mathcal{A}(\varphi) \cdot \chi_{\mathbb{R}^{d_Z} \setminus B_R(0)} \big\|^2_{L^2(F_Z)} \le \int_{\mathbb{R}^{d_Z} \setminus B_R(0)} f_Z(z)\, dz \to 0 \quad \text{as} \quad R \nearrow \infty.$$

This implies (SM.3) and that $\mathcal{A}$ is compact.

SM.2. CONSISTENCY

SM.2.1. *Existence of the Q-TiR Estimator*

Since $Q_T(\varphi) = \frac{1}{T\tau(1-\tau)} \sum_{t=1}^T \hat{m}(\varphi, Z_t)^2$ is positive, a function $\hat{\varphi} \in \Theta$ minimizes $Q_T(\varphi) + \lambda_T \|\varphi\|_H^2$ if and only if

$$\hat{\varphi} = \arg\inf_{\varphi\in\Theta} Q_T(\varphi) + \lambda_T \|\varphi\|_H^2 \quad \text{such that } \lambda_T \|\varphi\|_H^2 \leq L_T(\varphi_0). \tag{SM.6}$$

The solution $\hat{\varphi}$ in (SM.6) exists P-a.s. since (i) mapping $\varphi \to \|\varphi\|_H^2$ is lower semicontinuous on $H^l[0,1]$ w.r.t. the norm $\|\cdot\|$ (see Reed and Simon (1980, p. 358)) and mapping $\varphi \to Q_T(\varphi)$ is continuous on Θ w.r.t. the norm $\|\cdot\|$, P-a.s., for any T, and (ii) set $\{\varphi \in \Theta : \|\varphi\|_H^2 \leq \bar{L}\}$ is compact w.r.t. the norm $\|\cdot\|$ for any constant $0 < \bar{L} < \infty$ (compact embedding theorem; see Adams and Fournier (2003)).

SM.2.2. *Proof of Lemma A.3*

(i) We use the following lemma, which is proved by extending an argument in Hansen (2008).

LEMMA B.1: *Under Assumptions* A.1, A.2, *and* A.3(iii),

$$\sup_{x\in[0,1], y\in\mathbb{R}, z\in\mathcal{Z}} \left|\hat{f}_{X|Z}(x|z)\hat{F}_{Y|X,Z}(y|x,z) - f_{X|Z}(x|z)F_{Y|X,Z}(y|x,z)\right|^2 = O_p(a_T),$$

where $a_T := \frac{\log T}{T h_T^{d_Z+1}} + h_T^{2m}$.

We have $Q_T(\varphi_0) - Q_\infty(\varphi_0) = \frac{1}{T\tau(1-\tau)} \sum_{t=1}^T \Delta\hat{m}(\varphi_0, Z_t)^2$, where $\Delta\hat{m}(\varphi, \cdot) := \hat{m}(\varphi, \cdot) - m(\varphi, \cdot)$. Furthermore,

$$\begin{aligned} |\Delta\hat{m}(\varphi, \cdot)| &\leq \int_{\mathcal{X}} \left|\hat{f}_{X|Z}(x|\cdot)\hat{F}_{Y|X,Z}(\varphi(x)|x,\cdot) \right. \\ &\qquad \left. - f_{X|Z}(x|\cdot)F_{Y|X,Z}(\varphi(x)|x,\cdot)\right| dx \\ &\leq \sup_{x\in[0,1], y\in\mathbb{R}} \left|\hat{f}_{X|Z}(x|\cdot)\hat{F}_{Y|X,Z}(y|x,\cdot) \right. \\ &\qquad \left. - f_{X|Z}(x|\cdot)F_{Y|X,Z}(y|x,\cdot)\right| \end{aligned} \tag{SM.7}$$

uniformly in $\varphi \in \Theta$. Then (i) follows from Lemma B.1.

(ii) Using $\hat{m}(\varphi,\cdot)=\Delta\hat{m}(\varphi,\cdot)+m(\varphi,\cdot)$, we have

$$\begin{aligned}Q_T(\varphi)-Q_\infty(\varphi)&=\frac{1}{T\tau(1-\tau)}\sum_{t=1}^{T}\Delta\hat{m}(\varphi,Z_t)^2\\&\quad+\left\{\frac{1}{T\tau(1-\tau)}\sum_{t=1}^{T}m(\varphi,Z_t)^2-Q_\infty(\varphi)\right\}\\&\quad+2\frac{1}{T\tau(1-\tau)}\sum_{t=1}^{T}\Delta\hat{m}(\varphi,Z_t)m(\varphi,Z_t).\end{aligned}$$

From (SM.7) and Lemma B.1, the first term in the RHS is $O_p(a_T)$, uniformly in $\varphi\in\Theta$. By Cauchy–Schwarz inequality, we get that the third term in the RHS is $O_p(\sqrt{a_T}(\frac{1}{T\tau(1-\tau)}\sum_{t=1}^{T}m(\varphi,Z_t)^2)^{1/2})$, uniformly in $\varphi\in\Theta$. Thus, the conclusion follows if we show that $I_T:=\sup_{\varphi\in\Theta}|\frac{1}{T\tau(1-\tau)}\sum_{t=1}^{T}m(\varphi,Z_t)^2-Q_\infty(\varphi)|=O_p(\frac{1}{\sqrt{T}})$. To bound I_T, we use $m(\varphi,z)=\int_{\mathcal{X}}[F_{U|X,Z}(g^{-1}(x,\varphi(x))|x,z)-\tau]\times f_{X|Z}(x|z)\,dx$. Then

$$\begin{aligned}&\frac{1}{T}\sum_{t=1}^{T}m(\varphi,Z_t)^2-E[m(\varphi,Z)^2]\\&=\int_{\mathcal{X}}\int_{\mathcal{X}}\frac{1}{T}\sum_{t=1}^{T}\Big\{f_{X|Z}(x|Z_t)f_{X|Z}(\xi|Z_t)\\&\quad\times\big[F_{U|X,Z}\big(g^{-1}(x,\varphi(x))|x,Z_t\big)-\tau\big]\\&\quad\times\big[F_{U|X,Z}\big(g^{-1}(\xi,\varphi(\xi))|\xi,Z_t\big)-\tau\big]\\&\quad-E\Big[f_{X|Z}(x|Z)f_{X|Z}(\xi|Z)\\&\quad\times\big[F_{U|X,Z}\big(g^{-1}(x,\varphi(x))|x,Z\big)-\tau\big]\\&\quad\times\big[F_{U|X,Z}\big(g^{-1}(\xi,\varphi(\xi))|\xi,Z\big)-\tau\big]\Big]\Big\}\,dx\,d\xi.\end{aligned}$$

We get $I_T\leq\frac{1}{\tau(1-\tau)\sqrt{T}}\sup_{\varrho\in[0,1]^4}|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(a(Z_t,\varrho)-E[a(Z,\varrho)])|$, where $a(z,\varrho):=f_{X|Z}(x|z)f_{X|Z}(\xi|z)[F_{U|X,Z}(u|x,z)-\tau][F_{U|X,Z}(v|\xi,z)-\tau]$, $\varrho:=(x,\xi,u,v)\in[0,1]^4$. Using Assumption A.3(ii) and (iii), function a is bounded and Lipschitz w.r.t. ϱ: $|a(\cdot,\varrho_1)-a(\cdot,\varrho_2)|\leq C|\varrho_1-\varrho_2|$ for a constant C. By Andrews (1994, Theorem 2), the family $\mathcal{F}:=\{a(\cdot,\varrho):\varrho\in[0,1]^4\}$ satisfies the Pollard entropy condition. By Andrews (1994, Theorem 1), the empirical process $\nu_T(\varrho):=\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(a(Z_t,\varrho)-E[a(Z,\varrho)])$, $\varrho\in[0,1]^4$, is stochastically equicontinuous. Since we can apply a CLT for any $\varrho\in[0,1]^4$, by the fundamental convergence result for empirical processes (see, e.g., Andrews (1994, p. 2251)), $\nu_T(\cdot)$ converges weakly. By the continuous mapping theorem, $\sup_{\varrho\in[0,1]^4} |\nu_T(\varrho)| = O_p(1)$ and thus $I_T = O_p(1/\sqrt{T})$. Hence, the conclusion follows.

SM.3. FRECHET DERIVATIVE OF $\mathcal{A}$ AND CHARACTERIZATION OF ITS ADJOINTS

We start with the Frechet derivative A of operator $\mathcal{A}$ before characterizing the adjoints $\tilde{A}$ and A^*.

LEMMA A.16: *Under Assumption* A.4(ii) *and* (iii), *the Frechet derivative of* $\mathcal{A}$ *at* φ_0 *is the linear operator* $A := D\mathcal{A}(\varphi_0)$ *defined by* $A\varphi(z) = \int f_{X,Y|Z}(x, \varphi_0(x)|z)\varphi(x)\,dx$, $z \in \mathcal{Z}$, *for* $\varphi \in L^2[0,1]$. *Moreover, we have* $\mathcal{A}(\varphi) = \mathcal{A}(\varphi_0) + A\Delta\varphi + R(\varphi, \varphi_0)$, *where the residual* $R(\varphi, \varphi_0)$ *is such that* $\|R(\varphi, \varphi_0)\|_{L^2(F_Z,\tau)} \leq \frac{1}{2\sqrt{\tau(1-\tau)}} c\|\Delta\varphi\|^2$, $\Delta\varphi := \varphi - \varphi_0$, *and* $c := \sup_{x\in\mathcal{X},y\in\mathcal{Y},z\in\mathcal{Z}} |\nabla_y f_{X,Y|Z}(x, y|z)|$.

The characterization of the adjoint $\tilde{A}$ of operator A w.r.t. the L^2 scalar product $\langle\cdot,\cdot\rangle$ is obtained from

$$\begin{aligned}
&\langle\psi, A\varphi\rangle_{L^2(F_Z,\tau)} \\
&\quad= \frac{1}{\tau(1-\tau)} \int \psi(z)\left(\int f_{X,Y|Z}(x, \varphi_0(x)|z)\varphi(x)\,dx\right) f(z)\,dz \\
&\quad= \frac{1}{\tau(1-\tau)} \int \left(\int f_{X,Y,Z}(x, \varphi_0(x), z)\psi(z)dz\right)\varphi(x)\,dx = \langle\tilde{A}\psi, \varphi\rangle
\end{aligned}$$

for $\varphi \in L^2[0,1]$ and $\psi \in L^2(F_Z,\tau)$, which yields $\tilde{A}\psi(x) = \frac{1}{\tau(1-\tau)} \int f_{X,Y,Z}(x, \varphi_0(x), z)\psi(z)\,dz$.

The characterization of the adjoint A^* of operator A w.r.t. the Sobolev scalar product $\langle\cdot,\cdot\rangle_H$ involves the solution of a Dirichlet problem of finite, or infinite, order (see Dubinskij (1986, Chapter 2), for elliptic boundary value problems of infinite order). Let us introduce the function $p(\lambda) = \sum_{s=0}^{l} a_s\lambda^s$. When $l < \infty$, $a_s = 1$ for $s \leq l$, and $a_s = 0$ for $s > l$, the function $p(\lambda)$ is a polynomial of order l. When $l = \infty$ and $a_s = 1/s!$, we get $p(\lambda) = \exp(\lambda)$. Moreover, let us introduce the complete orthonormal system of $L^2[0,1]$ given by $\tilde{\psi}_1(x) = 1$, $\tilde{\psi}_j(x) = \sqrt{2}\cos(\pi(j-1)x)$, $j > 1$. Let us define

$$\mathcal{S}^l[0,1] = \left\{\varphi \in L^2[0,1] : \sum_{j=1}^{\infty}\left[p\left((\pi(j-1))^2\right)\langle\varphi, \tilde{\psi}_j\rangle\right]^2 < \infty\right\},$$

to $l \in \mathbb{N} \cup \{\infty\}$.

Space $\mathcal{S}^l[0,1]$ is a linear vector subspace of $L^2[0,1]$ made of functions whose basis coefficients $\langle\varphi, \tilde{\psi}_j\rangle$ feature rapid decay such that $p((\pi(j-1))^2)\langle\varphi, \tilde{\psi}_j\rangle$,

$j = 1, \dots,$ are square-summable. It is a Hilbert space w.r.t. the scalar product

$$\langle \varphi, \phi \rangle_{\mathcal{S}} := \sum_{j=1}^{\infty} \big[p\big((\pi(j-1))^2 \big) \big]^2 \langle \varphi, \tilde{\psi}_j \rangle \langle \phi, \tilde{\psi}_j \rangle.$$

We denote by $\|\varphi\|_{\mathcal{S}} := \langle \varphi, \varphi \rangle_{\mathcal{S}}^{1/2}$ the associated norm. When $l < \infty$, the space $\mathcal{S}^l[0,1]$ is equivalent to $H_0^{2l}[0,1] := \{\varphi \in H^{2l}[0,1] : \nabla^s \varphi(0) = \nabla^s \varphi(1) = 0, s = 1, 3, \dots, 2l-1\}$, that is, the Sobolev space of order $2l$ with boundary conditions for the odd-order derivatives (see Kress (1999, Chapter 8), for similar results with periodic functions). When $l = \infty$, the functions in $\mathcal{S}^\infty[0,1]$ are C^∞-functions with odd-order derivatives vanishing at the boundary and exponentially decaying basis coefficients.

LEMMA A.17: (i) *The differential operator* $\mathcal{D} := p(-\nabla^2)$ *is well defined on* $\mathcal{S}^l[0,1]$. (ii) *For any* $f \in L^2[0,1]$, *the* (*infinite-order*) *ordinary differential equation* (*ODE*)

$$\text{(SM.8)} \quad \mathcal{D}u = f, \quad u \in \mathcal{S}^l[0,1],$$

admits the unique solution $u = \sum_{j=1}^{\infty} \frac{1}{p((\pi(j-1))^2)} \langle f, \tilde{\psi}_j \rangle \tilde{\psi}_j$. (iii) *The operator* $\mathcal{D}$ *from* $\mathcal{S}^l[0,1]$ *to* $L^2[0,1]$ *is invertible and the inverse* $\mathcal{D}^{-1} : L^2[0,1] \to \mathcal{S}^l[0,1]$ *is continuous*.

For any $a \in [0,1]$ and $s = 1, \dots, l-1$, the linear functional $\varphi \mapsto \nabla^s \varphi(a)$ is continuous on $H^l[0,1]$. Denote by $\delta_a^{(s)} \in H^l[0,1]$ its Riesz representant, that is, $\langle \delta_a^{(s)}, \varphi \rangle_H = \nabla^s \varphi(a)$ for any $\varphi \in H^l[0,1]$. The next lemma gives the characterization of the adjoint A^* for $l \geq 1$.

LEMMA A.18: (i) *The adjoint of operator* A *w.r.t. the Sobolev scalar product* $\langle \cdot, \cdot \rangle_H$ *is* $A^* = \mathcal{E}\mathcal{D}^{-1}\tilde{A}$, *where the operator* $\mathcal{E} : \mathcal{S}^l[0,1] \to H^l[0,1]$ *is defined by* $\mathcal{E}u = u$ *if* $l = 1$; *by* $\mathcal{E}u = u - \sum_{j=1}^{\lfloor l/2 \rfloor} (-1)^j (\tilde{u}_j(1)\delta_1^{(2j-1)} - \tilde{u}_j(0)\delta_0^{(2j-1)})$, *where* $\tilde{u}_j = \sum_{i=j}^{l-j} (-\nabla^2)^i u$, *if* $2 \leq l < \infty$; *by* $\mathcal{E}u = u - \sum_{j=1}^{\infty} (-1)^j \frac{1}{j!} (\tilde{u}_j(1)\delta_1^{(2j-1)} - \tilde{u}_j(0)\delta_0^{(2j-1)})$, *where* $\tilde{u}_j = \sum_{i=j}^{\infty} \frac{j!}{(i+j)!} (-\nabla^2)^i u$, *if* $l = \infty$, *for* $u \in \mathcal{S}^l[0,1]$. (ii) *The operator* $\mathcal{E} : \mathcal{S}^l[0,1] \to H^l[0,1]$ *is continuous*.

SM.3.1. *Proof of Lemma A.16*

We have for $z \in \mathcal{Z}$,

$$\begin{aligned} R(\varphi, \varphi_0)(z) &= \int_{\mathcal{X}} f_{X|Z}(x|z) \big[F_{Y|X,Z}(\varphi(x)|x,z) - F_{Y|X,Z}(\varphi_0(x)|x,z) \\ &\quad - f_{Y|X,Z}(\varphi_0(x)|x,z) \Delta\varphi(x) \big] \, dx \\ &= \frac{1}{2} \int_{\mathcal{X}} f_{X|Z}(x|z) \nabla_y f_{Y|X,Z}(y(x,z)|x,z) \Delta\varphi(x)^2 \, dx, \end{aligned}$$

where $|y(x,z)| \leq |\Delta\varphi(x)|$. Thus, from Assumption A.4(iii),

$$|R(\varphi,\varphi_0)(z)| \leq \frac{1}{2}\int_{\mathcal{X}} \left|\nabla_y f_{X,Y|Z}(x,y(x,z)|z)\right| \Delta\varphi(x)^2\,dx$$
$$\leq \frac{1}{2}\left[\sup_{x,y}\left|\nabla_y f_{X,Y|Z}(x,y|z)\right|\right]\|\Delta\varphi\|^2.$$

Hence, $\|R(\varphi,\varphi_0)\|_{L^2(F_Z,\tau)} \leq \frac{1}{2\sqrt{\tau(1-\tau)}}c\|\Delta\varphi\|^2$. In particular, we get $\|\mathcal{A}(\varphi) - \mathcal{A}(\varphi_0) - A\Delta\varphi\|_{L^2(F_Z,\tau)}/\|\Delta\varphi\| \to 0$ as $\|\Delta\varphi\| \to 0$. Since operator A is bounded (Assumption A.4 (ii)), this implies that A is the Frechet derivative of $\mathcal{A}$ at φ_0.

SM.3.2. *Proof of Lemma A.17*

(i) The gradient operator ∇ is the weak (or distributional) derivative (see Adams and Fournier (2003, p. 22)). Then, when $l < \infty$, operator $p(-\nabla^2)$ is well defined for functions in $H^{2l}[0,1]$, and hence on the subspace $\mathcal{S}^l[0,1]$. Moreover, for $\varphi \in \mathcal{S}^l[0,1]$, we have

$$p(-\nabla^2)\varphi = \sum_{j=1}^{\infty} p\big((\pi(j-1))^2\big)\langle\varphi,\tilde{\psi}_j\rangle\tilde{\psi}_j \in L^2[0,1]. \tag{SM.9}$$

Indeed, for any $\phi \in C_0^\infty[0,1]$, we have

$$\int [p(-\nabla^2)\phi](x)\varphi(x)\,dx = \sum_{j=1}^{\infty}\langle\tilde{\psi}_j,\varphi\rangle \int [p(-\nabla^2)\phi](x)\tilde{\psi}_j(x)\,dx$$
$$= \sum_{j=1}^{\infty}\langle\tilde{\psi}_j,\varphi\rangle \int \phi(x)p(-\nabla^2)\tilde{\psi}_j(x)\,dx$$
$$= \sum_{j=1}^{\infty} p\big((\pi(j-1))^2\big)\langle\varphi,\tilde{\psi}_j\rangle\langle\phi,\tilde{\psi}_j\rangle,$$

where the first equality uses the equation $\varphi = \sum_{j=1}^{\infty}\langle\varphi,\tilde{\psi}_j\rangle\tilde{\psi}_j$, the second equality comes from partial integration with vanishing boundary terms, and the third equality exploits $p(-\nabla^2)\tilde{\psi}_j = p((\pi(j-1))^2)\tilde{\psi}_j$.

Let us now consider the case $l = \infty$. The operator $p(-\nabla^2) = \exp(-\nabla^2)$ is defined as an L^2-limit, that is, $p(-\nabla^2)\varphi := \lim_{n\to\infty}\sum_{s=0}^{n}\frac{1}{s!}(-\nabla^2)^s\varphi$ for $\varphi \in \mathcal{S}^\infty[0,1]$, where the limit is in the L^2-norm. Now, from the above arguments, we have $\sum_{s=0}^{n}\frac{1}{s!}(-\nabla^2)^s\varphi = \sum_{j=1}^{\infty}(\sum_{s=0}^{n}\frac{1}{s!}(\pi(j-1))^{2s})\langle\varphi,\tilde{\psi}_j\rangle\tilde{\psi}_j$ for any $n \in \mathbb{N}$. By using the Lebesgue theorem, it is seen that the L^2-limit of the RHS is $\sum_{j=1}^{\infty}\exp((\pi(j-1))^2)\langle\varphi,\tilde{\psi}_j\rangle\tilde{\psi}_j$. Thus, equation (SM.9) holds also in the case $l = \infty$.

(ii) By using (SM.9), it is seen that function u solves the ODE

(SM.10) $p(-\nabla^2)u = f.$

Let us now prove uniqueness. Suppose that $u, v \in \mathcal{S}^l[0,1]$ solve the ODE (SM.10). Thus, $\varphi := u - v \in \mathcal{S}^l[0,1]$ solves $p(-\nabla^2)\varphi = 0$. We now use the next lemma.

LEMMA B.2: *For* $\varphi, \phi \in \mathcal{S}^l[0,1]$, *we have* $\langle \phi, \varphi\rangle_H = \langle \phi, p(-\nabla^2)\varphi\rangle = \sum_{j=1}^{\infty} p((\pi(j-1))^2)\langle \phi, \tilde{\psi}_j\rangle\langle \varphi, \tilde{\psi}_j\rangle$.

From Lemma B.2, we have $0 = \langle \varphi, p(-\nabla^2)\varphi\rangle = \|\varphi\|_H^2$, which implies $\varphi = 0$, that is, $u = v$.

(iii) Let $f \in L^2[0,1]$ and $u = \mathcal{D}^{-1}f = \sum_{j=1}^{\infty} \frac{1}{p((\pi(j-1))^2)}\langle \tilde{\psi}_j, f\rangle \tilde{\psi}_j$. We have $\langle \tilde{\psi}_j, u\rangle = \frac{1}{p((\pi(j-1))^2)}\langle \tilde{\psi}_j, f\rangle$ and $\|u\|_{\mathcal{S}}^2 = \sum_{j=1}^{\infty}[p((\pi(j-1))^2)]^2\langle \tilde{\psi}_j, u\rangle^2 = \sum_{j=1}^{\infty}\langle \tilde{\psi}_j, f\rangle^2 = \|f\|^2$. The conclusion follows.

SM.3.3. *Proof of Lemma A.18*

(i) Let $\psi \in L^2(F_Z, \tau)$ be given. Define $f := \tilde{A}\psi \in L^2[0,1]$ and $u := \mathcal{D}^{-1}f \in \mathcal{S}^l[0,1]$. Let us first consider the case $l < \infty$. From partial integration (see also (SM.32) in Section SM.7.2), for any $\phi \in H^l[0,1]$, we have $\langle \nabla^s\phi, \nabla^s u\rangle = \langle \phi, (-\nabla^2)^s u\rangle + \sum_{i=0}^{s-1}(-1)^i \nabla^{s-1-i}\phi\nabla^{s+i}u|_0^1$, where the terms with $s+i$ odd vanish in the sum. Thus, we get

$$\begin{aligned}
\langle \phi, u\rangle_H &= \langle \phi, p(-\nabla^2)u\rangle + \sum_{s=0}^{l} a_s \sum_{i=0}^{s-1}(-1)^i \nabla^{s-1-i}\phi\nabla^{s+i}u|_0^1 \\
&= \langle \phi, p(-\nabla^2)u\rangle + \sum_{\substack{q\le l-1 \\ q \text{ odd}}} \sum_{i=0}^{l-1-q}(-1)^i a_{q+i+1}\nabla^q\phi\nabla^{q+2i+1}u|_0^1 \\
&= \langle \phi, p(-\nabla^2)u\rangle + \sum_{j=1}^{\lfloor l/2\rfloor}\sum_{i=0}^{l-2j}(-1)^i a_{2j+i}\nabla^{2j-1}\phi\nabla^{2j+2i}u|_0^1 \\
&= \langle \phi, p(-\nabla^2)u\rangle + \sum_{j=1}^{\lfloor l/2\rfloor}\sum_{i=j}^{l-j}(-1)^j a_{j+i}\nabla^{2j-1}\phi(-\nabla^2)^i u|_0^1.
\end{aligned}$$

Now, by using $\langle \phi, p(-\nabla^2)u\rangle = \langle \phi, f\rangle = \langle \phi, \tilde{A}\psi\rangle = \langle \phi, A^*\psi\rangle_H$, we get

$$\langle \phi, A^*\psi\rangle_H = \langle \phi, u\rangle_H - \sum_{j=1}^{\lfloor l/2\rfloor}\sum_{i=j}^{l-j}(-1)^j a_{j+i}\nabla^{2j-1}\phi(-\nabla^2)^i u|_0^1$$

for any $\phi \in H^l[0,1]$. Then, by using $\nabla^{2j-1}\phi(a) = \langle \phi, \delta_a^{(2j-1)}\rangle_H$ for $a = 0, 1$, we get

$$A^*\psi = u - \sum_{j=1}^{\lfloor l/2 \rfloor} \sum_{i=j}^{l-j} (-1)^j a_{j+i} \big([(-\nabla^2)^i u](1) \delta_1^{(2j-1)} - [(-\nabla^2)^i u](0) \delta_0^{(2j-1)} \big).$$

When $l = 1$, no boundary term appears. Then $A^*\psi = u$ and $A^* = \mathcal{D}^{-1}\tilde{A}$. When $2 \le l < \infty$, we set $a_s = 1$ for $s \le l$ and $a_s = 0$ for $s > l$. In the infinite-order case, we set $a_s = 1/s!$ and let $l \to \infty$. The conclusion follows.

(ii) We have to prove that $\|\mathcal{E}u\|_H \le C\|u\|_{\mathcal{S}}$ for a constant $C < \infty$. We consider the case $l = \infty$ (the case $l < \infty$ is similar). We have, for any $u \in \mathcal{S}^l[0,1]$,

$$\|\mathcal{E}u\|_H \le \|u\|_H + \sum_{j=1}^{\infty} \frac{1}{j!} \big(|\tilde{u}_j(1)| \big\| \delta_1^{(2j-1)} \big\|_H + |\tilde{u}_j(0)| \big\| \delta_0^{(2j-1)} \big\|_H \big). \tag{SM.11}$$

Moreover, $\|u\|_H \le \|u\|_{\mathcal{S}}$ from Lemma B.2. Let us now bound $|\tilde{u}_j(a)|$ and $\|\delta_a^{(2j-1)}\|_H$, $a \in \{0,1\}$, $j \in \mathbb{N}$. We use the next lemma.

LEMMA C.1: *For any* $\varphi \in H^1[0,1]$, *we have* $\sup_{x\in[0,1]} |\varphi(x)| \le 2\|\varphi\|_{H^1}$.

From Lemma C.1 and by using that $\|\varphi\|_H^2 \ge \frac{1}{(2j)!}\|\nabla^{2j-1}\varphi\|_{H^1}^2$ for any $\varphi \in H^\infty[0,1]$ and $j \in \mathbb{N}$, we get

$$\begin{aligned} \big\| \delta_a^{(2j-1)} \big\|_H &= \sup_{\substack{\varphi \in H^\infty[0,1]: \\ \|\varphi\|_H = 1}} \big\langle \delta_a^{(2j-1)}, \varphi \big\rangle \\ &= \sup_{\substack{\varphi \in H^\infty[0,1]: \\ \|\varphi\|_H = 1}} \nabla^{2j-1}\varphi(a) \le 2\sqrt{(2j)!}. \end{aligned} \tag{SM.12}$$

Moreover, $|\tilde{u}_j(a)| \le \sum_{i=j}^{\infty} \frac{j!}{(i+j)!} |(-\nabla^2)^i u(a)| \le \sum_{i=j}^{\infty} \frac{j!}{(i+j)!} \|(-\nabla^2)^i u\|_{H^1}$. Now, by using $\frac{j!}{(i+j)!} \le \frac{1}{(i+1)!}$ and the Cauchy–Schwarz inequality, we get, for any a and j,

$$\begin{aligned} |\tilde{u}_j(a)| &\le \sum_{i=0}^{\infty} \frac{1}{(i+1)!} \|(-\nabla^2)^i u\|_{H^1} \\ &\le \left(\sum_{i=0}^{\infty} \frac{2^{2i+1}}{(2i+1)!} \|(-\nabla^2)^i u\|_{H^1}^2 \right)^{1/2} \left(\sum_{i=0}^{\infty} \frac{(2i+1)!}{2^{2i+1}[(i+1)!]^2} \right)^{1/2}. \end{aligned}$$

Let us prove that the two series in the RHS are convergent. From Lemma B.2, we have

$$\begin{aligned}
\|(-\nabla^2)^i u\|_{H^1}^2 &= \langle u, (-\nabla^2)^{2i} u\rangle + \langle u, (-\nabla^2)^{2i+1} u\rangle \\
&= \sum_{j=1}^{\infty} \big[(\pi(j-1))^{4i} + (\pi(j-1))^{4i+2}\big] \langle \tilde{\psi}_j, u\rangle^2 \\
&\leq 2 \sum_{j=1}^{\infty} \big((\pi(j-1))^2\big)^{2i+1} \langle \tilde{\psi}_j, u\rangle^2.
\end{aligned}$$

Thus, we get

$$\begin{aligned}
&\sum_{i=0}^{\infty} \frac{2^{2i+1}}{(2i+1)!} \|(-\nabla^2)^i u\|_{H^1}^2 \\
&\quad \leq 2 \sum_{j=1}^{\infty} \left(\sum_{i=0}^{\infty} \frac{(2(\pi(j-1))^2)^{2i+1}}{(2i+1)!} \right) \langle \tilde{\psi}_j, u\rangle^2 \\
&\quad \leq \sum_{j=1}^{\infty} \big[\exp\big(2(\pi(j-1))^2\big) - \exp\big(-2(\pi(j-1))^2\big)\big] \langle \tilde{\psi}_j, u\rangle^2 \\
&\quad \leq \sum_{j=1}^{\infty} \big[\exp\big((\pi(j-1))^2\big)\big]^2 \langle \tilde{\psi}_j, u\rangle^2 = \|u\|_{\mathcal{S}}^2,
\end{aligned}$$

where we have used that $\sum_{i=0}^{\infty} \frac{z^{2i+1}}{(2i+1)!} = \frac{e^z - e^{-z}}{2}$. Let us now consider the series $\sum_{i=0}^{\infty} \frac{(2i+1)!}{2^{2i+1}[(i+1)!]^2}$. Note that $(2i+1)! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot \cdots \cdot (2i-1)(2i)(2i+1)$ and $2^{2i+1}[(i+1)!]^2 = 2(i+1)^2(2^i i!)^2 = 2(i+1)^2[2^2 \cdot 4^2 \cdot \cdots \cdot (2i)^2]$. Thus,

$$\begin{aligned}
\frac{(2i+1)!}{2^{2i+1}[(i+1)!]^2} &= \frac{1}{2(i+1)^2} \frac{1}{2} \cdot \frac{3}{4} \cdot \cdots \cdot \frac{2i-1}{2i} (2i+1) \\
&= \left(\frac{2i+3}{i+1} \right) \frac{1}{2} \cdot \frac{3}{4} \cdot \cdots \cdot \frac{2i-1}{2i} \cdot \frac{2i+1}{2(i+1)} \frac{1}{2i+3}.
\end{aligned}$$

Since $\frac{2i+3}{i+1} \leq 3$, the series $\sum_{i=0}^{\infty} \frac{(2i+1)!}{2^{2i+1}[(i+1)!]^2}$ converges if the series $\sum_{i=0}^{\infty} \frac{1}{2} \cdot \frac{3}{4} \cdot \cdots \cdot \frac{2i-1}{2i} \cdot \frac{2i+1}{2(i+1)} \frac{1}{2i+3} = \frac{1}{2}\frac{1}{3} + \frac{1}{2}\frac{3}{4}\frac{1}{5} + \frac{1}{2}\frac{3}{4}\frac{5}{6}\frac{1}{7} + \cdots$ converges. The Taylor series of the arcsin function is $\arcsin(z) = z + \frac{1}{2}\frac{z^3}{3} + \frac{1}{2}\frac{3}{4}\frac{z^5}{5} + \frac{1}{2}\frac{3}{4}\frac{5}{6}\frac{z^7}{7} + \cdots$. Thus, we get

$\sum_{i=0}^{\infty} \frac{1}{2} \cdot \frac{3}{4} \cdot \dots \cdot \frac{2i-1}{2i} \cdot \frac{2i+1}{2(i+1)} \frac{1}{2i+3} = \arcsin(1) - 1 = \pi/2 - 1$. It follows that the series $\sum_{i=0}^{\infty} \frac{(2i+1)!}{2^{2i+1}[(i+1)!]^2}$ converges. We get

$$\text{(SM.13)} \quad |\tilde{u}_j(a)| \le \left(\sum_{i=0}^{\infty} \frac{(2i+1)!}{2^{2i+1}[(i+1)!]^2} \right)^{1/2} \|u\|_{\mathcal{S}}$$

for any $a \in \{0, 1\}$ and $j \in \mathbb{N}$. Finally, from (SM.11), (SM.12), and (SM.13), we get $\|\mathcal{E}u\|_H \le C\|u\|_{\mathcal{S}}$, where $C = 1 + 4(\sum_{j=1}^{\infty} \frac{\sqrt{(2j)!}}{j!})(\sum_{i=0}^{\infty} \frac{(2i+1)!}{2^{2i+1}[(i+1)!]^2})^{1/2} < \infty$.

SM.4. ASYMPTOTIC DISTRIBUTION

SM.4.1. *Proof of Lemma A.4*

The statement follows from arguments similar to those developed in the proof of Lemma A.16 in Section SM.3.1.

SM.4.2. *Proof of Lemma A.5*

Operator $\hat{A}^*$ is the adjoint of $\hat{A}$ w.r.t. scalar products $\langle \cdot, \cdot \rangle_H$ and $\langle \cdot, \cdot \rangle_{L^2(\hat{F}_Z, \tau)}$, that is, $\langle \varphi, \hat{A}\psi \rangle_{L^2(\hat{F}_Z, \tau)} = \langle \hat{A}^*\varphi, \psi \rangle_H$. We have

$$\begin{aligned} \|\hat{\mathcal{K}}_T(\Delta\hat{\varphi})\| &\le \|\hat{\mathcal{K}}_T(\Delta\hat{\varphi})\|_H \\ &\le \|(\lambda_T + \hat{A}_0^*\hat{A}_0)^{-1}\hat{A}_0^*\|_{\mathcal{L}(L^2(\hat{F}_Z, \tau), H^l[0,1])} \|\hat{R}(\hat{\varphi}, \varphi_0)\|_{L^2(\hat{F}_Z, \tau)}. \end{aligned}$$

We will show below that P-a.s.,

$$\text{(i)} \quad \|(\lambda_T + \hat{A}_0^*\hat{A}_0)^{-1}\hat{A}_0^*\|_{\mathcal{L}(L^2(\hat{F}_Z, \tau), H^l[0,1])} \le 1/\sqrt{\lambda_T}$$

and

$$\begin{aligned} \text{(ii)} \quad & \|\hat{R}(\hat{\varphi}, \varphi_0)\|_{L^2(\hat{F}_Z, \tau)} \\ & \le \frac{1}{2\sqrt{\tau(1-\tau)}} \sup_{x \in \mathcal{X}, y \in \mathbb{R}, z \in \mathcal{Z}} \left| \nabla_y \hat{f}_{X,Y|Z}(x, y|z) \right| \|\Delta\hat{\varphi}\|^2, \end{aligned}$$

where $\|(\lambda_T + \hat{A}_0^*\hat{A}_0)^{-1}\hat{A}_0^*\|_{\mathcal{L}(L^2(\hat{F}_Z, \tau), H^l[0,1])} := \sup_{\psi \in L^2(\hat{F}_Z, \tau): \|\psi\|_{L^2(\hat{F}_Z, \tau)} = 1} \|(\lambda_T + \hat{A}_0^*\hat{A}_0)^{-1}\hat{A}_0^*\psi\|_H$ denotes the operator norm of $(\lambda_T + \hat{A}_0^*\hat{A}_0)^{-1}\hat{A}_0^*$.[1] Then it fol-

[1]For a linear operator $B: H_1 \to H_2$ between Hilbert spaces H_1 and H_2, we denote by $\|B\|_{\mathcal{L}(H_1, H_2)}$ the operator norm $\|B\|_{\mathcal{L}(H_1, H_2)} := \sup_{\varphi \in H_1: \|\varphi\|_{H_1} = 1} \|B\varphi\|_{H_2}$.

lows that

$$\|\hat{\mathcal{K}}_T(\Delta\hat{\varphi})\| \leq \frac{1}{\sqrt{\lambda_T}} \frac{1}{2\sqrt{\tau(1-\tau)}} \times \sup_{x\in\mathcal{X}, y\in\mathbb{R}, z\in\mathcal{Z}} \left|\nabla_y \hat{f}_{X,Y|Z}(x,y|z)\right| \|\Delta\hat{\varphi}\|^2, \quad P\text{-a.s.}$$

We deduce that for $\varepsilon := C - \frac{1}{2\sqrt{\tau(1-\tau)}} \sup_{x\in\mathcal{X}, y\in\mathbb{R}, z\in\mathcal{Z}} |\nabla_y f_{X,Y|Z}(x,y|z)| > 0$, we have

$$\begin{aligned}
&P\left[\|\hat{\mathcal{K}}_T(\Delta\hat{\varphi})\| > \frac{C}{\sqrt{\lambda_T}}\|\Delta\hat{\varphi}\|^2\right] \\
&\quad\leq P\left[\frac{1}{2\sqrt{\tau(1-\tau)}} \sup_{x\in\mathcal{X}, y\in\mathbb{R}, z\in\mathcal{Z}} \left|\nabla_y \hat{f}_{X,Y|Z}(x,y|z)\right| > C\right] \\
&\quad\leq P\Big[\sup_{x\in\mathcal{X}, y\in\mathbb{R}, z\in\mathcal{Z}} \left|\nabla_y \hat{f}_{X,Y|Z}(x,y|z)\right| \\
&\qquad > \sup_{x\in\mathcal{X}, y\in\mathbb{R}, z\in\mathcal{Z}} \left|\nabla_y f_{X,Y|Z}(x,y|z)\right| + 2\varepsilon\sqrt{\tau(1-\tau)}\Big] \\
&\quad\leq P\Big[\sup_{x\in\mathcal{X}, y\in\mathbb{R}, z\in\mathcal{Z}} \left|\nabla_y \hat{f}_{X,Y|Z}(x,y|z) - \nabla_y f_{X,Y|Z}(x,y|z)\right| \\
&\qquad \geq 2\varepsilon\sqrt{\tau(1-\tau)}\Big].
\end{aligned}$$

The latter probability converges to zero exponentially fast by an application of Bernstein's inequality similar to the proof of Lemma B.1, using that $\frac{\log T}{T h_T^{4+d_Z}} + h_T^{2m} = o(1)$, and that kernel K and variables X, Y have compact support.

Finally, let us prove statements (i) and (ii) above. To prove (i), define operator $\hat{U} := (\lambda_T + \hat{A}_0^* \hat{A}_0)^{-1} \hat{A}_0^*$. The adjoint of $\hat{U}$ w.r.t. scalar products $\langle \cdot, \cdot \rangle_{L^2(\hat{F}_Z,\tau)}$ and $\langle \cdot, \cdot \rangle_H$ is $\hat{U}^* = \hat{A}_0(\lambda_T + \hat{A}_0^* \hat{A}_0)^{-1}$. Further, we have $\|\hat{U}\|_{\mathcal{L}(L^2(\hat{F}_Z,\tau), H^l[0,1])} = \|\hat{U}^*\|_{\mathcal{L}(H^l[0,1], L^2(\hat{F}_Z,\tau))}$ from Kress (1999, Theorem 4.9). For any $\varphi \in H^2[0,1]$, we have

$$\begin{aligned}
&\|\hat{U}^*\varphi\|^2_{L^2(\hat{F}_Z,\tau)} \\
&\quad= \langle \hat{A}_0(\lambda_T + \hat{A}_0^*\hat{A}_0)^{-1}\varphi, \hat{A}_0(\lambda_T + \hat{A}_0^*\hat{A}_0)^{-1}\varphi \rangle_{L^2(\hat{F}_Z,\tau)} \\
&\quad= \langle (\lambda_T + \hat{A}_0^*\hat{A}_0)^{-1}\varphi, \hat{A}_0^*\hat{A}_0(\lambda_T + \hat{A}_0^*\hat{A}_0)^{-1}\varphi \rangle_H \\
&\quad= \left\|(\hat{A}_0^*\hat{A}_0)^{1/2}(\lambda_T + \hat{A}_0^*\hat{A}_0)^{-1}\varphi\right\|^2_H.
\end{aligned}$$

Moreover,

$$\begin{aligned}\big\|(\hat{A}_0^*\hat{A}_0)^{1/2}(\lambda_T+\hat{A}_0^*\hat{A}_0)^{-1}\varphi\big\|_H &= \big\|(\lambda_T+\hat{A}_0^*\hat{A}_0)^{-1}(\hat{A}_0^*\hat{A}_0)^{1/2}\varphi\big\|_H \\ &\le \big\|(\lambda_T+\hat{A}_0^*\hat{A}_0)^{-1/2}\big\|_{\mathcal{L}(H^l[0,1],H^l[0,1])} \\ &\quad\times \big\|(\lambda_T+\hat{A}_0^*\hat{A}_0)^{-1/2}(\hat{A}_0^*\hat{A}_0)^{1/2}\varphi\big\|_H.\end{aligned}$$

Using $\|(\lambda_T+\hat{A}_0^*\hat{A}_0)^{-1/2}\|_{\mathcal{L}(H^l[0,1],H^l[0,1])}\le\lambda_T^{-1/2}$ and $\|(\lambda_T+\hat{A}_0^*\hat{A}_0)^{-1/2}(\hat{A}_0^*\times\hat{A}_0)^{1/2}\varphi\|_H\le\|\varphi\|_H$, P-a.s., we get $\|\hat{U}^*\varphi\|_{L^2(\hat{F}_{Z,\tau})}\le\lambda_T^{-1/2}\|\varphi\|_H$, P-a.s.

Statement (ii) follows from Lemma A.4 with $\bar{\varphi}=\varphi_0$.

SM.4.3. *Proof of Lemma A.6*

SM.4.3.1. *Control of the Nonlinearity Term*

First we consider the nonstochastic analog of Equation (A.4).

LEMMA B.3: *Let function φ satisfy $\varphi=\psi+\varepsilon\mathcal{K}(\varphi)$, where ψ is a known function, $\mathcal{K}$ is a nonlinear operator such that $\|\mathcal{K}(\varphi)\|\le\|\varphi\|^2$, and $\varepsilon>0$. If $\varepsilon\|\psi\|<1/8$, then either $|\|\varphi\|^2-\|\psi\|^2|\le 32\varepsilon\|\psi\|^3$ or $\|\varphi\|^2\ge\frac{3}{8\varepsilon^2}$.*

We can use Lemma B.3 with $\varepsilon=\varepsilon_T=\frac{C}{\sqrt{\lambda_T}}$ to bound the difference $\|\Delta\hat{\varphi}\|^2-\|\Delta\hat{\psi}\|^2$ on the set $\{\varepsilon_T\|\Delta\hat{\psi}\|<1/8\wedge\|\Delta\hat{\varphi}\|^2<\frac{3}{8\varepsilon_T^2}\wedge\|\hat{\mathcal{K}}_T(\Delta\hat{\varphi})\|\le\varepsilon_T\|\Delta\hat{\varphi}\|^2\}$ and derive the following result.

LEMMA B.4: *Under Assumptions* A.1–A.3, *and* A.4(ii) *and* (iii), *and* $\eta<\frac{1}{4+d_Z}$, *we have for any $\bar{b}>0$, with* $C>\frac{1}{2\sqrt{\tau(1-\tau)}}\sup_{x\in\mathcal{X},y\in\mathcal{Y},z\in\mathcal{Z}}|\nabla_y f_{X,Y|Z}(x,y|z)|$,

$$\begin{aligned}&E[\|\Delta\hat{\varphi}\|^2]-E[\|\Delta\hat{\psi}\|^2]\\&\quad=O\bigg(\frac{1}{\sqrt{\lambda_T}}E[\|\Delta\hat{\psi}\|^3]+P\bigg[\|\Delta\hat{\psi}\|^2\ge\frac{\lambda_T}{64C^2}\bigg]+P\bigg[\|\Delta\hat{\varphi}\|^2\ge\frac{3\lambda_T}{8C^2}\bigg]\bigg)\\&\qquad+O(T^{-\bar{b}}).\end{aligned}\tag{SM.14}$$

Note that for large T, probability $P[\|\Delta\hat{\varphi}\|^2>\frac{3\lambda_T}{8C^2}]$ on the RHS of (SM.14) controls for both the event $\|\Delta\hat{\varphi}\|^2\ge\frac{3}{8\varepsilon_T^2}$ and the event $\|\Delta\hat{\varphi}\|\ge r$, in which the first-order condition (4.1) does not hold.

SM.4.3.2. *A Large Deviation Bound for Penalized Minimum Distance Estimators*

LEMMA B.5: *We have*

$$P[\|\hat{\varphi}-\varphi_0\|\ge\varepsilon_T]\le k_1(T,C(\varepsilon_T,\lambda_T))+k_2(T,C(\varepsilon_T,\lambda_T))$$

if $\varepsilon_T, \lambda_T > 0$ *are such that* $C(\varepsilon_T, \lambda_T) > 0$, *where*

$$C(\varepsilon, \lambda) := \inf_{\varphi \in \Theta: \|\varphi - \varphi_0\| \geq \varepsilon} Q_\infty(\varphi) + \lambda \|\varphi\|_H^2 - \lambda \|\varphi_0\|_H^2 \tag{SM.15}$$

and

$$k_1(T, \eta) := P\left[\sup_{\varphi \in \Theta} \sup_{z \in \mathcal{Z}} \frac{|\Delta \hat{m}(\varphi, z)|}{\sqrt{\tau(1-\tau)}} \geq \frac{\sqrt{\lambda_T \|\varphi_0\|_H^2 + 2\eta} - \sqrt{\lambda_T \|\varphi_0\|_H^2}}{4}\right], \tag{SM.16}$$

$$k_2(T, \eta) := P\left[\sup_{\varphi \in \Theta} \left| \frac{1}{T\tau(1-\tau)} \sum_{t=1}^T m(\varphi, Z_t)^2 - Q_\infty(\varphi) \right| \geq \eta/2 \right]. \tag{SM.17}$$

In an ill-posed setting, the usual "identifiable uniqueness" condition (White and Wooldridge (1991)) $\inf_{\varphi \in \Theta: \|\varphi - \varphi_0\| \geq \varepsilon} Q_\infty(\varphi) > Q_\infty(\varphi_0)$ does not hold (see Gagliardini and Scaillet (2012, GS)). It is replaced by the inequality $C(\varepsilon, \lambda) > 0$ for the penalized criterion, and the behavior of $C(\varepsilon, \lambda)$ as $\lambda, \varepsilon \to 0$ matters for the rate of convergence of $\hat{\varphi}$. A lower bound for the function $C(\varepsilon, \lambda)$ as $\lambda \to 0$ and $\varepsilon = O(\sqrt{\lambda})$ is given in the next result.

LEMMA B.6: *Under Assumptions* 1–4 *and* A.4(ii) *and* (iii), *we have for any* $c < d^2$,

$$\inf_{\substack{\varphi \in \Theta: \\ \|\varphi - \varphi_0\| \geq d\sqrt{\lambda}}} Q_\infty(\varphi) + \lambda \|\varphi\|_H^2 - \lambda \|\varphi_0\|_H^2 \geq c \lambda \Gamma(\lambda)$$

as $\lambda \to 0$, *where* $\Gamma(\lambda)$ *is defined in Assumption* 4.

From Lemmas B.5 and B.6, we deduce that for a constant $c > 0$,

$$\begin{aligned} P[\|\Delta \hat{\varphi}\|^2 \geq d^2 \lambda_T] &\leq P\left[\sup_{\varphi \in \Theta} \sup_{z \in \mathcal{Z}} |\Delta \hat{m}(\varphi, z)|^2 \geq c \lambda_T \Gamma(\lambda_T)^2\right] \\ &\quad + P\left[\sup_{\varphi \in \Theta} \left| \frac{1}{T\tau(1-\tau)} \sum_{t=1}^T m(\varphi, Z_t)^2 - Q_\infty(\varphi) \right| \geq c \lambda_T \Gamma(\lambda_T)\right]. \end{aligned} \tag{SM.18}$$

The next lemma is proved by bounding terms $\sup_{\varphi \in \Theta} \sup_{z \in \mathcal{Z}} |\Delta \hat{m}(\varphi, z)|^2$ and $\sup_{\varphi \in \Theta} |\frac{1}{T\tau(1-\tau)} \sum_{t=1}^T m(\varphi, Z_t)^2 - Q_\infty(\varphi)|$ in terms of suprema of suitable empirical processes over compact finite-dimensional sets.

LEMMA B.7: *Under Assumptions* 1–4, A.1–A.3, *and* A.4(ii) *and* (iii), *and* $0 < \gamma < \min\{\frac{1-\eta(d_Z+1)}{1+2a}, \frac{2m\eta}{1+2a}, \frac{1}{2(1+a)}\}$ *for any* $d, \bar{b} > 0$, *we have*

$$P[\|\Delta\hat{\varphi}\|^2 \geq d^2\lambda_T] = O(T^{-\bar{b}}), \quad P[\|\Delta\hat{\psi}\|^2 \geq d^2\lambda_T] = O(T^{-\bar{b}}).$$

By combining Lemmas B.4 and B.7, Lemma A.6 is proved.

SM.4.4. *Proof of Lemma A.7*

From Lemma A.6, the conclusion follows if

$$\text{(i)} \quad E[\|\Delta\hat{\psi}\|^2] = O(M_T(\lambda_T))$$

and

$$\text{(ii)} \quad \frac{1}{\sqrt{\lambda_T}} E[\|\Delta\hat{\psi}\|^3] = o(M_T(\lambda_T)).$$

To show these statements, we use the decomposition (A.5) and give a series of inequalities and bounds to show that the remainder term $\mathcal{R}_T$ given in (A.6) can be neglected. First, from Cauchy–Schwarz inequality,

$$\begin{aligned} E[\|\Delta\hat{\psi}\|^2] &= E[\|\mathcal{V}_T + \mathcal{B}_T\|^2] + E[\|\mathcal{R}_T\|^2] \\ &\quad + O\big(E[\|\mathcal{V}_T + \mathcal{B}_T\|^2]^{1/2} E[\|\mathcal{R}_T\|^2]^{1/2}\big), \end{aligned} \tag{SM.19}$$

where $\mathcal{V}_T := (\lambda_T + A^*A)^{-1}A^*\hat{\zeta}$ and

$$\begin{aligned} E[\|\Delta\hat{\psi}\|^3] &\leq E[\|\Delta\hat{\psi}\|^4]^{1/2} E[\|\Delta\hat{\psi}\|^2]^{1/2} \\ &\leq C\big(E[\|\mathcal{V}_T + \mathcal{B}_T\|^4] + E[\|\mathcal{R}_T\|^4]\big)^{1/2} E[\|\Delta\hat{\psi}\|^2]^{1/2} \end{aligned} \tag{SM.20}$$

for a constant C. Second, we can isolate the estimation bias by writing

$$\begin{aligned} \mathcal{V}_T + \mathcal{B}_T &= (\lambda_T + A^*A)^{-1}A^*(\hat{\zeta} - E\hat{\zeta}) + (\lambda_T + A^*A)^{-1}A^*E\hat{\zeta} \\ &\quad + [(\lambda_T + A^*A)^{-1}A^*A - 1]\varphi_0. \end{aligned}$$

Thus,

$$\begin{aligned} E[\|\mathcal{V}_T + \mathcal{B}_T\|^2] &= E\big[\|(\lambda_T + A^*A)^{-1}A^*(\hat{\zeta} - E\hat{\zeta})\|^2\big] \\ &\quad + \big\|(\lambda_T + A^*A)^{-1}A^*E\hat{\zeta} \\ &\quad + [(\lambda_T + A^*A)^{-1}A^*A - 1]\varphi_0\big\|^2 \end{aligned} \tag{SM.21}$$

and, for a constant C,

$$\|\mathcal{V}_T+\mathcal{B}_T\|^4 \leq C\big(\|(\lambda_T+A^*A)^{-1}A^*(\hat{\zeta}-E\hat{\zeta})\|^4 + \big\|(\lambda_T+A^*A)^{-1}A^*E\hat{\zeta} + [(\lambda_T+A^*A)^{-1}A^*A-1]\varphi_0\big\|^4\big). \tag{SM.22}$$

In Lemma B.8, we give the asymptotic behavior of $E[\|(\lambda_T+A^*A)^{-1}A^*(\hat{\zeta}-E\hat{\zeta})\|^2]$ and $E[\|(\lambda_T+A^*A)^{-1}A^*(\hat{\zeta}-E\hat{\zeta})\|^4]$. In Lemma B.9, we prove that estimation bias is negligible compared to regularization bias, and in Lemma B.10, we give bounds on the remainder term. Combining Lemmas B.8(i), B.9, and B.10(i) with Equations (SM.19) and (SM.21), and using $M_T(\lambda_T)=\frac{1}{T}\sum_{j=1}^{\infty}\frac{\nu_j}{(\nu_j+\lambda_T)^2}\|\phi_j\|^2+\int\mathcal{B}_T(x)^2\,dx$ yields statement (i) above. Then combining Lemmas B.8(ii), B.9, and B.10(ii) with inequalities (SM.20) and (SM.22) yields $E[\|\Delta\hat{\psi}\|^3]=O(M_T(\lambda_T)^{3/2})$. The latter in turn implies statement (ii) by using $M_T(\lambda_T)=V_T(\lambda_T)+O(\lambda_T^{2\delta})=o(\lambda_T)$.

LEMMA B.8: *Under Assumptions* A.1, A.2, A.4(i), *and* A.5, *and if* $h_T^{1/4}\times\frac{V_T(\lambda_T;2)}{V_T(\lambda_T)}=o(1)$, *then*

(i) $$E\big[\|(\lambda_T+A^*A)^{-1}A^*(\hat{\zeta}-E\hat{\zeta})\|^2\big] = O\left(\frac{1}{T}\sum_{j=1}^{\infty}\frac{\nu_j}{(\nu_j+\lambda_T)^2}\|\phi_j\|^2\right),$$

(ii) $$E\big[\|(\lambda_T+A^*A)^{-1}A^*(\hat{\zeta}-E\hat{\zeta})\|^4\big] = O\left(\left(\frac{1}{T}\sum_{j=1}^{\infty}\frac{\nu_j}{(\nu_j+\lambda_T)^2}\|\phi_j\|^2\right)^2\right).$$

LEMMA B.9: *Under Assumptions* A.1, A.2, A.3(ii), *and* A.4(i), *and* $\gamma<\frac{\eta m}{2}$,

$$\big\|(\lambda_T+A^*A)^{-1}A^*E\hat{\zeta}+[(\lambda_T+A^*A)^{-1}A^*A-1]\varphi_0\big\|^2 = (1+o(1))\int\mathcal{B}_T(x)^2\,dx.$$

LEMMA B.10: *Under Assumption* A.1–A.5, *and* $\eta<\frac{1}{2(d_Z+2)}$ *and* $\gamma<\frac{1}{2}\min\{1-(d_Z+1)\eta, m\eta\}$, *we have* (i) $E[\|\mathcal{R}_T\|^2]=o(M_T(\lambda_T))$ *and* (ii) $E[\|\mathcal{R}_T\|^4]=o(M_T(\lambda_T)^2)$.

SM.4.5. *Proof of Lemma A.8*

We have

$$\begin{aligned}
&\sum_{j=1}^{\infty} w_{j,T}(x) Z_{j,T} \\
&\quad = \sum_{j=1}^{\infty} w_{j,T}(x)\sqrt{T}\int_{\mathcal{S}} g_j(s)[\hat{f}_{X,Y,Z}(s) - E\hat{f}_{X,Y,Z}(s)]\, ds,
\end{aligned}$$

where $s = (x, y, z)$ and $\mathcal{S} = \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$. The estimator $\hat{f}_{X,Y,Z}(s)$ is defined by $\hat{f}_{X,Y,Z}(s) = \frac{\hat{f}_{X,Y,Z^*}(s)}{\int_{\mathcal{Z}} \hat{f}_{Z^*}(z)\, dz}$, where $\hat{f}_{X,Y,Z^*}(s)$ is the kernel estimator $\hat{f}_{X,Y,Z^*}(s) = \frac{1}{T^* h_T^d} \sum_{t=1}^{T^*} K(\frac{S_t - s}{h_T})$, $S_t = (X_t, Y_t, Z_t^*)$. We use that both T/T^* and $\int_{\mathcal{Z}} \hat{f}_{Z^*}(z)\, dz$ converge to $P[Z^* \in \mathcal{Z}]$. Then we get

$$\begin{aligned}
&\sum_{j=1}^{\infty} w_{j,T}(x)\sqrt{T}\int_{\mathcal{S}} g_j(s)\hat{f}_{X,Y,Z}(s)\, ds \\
&\quad = \frac{1}{\sqrt{T}} \sum_{j=1}^{\infty} \sum_{t=1}^{T^*} w_{j,T}(x) \frac{1}{h_T^d} \int_{\mathcal{S}} g_j(s) K\left(\frac{S_t - s}{h_T}\right) ds \cdot (1 + o_p(1)).
\end{aligned}$$

Moreover, by a change of variable,

$$\begin{aligned}
&\frac{1}{h_T^d}\int_S g_j(r) K\left(\frac{S_t - s}{h_T}\right) ds \\
&\quad = g_j(S_t) 1_S(S_t) \\
&\qquad + \int [g_j(S_t - h_T v) 1_S(S_t - h_T v) - g_j(S_t) 1_S(S_t)] K(v)\, dv \\
&\quad = g_j(S_t) 1_S(S_t) + 1_S(S_t) \int [g_j(S_t - h_T v) - g_j(S_t)] K(v)\, dv \\
&\qquad + \int g_j(S_t - h_T v)[1_S(S_t - h_T v) - 1_S(S_t)] K(v)\, dv \\
&\quad =: g_j(S_t) 1_S(S_t) + v_{j,T}(S_t) + u_{j,T}(S_t)
\end{aligned}$$

and similarly

$$\frac{1}{h_T^d}\int_S g_j(s) E\left[K\left(\frac{S - s}{h_T}\right)\right] ds = E[v_{j,T}(S)] + E[u_{j,T}(S)].$$

Since $\sum_{t=1}^{T^*} g_j(S_t)1_S(S_t) = \sum_{t=1}^{T} g_j(R_t)$, where $R_t = (X_t, Y_t, Z_t)$, the result follows if we can prove that

$$J_T(x) := \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \sum_{j=1}^{\infty} w_{j,T}(x)\{v_{j,T}(R_t) - E[v_{j,T}(R)]\} = o_p(1) \tag{SM.23}$$

and

$$I_T(x) := \frac{1}{\sqrt{T}} \sum_{t=1}^{T^*} \sum_{j=1}^{\infty} w_{j,T}(x)\{u_{j,T}(S_t) - E[u_{j,T}(S)]\} = o_p(1). \tag{SM.24}$$

Let us first focus on $J_T(x)$. Write $g_j(s) =: \frac{1}{\tau(1-\tau)}\psi_j(z)1_{\varphi_0}(w)$, $v = (\xi, \eta, \zeta)$ and

$$\begin{aligned} &g_j(R_t - h_T v) - g_j(R_t) \\ &\quad = \frac{1}{\tau(1-\tau)}\psi_j(Z_t - h_T\zeta)\big[1\{Y_t - h_T\eta \le \varphi_0(X_t - h_T\xi)\} - 1_{\varphi_0}(W_t)\big] \\ &\qquad + \frac{1}{\tau(1-\tau)}[\psi_j(Z_t - h_T\zeta) - \psi_j(Z_t)]1_{\varphi_0}(W_t). \end{aligned}$$

Then we have

$$\begin{aligned} v_{j,T}(R_t) &= \frac{1}{\tau(1-\tau)} \int \psi_j(Z_t - h_T\zeta) \\ &\quad \times \big[1\{Y_t - h_T\eta \le \varphi_0(X_t - h_T\xi)\} - 1_{\varphi_0}(W_t)\big]K(v)\,dv \\ &\quad + \frac{1}{\tau(1-\tau)}1_{\varphi_0}(W_t)\int [\psi_j(Z_t - h_T\zeta) - \psi_j(Z_t)]K(\zeta)\,d\zeta \\ &=: \bar{v}_{j,T}(R_t) + \tilde{v}_{j,T}(R_t). \end{aligned}$$

Inserting into (SM.23), this yields an analogous decomposition $J_T(x) = \bar{J}_T(x) + \tilde{J}_T(x)$ and we have to prove that both terms are $o_p(1)$.

For $\bar{J}_T(x)$, we have $E[\bar{J}_T(x)^2] = E[(\sum_{j=1}^{\infty} w_{j,T}(x)\{\bar{v}_{j,T}(R) - E[\bar{v}_{j,T}(R)]\})^2] = \sum_{j,l=1}^{\infty} w_{j,T}(x)w_{l,T}(x)\,\mathrm{Cov}[\bar{v}_{j,T}(R), \bar{v}_{l,T}(R)]$ from the i.i.d. assumption. Using the same arguments as in the proof of Lemma C.2 (Section SM.8.17), we have $E[\bar{v}_{j,T}(R)^2] \le C\|\psi_j\|^2_{L^2(F_Z)} \int 1\{|y - \varphi_0(x)| \le C\sqrt{h_T}\} f_W(w)\,dw = O(\sqrt{h_T})$. Thus, by the Cauchy–Schwarz inequality, $\mathrm{Cov}[\bar{v}_{j,T}(R), \bar{v}_{l,T}(R)] = O(\sqrt{h_T})$ uniformly in j, l. Moreover, $\sum_{j,l=1}^{\infty} |w_{j,T}(x)w_{l,T}(x)| \le (\sum_{j=1}^{\infty} |w_{j,T}(x)|)^2$. Thus, we get $E[\bar{J}_T(x)^2] = O(\sqrt{h_T}(\sum_{j=1}^{\infty} |w_{j,T}(x)|)^2)$. We use Cauchy–Schwarz inequality to get

$$\left(\sum_{j=1}^{\infty}|w_{j,T}(x)|\right)^2 = \frac{\left(\sum_{j=1}^{\infty}\frac{\sqrt{\nu_j}}{\lambda_T+\nu_j}|\phi_j(x)|\right)^2}{\sum_{j=1}^{\infty}\frac{\nu_j}{(\lambda_T+\nu_j)^2}\phi_j(x)^2}$$

$$\leq \frac{\sum_{j=1}^{\infty}\frac{\nu_j}{(\lambda_T+\nu_j)^2}|\phi_j(x)|^2 j^{\varepsilon_1}\sum_{j=1}^{\infty}j^{-\varepsilon_1}}{\sum_{j=1}^{\infty}\frac{\nu_j}{(\lambda_T+\nu_j)^2}\phi_j(x)^2}$$

for $\varepsilon_1 > 1$. Using $\frac{V_T(\lambda_T)}{\sigma_T^2(x)/T} = O(1)$, we get

$$E[\bar{J}_T(x)^2] = O\left(\sqrt{h_T}\frac{1/T\sum_{j=1}^{\infty}(\nu_j/(\lambda_T+\nu_j)^2)|\phi_j(x)|^2 j^{\varepsilon_1}}{V_T(\lambda_T)}\right).$$

Now, since $\int_0^1(\frac{1}{T}\sum_{j=1}^{\infty}\frac{\nu_j}{(\lambda_T+\nu_j)^2}|\phi_j(x)|^2 j^{\varepsilon_1})\,dx = V_T(\lambda_T;\varepsilon_1)$, we get $E[\bar{J}_T(x)^2] = O(\sqrt{h_T}\frac{V_T(\lambda_T;\varepsilon_1)}{V_T(\lambda_T)})$. We deduce $E[\bar{J}_T(x)^2] = o(1)$, which implies $\bar{J}_T(x) = o_p(1)$.

For $\tilde{J}_T(x)$, as above, $E[\tilde{J}_T(x)^2] = \sum_{j,l=1}^{\infty} w_{j,T}(x)w_{l,T}(x)\operatorname{Cov}[\tilde{v}_{j,T}(R), \tilde{v}_{l,T}(R)]$. By the Taylor theorem, using boundedness of 1_{φ_0} and Assumption A.2(i) and (ii), we get

$$\begin{aligned}&\operatorname{Cov}[\tilde{v}_{j,T}(R), \tilde{v}_{l,T}(R)]\\ &\quad = O\Big(h_T^{2m}\sup_{|\zeta|\leq\varepsilon}\sup_{\alpha\in\mathbb{N}^{d_Z}:|\alpha|=m} E\big[|\nabla^{\alpha}\psi_j(Z-\zeta)|^2\big]^{1/2}E\big[|\nabla^{\alpha}\psi_l(Z-\zeta)|^2\big]^{1/2}\Big)\end{aligned}$$

uniformly in j, l, for $\varepsilon > 0$. To bound $E[|\nabla^{\alpha}\psi_j(Z-\zeta)|^2]$, we use

$$\begin{aligned}&E\big[|\nabla^{\alpha}\psi_j(Z-\zeta)|^2\big]\\ &\quad = \int |\nabla^{\alpha}\psi_j(z)|^2 f_Z(z+\zeta)\,dz\\ &\quad = E\big[|\nabla^{\alpha}\psi_j(Z)|^2\big] + \int |\nabla^{\alpha}\psi_j(z)|^2[f_Z(z+\zeta)-f_Z(z)]\,dz\\ &\quad = E\big[|\nabla^{\alpha}\psi_j(Z)|^2\big]\\ &\qquad + \int\int |\nabla^{\alpha}\psi_j(z)|^2[f_{X,Y,Z}(w,z+\zeta)-f_{X,Y,Z}(w,z)]\,dz\,dw.\end{aligned}$$

By the mean-value theorem, Cauchy–Schwarz inequality and Assumption A.4(i), we get

$$
\begin{aligned}
&E\big[|\nabla^\alpha \psi_j(Z-\zeta)|^2\big] \\
&\quad \le E\big[|\nabla^\alpha \psi_j(Z)|^2\big] + |\zeta| \int\int |\nabla^\alpha \psi_j(z)|^2 q(w,z)\,dz\,dw \\
&\quad \le E\big[|\nabla^\alpha \psi_j(Z)|^2\big] + |\zeta| E\big[|\nabla^\alpha \psi_j(Z)|^4\big]^{1/2} \left(\int \frac{q(s)^2}{f_{XYZ}(s)}\,ds\right)^{1/2}.
\end{aligned}
$$

From Assumptions A.4(i) and A.5(iii), we get $\mathrm{Cov}[\tilde{v}_{j,T}(R), \tilde{v}_{l,T}(R)] = O(h_T^{2m} \times j^m l^m)$. Thus, $E[\tilde{J}_T(x)^2] = O(h_T^{2m}(\sum_{j=1}^\infty w_{j,T}(x) j^m)^2)$. By the same arguments as above, $E[\tilde{J}_T(x)^2] = O(h_T^{2m}\frac{V_T(\lambda_T;2m+\varepsilon_1)}{V_T(\lambda_T)}) = o(1)$, $\varepsilon_1 > 1$. Similar arguments can be used to prove (SM.24), and the conclusion follows.

SM.4.6. *Proof of Lemma A.9*

From Lemma C.1, Lemma A.9 is proved if we show

$$
\text{(SM.25)} \quad \|\hat{\mathcal{K}}_T(\Delta\hat{\varphi})\|_H = O_p\left(\frac{1}{\sqrt{\lambda_T}}\|\Delta\hat{\varphi}\|^2\right).
$$

From the proof of Lemma A.5, we know that

$$
\begin{aligned}
\|\hat{\mathcal{K}}_T(\Delta\hat{\varphi})\|_H \le{}& \frac{1}{\sqrt{\lambda_T}} \frac{1}{2\sqrt{\tau(1-\tau)}} \\
&\times \sup_{x\in\mathcal{X}, y\in\mathbb{R}, z\in\mathcal{Z}} \big|\nabla_y \hat{f}_{X,Y|Z}(x,y|z)\big| \|\Delta\hat{\varphi}\|^2, \quad P\text{-a.s.},
\end{aligned}
$$

and from Hansen (2008), we know that $\sup_{x\in\mathcal{X}, y\in\mathbb{R}, z\in\mathcal{Z}} |\nabla_y \hat{f}_{X,Y|Z}(x,y|z) - \nabla_y f_{X,Y|Z}(x,y|z)|^2 = O_p(\frac{\log T}{T h_T^{4+d_Z}} + h_T^{2m})$. Since $\eta < \frac{1}{4+d_Z}$ and by using Assumption A.4(iii), $\sup_{x\in\mathcal{X}, y\in\mathbb{R}, z\in\mathcal{Z}} |\nabla_y \hat{f}_{X,Y|Z}(x,y|z)| = O_p(1)$, the conclusion follows.

SM.4.7. *Proof of Lemma A.10*

Using the same arguments as in the proof of Lemma A.8, we have $\sqrt{T/\sigma_T^2(x)}(\lambda_T + A^*A)^{-1}A^*E\hat{\zeta}(x) = \sqrt{T}\sum_{j=1}^\infty w_{jT}(x)\frac{1}{h_T^d}\int_{\mathcal{S}} g_j(s)E[K(\frac{S-s}{h_T})]\,ds \cdot (1+o(1))$. Further, we can use $\frac{1}{h_T^d}\int_{\mathcal{S}} g_j(s)E[K(\frac{S-s}{h_T})]\,ds = \int_{\mathcal{S}} g_j(s)\int K(v) \times [f_{XYZ^*}(s+h_T v) - f_{XYZ^*}(s)]\,dv\,ds$ and Assumptions A.2 and A.4(i) to get $\frac{1}{h_T^d}\int_{\mathcal{S}} g_j(s)E[K(\frac{S-s}{h_T})]\,ds = O(h_T^m)$ uniformly in j. From $\sum_{j=1}^\infty |w_{j,T}(x)| =$

$O(\frac{1}{\sigma_T(x)}\frac{1}{\lambda_T})$, we get $\sqrt{T/\sigma_T^2(x)}(\lambda_T + A^*A)^{-1}A^*E\hat{\zeta}(x) = O(\frac{1}{\sqrt{\sigma_T^2(x)/T}}\frac{h_T^m}{\lambda_T})$. By using $\frac{1}{\sigma_T^2(x)/T}\frac{h_T^{2m}}{\lambda_T^2} = O(\frac{1}{V_T(\lambda_T)}\frac{h_T^{2m}}{\lambda_T^2}) = O(\frac{h_T^{2m}}{\lambda_T^4}) = o(1)$, the conclusion follows.

SM.4.8. *Proof of Lemma A.11*

First, by the same argument as in the proof of Lemma A.9 (use of Lemma C.1), it is enough to show that $\sqrt{T/\sigma_T^2(x)}\|\mathcal{R}_T\|_H = o_p(1)$. Second, from the proof of Lemma B.10, we have the decomposition $\mathcal{R}_T = \sum_{j=1}^5 \mathcal{R}_{j,T}$, where

$$\|\mathcal{R}_{1,T}\|_H = O_p\bigg(\frac{a_T^{1/2}}{\lambda_T}b(\lambda_T)\bigg),$$

$$\|\mathcal{R}_{2,T}\|_H = O_p\bigg(\frac{a_T^{1/2}}{\lambda_T}b(\lambda_T)\bigg),$$

$$\|\mathcal{R}_{3,T}\|_H = O_p\bigg(\frac{a_T^{1/2}}{\lambda_T}\sqrt{V_T(\lambda_T)}\bigg),$$

$$\|\mathcal{R}_{4,T}\|_H = O_p\bigg(\frac{1}{\lambda_T}\bigg(\frac{(\log T)^2}{Th_T^{d_Z+1}} + h_T^m\bigg)\bigg),$$

and

$$\|\mathcal{R}_{5,T}\|_H = O_p\bigg(\frac{1}{\lambda_T}\|\Delta\hat{\varphi}\|\sqrt{\frac{\log T}{Th_T^{d_Z}} + h_T^{2m}}\bigg), \quad a_T := \frac{1}{Th_T^2} + h_T^{2m}.$$

Then using $b(\lambda_T) = O(\lambda_T^\delta) = O(\sqrt{V_T(\lambda_T)})$ and $\|\Delta\hat{\varphi}\| = O_p(M_T(\lambda_T)^{1/2}) = O_p(\sqrt{V_T(\lambda_T)})$, by Lemma A.7, we get

$$\|\mathcal{R}_T\|_H = O_p\bigg(\frac{a_T^{1/2}}{\lambda_T}\sqrt{V_T(\lambda_T)}\bigg) + O_p\bigg(\frac{1}{\lambda_T}\bigg(\frac{(\log T)^2}{Th_T^{d_Z+1}} + h_T^m\bigg)\bigg) + O_p\bigg(\frac{\sqrt{V_T(\lambda_T)}}{\lambda_T}\sqrt{\frac{\log T}{Th_T^{d_Z}} + h_T^{2m}}\bigg).$$

The conclusion follows since

$$\sqrt{T/\sigma_T^2(x)}\frac{a_T^{1/2}}{\lambda_T}\sqrt{V_T(\lambda_T)} = o\bigg(\frac{a_T^{1/2}}{\lambda_T}\bigg) = o(1),$$

$$\frac{\sqrt{T/\sigma_T^2(x)}}{\lambda_T}\bigg(\frac{(\log T)^2}{Th_T^{d_Z+1}} + h_T^m\bigg) = o\bigg(\frac{\sqrt{V_T(\lambda_T)}}{\lambda_T}\bigg) = o(1),$$

and

$$\sqrt{T/\sigma_T^2(x)}\frac{\sqrt{V_T(\lambda_T)}}{\lambda_T}\sqrt{\frac{\log T}{Th_T^{d_Z}}+h_T^{2m}} = O\left(\frac{1}{\lambda_T}\sqrt{\frac{\log T}{Th_T^{d_Z}}+h_T^{2m}}\right) = o(1).$$

SM.5. ESTIMATION OF THE ASYMPTOTIC VARIANCE

SM.5.1. *Proof of Lemma A.12*

We have

$$\hat{\sigma}_T^2(x) - \sigma_{0,T}^2(x) = \sum_{j=1}^{N_T}\frac{\nu_j}{(\nu_j+\lambda_T)^2}(\hat{\phi}_j(x)^2 - \phi_j(x)^2) + \sum_{j=1}^{N_T}\left(\frac{\hat{\nu}_j}{(\hat{\nu}_j+\lambda_T)^2} - \frac{\nu_j}{(\nu_j+\lambda_T)^2}\right)\hat{\phi}_j(x)^2.$$

Moreover,

$$\frac{\hat{\nu}_j}{(\hat{\nu}_j+\lambda_T)^2} - \frac{\nu_j}{(\nu_j+\lambda_T)^2} = \frac{1}{(\hat{\nu}_j+\lambda_T)^2(\nu_j+\lambda_T)^2}(\lambda_T^2 - \hat{\nu}_j\nu_j)(\hat{\nu}_j - \nu_j)$$

and

$$\left|\frac{\hat{\nu}_j}{(\hat{\nu}_j+\lambda_T)^2} - \frac{\nu_j}{(\nu_j+\lambda_T)^2}\right| \le \frac{1}{(\nu_j+\lambda_T)^2}|\hat{\nu}_j - \nu_j| + \frac{1}{(\hat{\nu}_j+\lambda_T)(\nu_j+\lambda_T)}|\hat{\nu}_j - \nu_j|.$$

Thus,

$$\delta_T \le \frac{1}{\sigma_T^2(x)}\sum_{j=1}^{N_T}\frac{\nu_j}{(\nu_j+\lambda_T)^2}|\hat{\phi}_j(x)^2 - \phi_j(x)^2| + \frac{1}{\sigma_T^2(x)}\sum_{j=1}^{N_T}\frac{1}{(\nu_j+\lambda_T)^2}|\hat{\nu}_j - \nu_j|\hat{\phi}_j(x)^2 + \frac{1}{\sigma_T^2(x)}\sum_{j=1}^{N_T}\frac{1}{(\hat{\nu}_j+\lambda_T)(\nu_j+\lambda_T)}|\hat{\nu}_j - \nu_j|\hat{\phi}_j(x)^2. \qquad \text{(SM.26)}$$

For any $j \le N_T$, we have $|\hat{\phi}_j(x)^2 - \phi_j(x)^2| \le \xi_j^* \epsilon_{2,T}$, and $|\hat{\nu}_j - \nu_j|\hat{\phi}_j(x)^2 \le \nu_j \epsilon_{1,T}(\phi_j(x)^2 + \xi_j^* \epsilon_{2,T})$. Moreover, by Cauchy–Schwarz inequality,

$$\begin{aligned}
&\sum_{j=1}^{N_T} \frac{1}{(\hat{\nu}_j + \lambda_T)(\nu_j + \lambda_T)} |\hat{\nu}_j - \nu_j| \hat{\phi}_j(x)^2 \\
&\quad \le \sum_{j=1}^{N_T} \frac{\sqrt{\nu_j \hat{\nu}_j} |\hat{\phi}_j(x)| (|\phi_j(x)| + \sqrt{\xi_j^* \epsilon_{2,T}})}{(\hat{\nu}_j + \lambda_T)(\nu_j + \lambda_T)} \frac{\epsilon_{1,T}}{\sqrt{1 - \epsilon_{1,T}}} \\
&\quad \le \left(\sum_{j=1}^{N_T} \frac{\hat{\nu}_j \hat{\phi}_j(x)^2}{(\hat{\nu}_j + \lambda_T)^2} \right)^{1/2} \\
&\qquad \times \left[\left(\sum_{j=1}^{N_T} \frac{\nu_j \phi_j(x)^2}{(\nu_j + \lambda_T)^2} \right)^{1/2} + \sqrt{\epsilon_{2,T}} \left(\sum_{j=1}^{N_T} \frac{\nu_j \xi_j^*}{(\nu_j + \lambda_T)^2} \right)^{1/2} \right] \\
&\qquad \times \frac{\epsilon_{1,T}}{\sqrt{1 - \epsilon_{1,T}}}.
\end{aligned}$$

Thus, from (SM.26), we get

$$\begin{aligned}
\delta_T \le{}& \frac{\sigma_{*,0,T}^2(x)}{\sigma_T^2(x)} (1 + \epsilon_{1,T}) \epsilon_{2,T} + \frac{\sigma_{0,T}^2(x)}{\sigma_T^2(x)} \epsilon_{1,T} \\
&+ \frac{\hat{\sigma}_T(x) \sigma_{0,T}(x)}{\sigma_T^2(x)} \frac{\epsilon_{1,T}}{\sqrt{1 - \epsilon_{1,T}}} + \frac{\hat{\sigma}_T(x) \sigma_{*,0,T}(x)}{\sigma_T^2(x)} \frac{\epsilon_{1,T} \sqrt{\epsilon_{2,T}}}{\sqrt{1 - \epsilon_{1,T}}},
\end{aligned}$$

where $\sigma_{*,0,T}^2(x) := \sum_{j=1}^{N_T} \frac{\nu_j \xi_j^*}{(\nu_j + \lambda_T)^2}$. Now, by using $\sigma_{0,T}^2(x) \le \sigma_T^2(x)$, $\sigma_{*,0,T}^2(x) \le \sigma_{*,T}^2(x) \le C \sigma_T^2(x)$ for a constant C, and $\hat{\sigma}_T^2(x) \le \sigma_T^2(x)(\delta_T + 1 - \xi_T)$, the conclusion follows.

SM.5.2. *Proof of Lemma A.13*

Point (i) follows from Lemma 4.2 in Bosq (2000). Point (ii) follows from Lemma 4.3 in Bosq (2000) by using that $\|\hat{\phi}_j - \phi_j\| \le \|\hat{\phi}_j - \phi_j\|_H$.

SM.5.3. *Proof of Lemma A.14*

Let $\bar{A} = D\hat{\mathcal{A}}(\bar{\varphi})$ and $\hat{A}_0 = D\hat{\mathcal{A}}(\varphi_0)$ be the Frechet derivatives of operator $\hat{\mathcal{A}}$ at $\bar{\varphi}$ and φ_0, respectively. We have

$$\hat{D} = \bar{A}^* \bar{A} - A^* A = (\hat{A}_0^* \hat{A}_0 - A^* A) + (\bar{A}^* \bar{A} - \hat{A}_0^* \hat{A}_0).$$

Let us consider the first term, $\hat{A}_0^* \hat{A}_0 - A^* A$, that does not involve the pilot estimator $\bar{\varphi}$. We have $\hat{A}_0^* \hat{A}_0 - A^* A = \mathcal{E}\mathcal{D}^{-1}(\hat{\tilde{A}}_0 \hat{A}_0 - \tilde{A} A)$, where

$$\begin{aligned}
&(\hat{\tilde{A}}_0 \hat{A}_0 - \tilde{A} A)\phi(x) \\
&\quad = \frac{1}{\tau(1-\tau)} \int_0^1 \left\{ \frac{1}{T} \sum_{t=1}^{T} \hat{f}(x, \varphi_0(x)|Z_t) \hat{f}(\xi, \varphi_0(\xi)|Z_t) \right. \\
&\qquad \left. - \int f(x, \varphi_0(x)|z) f(\xi, \varphi_0(\xi)|z) f(z)\, dz \right\} \phi(\xi)\, d\xi.
\end{aligned}$$

Since operator $\mathcal{E}\mathcal{D}^{-1} : L^2[0,1] \to H^l[0,1]$ is bounded w.r.t. the norms $\|\cdot\|$ in $L^2[0,1]$ and $\|\cdot\|_H$ in $H^l[0,1]$ (see Lemmas A.17(iii) and A.18(ii) in Section SM.3), we get

$$\|\hat{A}_0^* \hat{A}_0 - A^* A\|_H \le C \|\hat{\tilde{A}}_0 \hat{A}_0 - \tilde{A} A\|.$$

From Lemma C.8 (see Section SM.8.10.2) we have

$$\|\hat{\tilde{A}}_0 \hat{A}_0 - \tilde{A} A\| = O_p\left(\frac{1}{\sqrt{T h_T^2}} + h_T^m \right).$$

Thus, $\|\hat{A}_0^* \hat{A}_0 - A^* A\|_H = O_p(\frac{1}{\sqrt{T h_T^2}} + h_T^m)$.

Let us now consider the term $\bar{A}^* \bar{A} - \hat{A}_0^* \hat{A}_0$ that involves the pilot estimator. It is equal to $\bar{A}^* \bar{A} - \hat{A}_0^* \hat{A}_0 = \mathcal{E}\mathcal{D}^{-1}(\bar{\tilde{A}} \bar{A} - \hat{\tilde{A}}_0 \hat{A}_0)$, where

$$\begin{aligned}
&(\bar{\tilde{A}} \bar{A} - \hat{\tilde{A}}_0 \hat{A}_0)\phi(x) \\
&\quad = \frac{1}{\tau(1-\tau)} \int_0^1 \left\{ \frac{1}{T} \sum_{t=1}^{T} \big[\hat{f}(x, \bar{\varphi}(x)|Z_t) \hat{f}(\xi, \bar{\varphi}(\xi)|Z_t) \right. \\
&\qquad \left. - \hat{f}(x, \varphi_0(x)|Z_t) \hat{f}(\xi, \varphi_0(\xi)|Z_t) \big] \right\} \phi(\xi)\, d\xi.
\end{aligned}$$

By the convergence of the sample average over the Z_t's and the convergence of the kernel, the dominating term in $\bar{A}^* \bar{A} - \hat{A}_0^* \hat{A}_0$ is $A_{\bar{\varphi}}^* A_{\bar{\varphi}} - A^* A$, where $A_{\bar{\varphi}} = D\mathcal{A}(\bar{\varphi})$ is the Frechet derivative of $\mathcal{A}$ at $\bar{\varphi}$. Split $A_{\bar{\varphi}}^* A_{\bar{\varphi}} - A^* A$ as

$$A_{\bar{\varphi}}^* A_{\bar{\varphi}} - A^* A = (A_{\bar{\varphi}}^* - A^*) A + A^* (A_{\bar{\varphi}} - A) + (A_{\bar{\varphi}}^* - A^*)(A_{\bar{\varphi}} - A).$$

Now, from the mean-value theorem and the Cauchy–Schwarz inequality, for any $\phi \in L^2[0,1]$, we have

$$\begin{aligned}|(A_{\bar{\varphi}} - A)\phi(z)| &\le \int \big|f(x, \bar{\varphi}(x)|z) - f(x, \varphi_0(x)|z)\big||\phi(x)|\,dx \\ &\le \sup_{x,y,z}\big|\nabla_y f(x, y|z)\big| \int |\bar{\varphi}(x) - \varphi_0(x)||\phi(x)|\,dx \\ &\le C_1\|\bar{\varphi} - \varphi_0\|\|\phi\|,\end{aligned}$$

where $C_1 := \sup_{x,y,z}|\nabla_y f(x, y|z)| < \infty$ by Assumption A.4(iii). Thus, $\|A_{\bar{\varphi}} - A\| \le C_1\|\bar{\varphi} - \varphi_0\|$. We get $\|A^*_{\bar{\varphi}}A_{\bar{\varphi}} - A^*A\|_H = O_p(\|\bar{\varphi} - \varphi_0\|)$. The conclusion follows.

SM.5.4. *Proof of Lemma A.15*

To prove parts (i) and (ii) of Lemma A.15, we use the next lemma.

LEMMA B.11: *Let $v_j' = -(j, \log j)$. Then* (i) *matrix $D_{(T)}(V_{(T)}'V_{(T)})D_{(T)}$ converges to a positive definite matrix as $n_T \to \infty$, where $V_{(T)} = (v_{n_T/2} - v_{n_T}, \dots, v_{n_T-1} - v_{n_T})'$ and $D_{(T)}$ denotes the diagonal matrix with diagonal elements $D_{(T),kk} = (\sum_{j=n_T/2}^{n_T-1}(v_{j,k} - v_{n_T,k}))^{-1/2}$;* (ii) *$\zeta_{(T)}' d_{(T)} d_{(T)}' m_{(T)} = O(1)$ as $n_T, N_T \to \infty$ such that $N_T = O(n_T)$, where $d_{(T)}$ is the vector with elements $d_{(T),k} = D_{(T),kk}$, vector $\zeta_{(T)}$ has elements $\zeta_{(T),k} = \sup_{n_T<j\le N_T}|v_{j,k} - v_{n_T,k}|$, and vector $m_{(T)}$ has elements $m_{(T),k} = \sum_{j=n_T/2}^{n_T-1}|v_{j,k} - v_{n_T,k}|$.*

The results in Lemma B.11 hold also for $w_j = -\log j$, which is an element of v_j.

SM.5.4.1. *Proof of Part (i)*

We have $\hat{\nu}_j = \bar{\nu}_{n_T}\exp((v_j - v_{n_T})'\hat{\alpha})$ and $\nu_j = \nu_{n_T}\exp((v_j - v_{n_T})'\alpha)c_{1,j}/c_{1,n_T}$. Thus

$$\frac{\hat{\nu}_j}{\nu_j} = \frac{\bar{\nu}_{n_T}}{\nu_{n_T}}\exp\big((v_j - v_{n_T})'(\hat{\alpha} - \alpha)\big)\frac{c_{1,j}}{c_{1,n_T}}.$$

Since $\frac{\bar{\nu}_{n_T}}{\nu_{n_T}} - 1 = o_p(1)$ from Appendix A.5, the conclusion follows if

$$\sup_{n_T<j\le N_T}\left|\frac{c_{1,j}}{c_{1,n_T}} - 1\right| = o(1)$$

and

$$\sup_{n_T<j\le N_T}\big|\exp\big((v_j - v_{n_T})'(\hat{\alpha} - \alpha)\big) - 1\big| = o_p(1).$$

The first condition is satisfied since

$$\sup_{n_T<j\leq N_T}\left|\frac{c_{1,j}}{c_{1,n_T}}-1\right|\leq\frac{1}{c_{1,n_T}}\sup_{n_T<j}|c_{1,j}-\bar{c}_1|+\frac{|c_{1,n_T}-\bar{c}_1|}{c_{1,n_T}}=o(1)$$

by the convergences $c_{1,j}\to\bar{c}_1$, with $\bar{c}_1>0$, and $n_T\to\infty$. Let us now consider the second condition. We have $\hat{\alpha}-\alpha=(V'_{(T)}V_{(T)})^{-1}V'_{(T)}a_{(T)}$, where matrix $V_{(T)}$ is defined in Lemma B.11 and vector $a_{(T)}=(a_{n_T/2},\dots,a_{n_T-1})'$ with $a_j=\log(\bar{\nu}_j/\bar{\nu}_{n_T})-\log(\nu_j/\nu_{n_T})+\log(c_{1,j}/c_{1,n_T})$. By using that $|e^x-1|\leq 2|x|$ for small x, we get

$$\text{(SM.27)}\quad \sup_{n_T<j\leq N_T}\left|\exp\left((v_j-v_{n_T})'(\hat{\alpha}-\alpha)\right)-1\right| = O_p\left(\sup_{n_T<j\leq N_T}\left|(v_j-v_{n_T})'\left(V'_{(T)}V_{(T)}\right)^{-1}V'_{(T)}a_{(T)}\right|\right).$$

Let us now bound the RHS. From Lemma B.11(ii), vector $v_j-v_{n_T}$ is bounded componentwise by $\zeta_{(T)}$ for any j such that $n_T<j\leq N_T$. Moreover, the lth component of $V'_{(T)}a_{(T)}$ is such that

$$\left[V'_{(T)}a_{(T)}\right]_l=\sum_{j=n_T/2}^{n_T-1}(v_{j,l}-v_{n_T,l})a_j=O_p\left(m_{(T),l}\bar{a}_T\right),$$

where $\bar{a}_T=\sup_{n_T/2\leq j<n_T}|a_j|$ and vector $m_{(T)}$ is defined in Lemma B.11(ii). From $(V'_{(T)}V_{(T)})^{-1}=D_{(T)}(D_{(T)}(V'_{(T)}V_{(T)})D_{(T)})^{-1}D_{(T)}$ and Lemma B.11(i) we get

$$\left(V'_{(T)}V_{(T)}\right)^{k,l}=O\left(d_{(T),k}d_{(T),l}\right),$$

where $(V'_{(T)}V_{(T)})^{k,l}$ denotes the (k,l) element of matrix $(V'_{(T)}V_{(T)})^{-1}$. We deduce

$$\sup_{n_T<j\leq N_T}\left|(v_j-v_{n_T})'\left(V'_{(T)}V_{(T)}\right)^{-1}V'_{(T)}a_{(T)}\right|=O_p\left(\zeta'_{(T)}d_{(T)}d'_{(T)}m_{(T)}\bar{a}_T\right).$$

By (SM.27) and Lemma B.11(ii), we get

$$\sup_{n_T<j\leq N_T}\left|\exp\left((v_j-v_{n_T})'(\hat{\alpha}-\alpha)\right)-1\right|=O_p(\bar{a}_T).$$

By using $a_j=\log(1+\frac{\bar{\nu}_j-\nu_j}{\nu_j})-\log(1+\frac{\bar{\nu}_{n_T}-\nu_{n_T}}{\nu_{n_T}})+\log(c_{1,j}/c_{1,n_T})$ and $|\log(1+x)|\leq 2|x|$ for small x, we get

$$\bar{a}_T=O_p\left(\sup_{1\leq j\leq n_T}\frac{|\bar{\nu}_j-\nu_j|}{\nu_j}\right)+o(1)=o_p(1)$$

from Appendix A.5. The conclusion follows.

SM.5.4.2. *Proof of Part (ii)*

Without loss of generality, we can normalize the periodic component such that $\sum_{j=1}^{S}\chi_j = 1$. Let $\phi_{S,j}(x)^2 = \sum_{k=0}^{S-1}\phi_{j-k}(x)^2$.

LEMMA B.12: *We have* $\phi_{S,j}(x)^2 = c_{2,j}^{*}\exp(w_j\beta)$, *where* $c_{2,j}^{*}/c_{2,j} = 1 + o(1)$ *as* $j \to \infty$.

From Lemma B.12, the filtered spectral coefficients $\phi_{S,j}(x)^2$ satisfy a decay behavior compatible with Assumption 5(i) with v_j replaced by $w_j = -\log j$. By applying Lemma B.11 to w_j and an argument similar to part (i), we deduce

$$\sup_{n_T < j \le N_T} \frac{|\hat{\phi}_{S,j}(x)^2 - \phi_{S,j}(x)^2|}{\phi_{S,j}(x)^2} = o_p(1), \tag{SM.28}$$

where $\hat{\phi}_{S,j}(x)^2 = \bar{\phi}_{S,n_T}(x)^2\exp((w_j - w_{n_T})\hat{\beta})$ for $n_T < j \le N_T$. The consistency of the estimator of the periodic component is proved in the next lemma.

LEMMA B.13: *We have* $\hat{\chi}_j - \chi_j = o_p(1)$ *for* $j = 1, \dots, S$.

Now, by using $\hat{\phi}_j(x)^2 = \hat{\phi}_{S,j}(x)^2\hat{\chi}_{j \bmod S}$ and $\phi_j(x)^2 = \xi_j^{*}\chi_j$, where $\xi_j^{*} = c_{2,j}\exp(w_j\beta)$, we have

$$\begin{aligned}\hat{\phi}_j(x)^2 - \phi_j(x)^2 &= [\hat{\phi}_{S,j}(x)^2 - \phi_{S,j}(x)^2]\hat{\chi}_{j \bmod S} \\ &\quad + \hat{\chi}_{j \bmod S}[\phi_{S,j}(x)^2 - \xi_j^{*}] + \xi_j^{*}(\hat{\chi}_{j \bmod S} - \chi_j).\end{aligned}$$

By combining bound (SM.28) and Lemma B.13, and using $\phi_{S,j}(x)^2/\xi_j^{*} = 1 + o(1)$ from Lemma B.12, the conclusion follows.

SM.6. CHARACTERIZATION OF THE ASYMPTOTIC MISE

The asymptotic MISE of $\hat{\varphi}$ can be characterized under a strengthening of Assumption A.5(iii):

ASSUMPTION A.5(iii′): *Functions* ψ_j *are in class* $C^m(\mathbb{R}^{d_Z})$ *such that*

$$\sup_{j \in N} E\big[|\nabla^{\alpha}\psi_j(Z)|^{\bar{s}}\big]^{1/\bar{s}} < \infty$$

for $\bar{s} \ge 4$ *and any* $\alpha \in \mathbb{N}^{d_Z}$ *with* $|\alpha| \le m$.

Then, under Assumptions A.1(i)–A.5(iii′), we have $E[\|\hat{\varphi} - \varphi_0\|^2] = M_T(\lambda_T)(1 + o(1))$. The proof follows by the same arguments as in the proof of Lemma A.7 and by replacing Lemma B.8(i) with the next lemma.

LEMMA B.14: *Under Assumptions* A.1, A.2, A.4(i), A.5(i)–(iii′),

$$E\big[\|(\lambda_T + A^*A)^{-1}A^*(\hat{\zeta} - E\hat{\zeta})\|^2\big]$$
$$= \frac{1}{T}\sum_{j=1}^{\infty}\frac{\nu_j}{(\nu_j + \lambda_T)^2}\|\phi_j\|^2(1 + o(1)).$$

SM.7. AN EXAMPLE

SM.7.1. *The Model*

Let us define the functions $\tilde{\phi}_j$ by $\tilde{\phi}_1(x) = 1$ and $\tilde{\phi}_j(x) = \sqrt{2}\cos[\pi(j-1)x]$ for $j \geq 2$. They build a complete orthonormal system in $L^2[0,1]$ w.r.t. norm $\|\cdot\|$. Let further $\mu_j = e^{-\alpha_1(j-1)}$ for $j \geq 1$ and $\alpha_1 > 0$. Let us define the conditional density of (X, Z) given $U = \tau$ by the bivariate p.d.f. (see Hall and Horowitz (2005) and Horowitz and Lee (2007) for similar constructions)

$$f_{X,Z|U}(x, z|\tau) = \sum_{j=1}^{\infty}\mu_j^{1/2}\tilde{\phi}_j(x)\tilde{\phi}_j(z), \quad x, z \in [0,1]. \tag{SM.29}$$

A sufficient condition for function $f_{X,Z|U}$ to be positive is $\sum_{j=2}^{\infty}\mu_j^{1/2} \leq 1/2$, that is, $\alpha_1 \geq 2\log 3$. Moreover, by using that $\int_0^1 \tilde{\phi}_j(x)\,dx = 0$ for $j \geq 2$, it follows that the total mass of $f_{X,Z|U}$ is 1 and that the distributions of X given $U = \tau$ and of Z given $U = \tau$ are uniform.

We consider the separable model

$$Y = \varphi_0(X) + U^*,$$

where $U^* = G^{-1}(U)$, variable U is uniformly distributed on [0, 1] and independent of Z, the distribution of (X, Z) given $U = \tau$ admits the p.d.f. (SM.29), and G is a c.d.f. such that $G(0) = \tau$.

SM.7.2. *The Spectrum of Operator* $\tilde{A}A$

The operator A is given by

$$A\varphi(z) = G'(0)\int_0^1 f_{X|Z,U}(x|z,\tau)\varphi(x)\,dx,$$

and its adjoint w.r.t. the L^2 norm is

$$\tilde{A}\psi(x) = \frac{G'(0)}{\tau(1-\tau)}\int_0^1 f_{Z|X,U}(z|x,\tau)\psi(z)\,dz.$$

Then the operator $\tilde{A}A$ is such that

$$\tilde{A}A\varphi(x) = \frac{G'(0)^2}{\tau(1-\tau)} \int_0^1 a(x,\xi)\varphi(\xi)\,d\xi,$$

where

$$a(x,\xi) = \int_0^1 f_{Z|X,U}(z|x,\tau) f_{X|Z,U}(\xi|z,\tau)\,dz = \sum_{j=1}^{\infty} \mu_j \tilde{\phi}_j(x)\tilde{\phi}_j(\xi).$$

Thus, the operator $\tilde{A}A$ admits normalized eigenfunctions $\tilde{\phi}_j$ in $L^2[0,1]$ and eigenvalues

$$\tilde{\nu}_j = \frac{G'(0)^2}{\tau(1-\tau)}\mu_j = \frac{G'(0)^2}{\tau(1-\tau)} e^{-\alpha_1(j-1)}.$$

Functions $\tilde{\phi}_j$ are eigenfunctions of operator $\mathcal{D} = 1 - \nabla^2$ such that $\mathcal{D}\tilde{\phi}_j = [1+\pi^2(j-1)^2]\tilde{\phi}_j$. Moreover, we have

$$\|\tilde{\phi}_j\|_H^2 = \langle \tilde{\phi}_j, \mathcal{D}\tilde{\phi}_j\rangle = 1 + \pi^2(j-1)^2.$$

SM.7.3. *The Spectrum of Operator A^*A*

The operator $A^*A = \mathcal{D}^{-1}\tilde{A}A$ admits normalized eigenfunctions

$$\phi_j(x) = \frac{1}{\sqrt{1+\pi^2(j-1)^2}}\tilde{\phi}_j(x) = \begin{cases} 1, & j = 1, \\ \sqrt{\dfrac{2}{1+\pi^2(j-1)^2}}\cos[\pi(j-1)x], & j \geq 2, \end{cases}$$

in $H^1[0,1]$ and eigenvalues

$$\nu_j = \frac{G'(0)^2}{\tau(1-\tau)} \frac{1}{1+\pi^2(j-1)^2} e^{-\alpha_1(j-1)}.$$

Moreover,

$$\|\phi_j\|^2 = \frac{1}{1+\pi^2(j-1)^2}.$$

The asymptotic behavior of the spectrum is such that

$$\nu_j \asymp j^{-2}e^{-\alpha_1 j}, \quad \phi_j(x)^2 \asymp j^{-2}\cos^2[\pi(j-1)x], \quad \|\phi_j\|^2 \asymp j^{-2}.$$

Thus, ν_j and $\phi_j(x)^2$ satisfy Assumption 5, and $\|\phi_j\|^2 \asymp j^{-\beta}$ with $\beta = 2$.

SM.7.4. *Asymptotic Behavior of $V_T(\lambda_T;\varepsilon)$, $\sigma_T^2(x)$, and $\sigma_{*,T}^2(x)$*

The asymptotic behavior of $V_T(\lambda_T;\varepsilon)$, $\sigma_T^2(x)$, and $\sigma_{*,T}^2(x)$ as $T\to\infty$ can be derived using the next lemma, which is a generalization of Lemma A.8 in GS.

LEMMA B.15: *Let $\nu_j \asymp j^{-\alpha_2}e^{-\alpha_1 j}$ and $a_j \asymp j^{-\alpha_3}e^{-\alpha_4 j}\chi_j$ for $\alpha_2,\alpha_3\geq 0$, $\alpha_1,\alpha_4>0$, where χ_j is positive, bounded, and such that $\sup_{i\geq 1:\chi_i=0}\min_{j\geq 1:\chi_j\geq\varepsilon}|j-i|<\infty$ for some $\varepsilon>0$. Let $n_\lambda\in\mathbb{N}$ be such that $\nu_{n_\lambda}\asymp\lambda$ as $\lambda\to 0$. Then, as $\lambda\to 0$,*

$$\sum_{j=1}^{\infty}\frac{a_j}{(\lambda+\nu_j)^2}\asymp\begin{cases}\lambda^{-2+\alpha_4/\alpha_1}n_\lambda^{\alpha_2\alpha_4/\alpha_1-\alpha_3}, & \text{if } \alpha_4<2\alpha_1,\\ 1, & \text{if } \alpha_4>2\alpha_1.\end{cases}$$

By using $a_j=\nu_j\|\phi_j\|^2 j^{\varepsilon}$, $\alpha_3=\alpha_2+\beta-\varepsilon$, $\alpha_4=\alpha_1$, and $n_\lambda\asymp\log(1/\lambda)$, we get

$$V_T(\lambda_T;\varepsilon)=\frac{1}{T}\sum_{j=1}^{\infty}\frac{\nu_j}{(\lambda_T+\nu_j)^2}\|\phi_j\|^2 j^{\varepsilon}\asymp\frac{1}{T}\lambda_T^{-1}[\log(1/\lambda_T)]^{-\beta+\varepsilon}.$$

By using $a_j=\nu_j\phi_j(x)^2=\nu_j c_{2,j}j^{-\beta}\chi_j$, $\alpha_3=\alpha_2+\beta$, $\alpha_4=\alpha_1$, and $n_\lambda\asymp\log(1/\lambda)$, we get

$$\sigma_T^2(x)=\sum_{j=1}^{\infty}\frac{\nu_j}{(\lambda_T+\nu_j)^2}\phi_j(x)^2\asymp\lambda_T^{-1}[\log(1/\lambda_T)]^{-\beta},$$

where we have used that a periodic function χ_j satisfies the condition in Lemma B.15. Finally, by using $a_j=\nu_j c_{2,j}j^{-\beta}$, $\alpha_3=\alpha_2+\beta$, $\alpha_4=\alpha_1$, and $n_\lambda\asymp\log(1/\lambda)$, we get

$$\sigma_{*,T}^2(x)=\sum_{j=1}^{\infty}\frac{\nu_j}{(\lambda_T+\nu_j)^2}c_{2,j}j^{-\beta}\asymp\lambda_T^{-1}[\log(1/\lambda_T)]^{-\beta}.$$

Thus, $\frac{V_T(\lambda_T)}{\sigma_T^2(x)/T}=O(1)$ and $\sigma_{*,T}^2(x)/\sigma_T^2(x)=O(1)$.

SM.7.5. *Hyberbolic Spectrum*

A similar example can be developed if we choose $\mu_1=1$ and $\mu_j=Cj^{-\alpha_1}$, $j>1$, with $C>0$ and $\alpha_1>2$, in (SM.29). The p.d.f. (SM.29) is well defined if $C\leq(\alpha_1-2)/4$. The eigenvalues of $\tilde{A}A$ are

$$\tilde{\nu}_j=\frac{G'(0)^2}{\tau(1-\tau)}\mu_j.$$

SM.8. PROOF OF THE SECONDARY TECHNICAL LEMMAS

SM.8.1. *Proof of Lemma B.1*

Let $T^* = T$ to ease notation. Write

$$f_{X|Z}(x|z)F_{Y|X,Z}(y|x,z) = \frac{\Psi(x,y,z)}{f_{Z^*}(z)}$$

and similarly

$$\hat{f}_{X|Z}(x|z)\hat{F}_{Y|X,Z}(y|x,z) = \frac{\hat{\Psi}(x,y,z)}{\hat{f}_{Z^*}(z)},$$

where

$$\Psi(x,y,z) := \int_{-\infty}^{y} f_{X,Y,Z^*}(x,v,z)\,dv,$$

$$\begin{aligned}\hat{\Psi}(x,y,z) &:= \int_{-\infty}^{y} \hat{f}_{X,Y,Z^*}(x,v,z)\,dv \\ &= \frac{1}{Th_T^{1+d_Z}}\sum_{t=1}^{T} IK\left(\frac{y-Y_t}{h_T}\right)K\left(\frac{x-X_t}{h_T}\right)K\left(\frac{z-Z_t^*}{h_T}\right),\end{aligned}$$

and $IK(y) := \int_{-\infty}^{y} K(v)\,dv$. Since

$$\begin{aligned}&\frac{\hat{\Psi}(x,y,z)}{\hat{f}_{Z^*}(z)} - \frac{\Psi(x,y,z)}{f_{Z^*}(z)} \\ &\quad= \frac{1}{f_{Z^*}(z)}\frac{\hat{\Psi}(x,y,z)-\Psi(x,y,z)}{1+\frac{\hat{f}_{Z^*}(z)-f_{Z^*}(z)}{f_{Z^*}(z)}} \\ &\qquad- \left(\frac{\Psi(x,y,z)}{f_{Z^*}(z)}\right)\frac{1}{f_{Z^*}(z)}\frac{\hat{f}_{Z^*}(z)-f_{Z^*}(z)}{1+\frac{\hat{f}_{Z^*}(z)-f_{Z^*}(z)}{f_{Z^*}(z)}},\end{aligned}$$

$\sup_{z\in\mathcal{Z}}|\hat{f}_{Z^*}(z)-f_{Z^*}(z)|^2 = O_p(\frac{\log T}{Th_T^{d_Z}} + h_T^{2m})$ by Assumptions A.1 and A.2, and Theorem 6 in Hansen (2008), and $\inf_{z\in\mathcal{Z}} f_{Z^*}(z) > 0$ by Assumption A.1(iii), the conclusion follows if

$$\sup_{x\in[0,1],y\in\mathbb{R},z\in\mathcal{Z}}|\hat{\Psi}(x,y,z)-\Psi(x,y,z)|^2 = O_p\left(\frac{\log T}{Th_T^{1+d_Z}} + h_T^{2m}\right). \tag{SM.30}$$

To show this result, as usual we distinguish the bias term and the stochastic term. The bias term is $E[\hat{\Psi}(x, y, z)] - \Psi(x, y, z) = \int_{-\infty}^{y} (E[\hat{f}_{X,Y,Z^*}(x, v, z)] - f_{X,Y,Z^*}(x, v, z))\, dv$. Since Y has compact support in $[0, 1]$ and kernel K has compact support (Assumption A.2), $\hat{f}_{X,Y,Z^*}(x, v, z) = f_{X,Y,Z^*}(x, v, z) = 0$ for $v \notin [-\varepsilon, 1+\varepsilon]$ for large T and any $\varepsilon > 0$. It follows that $\sup_{x\in[0,1], y\in\mathbb{R}, z\in\mathcal{Z}} |E[\hat{\Psi}(x, y, z)] - \Psi(x, y, z)| = O(h_T^m)$ by a standard bias expansion using Assumption A.1. Similarly, for the stochastic term

$$\sup_{x\in[0,1], y\in\mathbb{R}, z\in\mathcal{Z}} \left|\hat{\Psi}(x, y, z) - E[\hat{\Psi}(x, y, z)]\right| \leq \sup_{x\in[0,1], y\in[-\varepsilon,1+\varepsilon], z\in\mathcal{Z}} \left|\hat{\Psi}(x, y, z) - E[\hat{\Psi}(x, y, z)]\right|.$$

To show that the RHS is $O_p(\sqrt{\frac{\log T}{T h_T^{1+d_Z}}})$, we follow Hansen (2008, Proof of Theorem 1) and Bosq (1998, Proof of Theorem 2.2). Write $u := (x, y, z) \in \mathbb{R}^d$ and introduce a covering of $\Xi := [0, 1] \times [-\varepsilon, 1 + \varepsilon] \times \mathcal{Z}$ by n_T balls $B_{j,T} := \{u \in \mathbb{R}^d : |u - u_{j,T}| \leq \frac{C}{n_T}\}$, $j = 1, \dots, n_T$, where C is a constant, $u_{j,T} \in \Xi$, and n_T is such that $n_T = O(T^c)$ for some $c > 0$ and $\frac{1}{h_T^{d_Z+2} n_T} = o(\sqrt{\frac{\log T}{T h_T^{1+d_Z}}})$. Kernel K is bounded and Lipschitz (Assumption A.2), and IK is also Lipschitz. It follows that

$$|\hat{\Psi}(u) - \hat{\Psi}(u_{j,T})| \leq \frac{C_1}{n_T h_T^{d_Z+2}} \quad \text{and}$$

$$\left|E[\hat{\Psi}(u)] - E[\hat{\Psi}(u_{j,T})]\right| \leq \frac{C_1}{n_T h_T^{d_Z+2}}, \quad \text{if } u \in B_{j,T},$$

for a constant C_1 and $j = 1, \dots, n_T$. Thus

$$\sup_{u\in\Xi} \left|\hat{\Psi}(u) - E[\hat{\Psi}(u)]\right| \leq \frac{2C_1}{n_T h_T^{d_Z+2}} + \sup_{j=1,\dots,n_T} \left|\hat{\Psi}(u_{j,T}) - E[\hat{\Psi}(u_{j,T})]\right|.$$

Then

$$\text{(SM.31)} \quad P\Big[\sup_{u\in\Xi} \left|\hat{\Psi}(u) - E[\hat{\Psi}(u)]\right| \geq \eta_T\Big] \leq \sum_{j=1}^{n_T} P\left[\left|\hat{\Psi}(u_{j,T}) - E[\hat{\Psi}(u_{j,T})]\right| \geq \frac{1}{2}\eta_T\right] \leq n_T \sup_{u\in\Xi} P\left[\left|\hat{\Psi}(u) - E[\hat{\Psi}(u)]\right| \geq \frac{1}{2}\eta_T\right]$$

for $\eta_T := \eta\sqrt{\frac{\log T}{Th_T^{1+d_Z}}}$ and any constant $\eta > 0$. To bound $P[|\hat{\Psi}(u) - E[\hat{\Psi}(u)]| \geq \frac{1}{2}\eta_T]$, $u \in \Xi$, write

$$\hat{\Psi}(u) - E[\hat{\Psi}(u)] = \frac{1}{T}\sum_{t=1}^{T}\kappa_t(u),$$

where

$$\kappa_t(u) := \frac{1}{h_T^{d_Z+1}} IK\left(\frac{y - Y_t}{h_T}\right)K\left(\frac{x - X_t}{h_T}\right)K\left(\frac{z - Z_t^*}{h_T}\right) - E\left[\frac{1}{h_T^{d_Z+1}} IK\left(\frac{y - Y}{h_T}\right)K\left(\frac{x - X}{h_T}\right)K\left(\frac{z - Z^*}{h_T}\right)\right]$$

and apply Bernstein's inequality (e.g., Bosq (1998, Theorem 1.2(2))). It is possible to show that Cramer's condition is satisfied, with $E[|\kappa_t(u)|^n] \leq E[|\kappa_t(u)|^2](\frac{c_1}{h_T^{d_Z+1}})^{n-2}n!$, $n \in \mathbb{N}$, for a constant c_1. Then we get, for any $u \in \Xi$,

$$P\left[\left|\sum_{t=1}^{T}\kappa_t(u)\right| \geq \frac{1}{2}T\eta_T\right] \leq 2\exp\left(-\frac{\left(\frac{1}{2}T\eta_T\right)^2}{4\sum_{t=1}^{T}E[|\kappa_t(u)|^2] + 2\left(\frac{c_1}{h_T^{1+d_Z}}\right)\frac{1}{2}T\eta_T}\right).$$

Using $E[|\kappa_t(u)|^2] \leq c_2/h_T^{1+d_Z}$ for large T, then $P[|\sum_{t=1}^{T}\kappa_t(u)| \geq \frac{1}{2}T\eta_T] \leq 2\exp(-c_3\eta^2\log T)$ uniformly in $u \in \Xi$. From (SM.31), $P[\sup_{u\in\Xi}|\hat{\Psi}(u) - E[\hat{\Psi}(u)]| \geq \eta\sqrt{\frac{\log T}{Th_T^{1+d_Z}}}] = o(1)$ for η sufficiently large. Then (SM.30) is proved.

SM.8.2. *Proof of Lemma B.2*

The second equality follows from (SM.9). Let us now prove the first equality. We first consider the case $l < \infty$. Since the odd-order derivatives of functions φ and ϕ vanish at the boundary, we have

$$\begin{aligned}\langle\nabla^s\phi, \nabla^s\varphi\rangle &= \nabla^{s-1}\phi\nabla^s\varphi|_0^1 - \langle\nabla^{s-1}\phi, \nabla^{s+1}\varphi\rangle \\ &= -\nabla^{s-2}\phi\nabla^{s+1}\varphi|_0^1 + \langle\nabla^{s-2}\phi, \nabla^{s+2}\varphi\rangle\end{aligned} \tag{SM.32}$$

$$\begin{aligned}&\vdots\\&= (-1)^{s-1}\phi\nabla^{2s-1}\varphi|_0^1 + (-1)^s\langle\phi, \nabla^{2s}\varphi\rangle\\&= \langle\phi, (-\nabla^2)^s\varphi\rangle\end{aligned}$$

for $s = 0, 1, \ldots, l$. By summing over s, we get $\langle\phi, \varphi\rangle_H = \langle\phi, p(-\nabla^2)\varphi\rangle$. Let us now consider the case $l = \infty$. We have for any $n \in \mathbb{N}$, $\sum_{s=0}^n \frac{1}{s!}\langle\nabla^s\phi, \nabla^s\varphi\rangle = \langle\phi, \sum_{s=0}^n \frac{1}{s!}(-\nabla^2)^s\varphi\rangle$. Now take the limit $n \to \infty$: the LHS converges to $\langle\phi, \varphi\rangle_H$, while the RHS converges to $\langle\phi, p(-\nabla^2)\varphi\rangle$ from Lemma A.17(i) in Section SM.3. The conclusion follows.

SM.8.3. *Proof of Lemma B.3*

By Cauchy–Schwarz inequality, we have

$$\begin{aligned}\|\varphi\|^2 &= \|\psi\|^2 + 2\varepsilon\langle\psi, \mathcal{K}(\varphi)\rangle + \varepsilon^2\|\mathcal{K}(\varphi)\|^2\\&\leq \|\psi\|^2 + 2\varepsilon\|\psi\|\|\varphi\|^2 + \varepsilon^2\|\varphi\|^4.\end{aligned} \tag{SM.33}$$

Thus, the squared norm $\|\varphi\|^2$ satisfies the quadratic inequality

$$\varepsilon^2\|\varphi\|^4 - (1 - 2\varepsilon\|\psi\|)\|\varphi\|^2 + \|\psi\|^2 \geq 0.$$

Let $z_1 < z_2$ denote the roots of $\varepsilon^2 z^2 - (1 - 2\varepsilon\|\psi\|)z + \|\psi\|^2 = 0$:

$$z_{1,2} = \frac{1 - 2\varepsilon\|\psi\| \mp \sqrt{1 - 4\varepsilon\|\psi\|}}{2\varepsilon^2}.$$

Then either $\|\varphi\|^2 \leq z_1$ or $\|\varphi\|^2 \geq z_2$. Using $\varepsilon\|\psi\| < 1/8$ and that $\sqrt{1+a} \geq 1 + \frac{1}{2}a - a^2$ for any $a \in [-1/2, 0]$, we deduce that $z_1 \leq 8\|\psi\|^2$ and $z_2 \geq \frac{3}{8\varepsilon^2}$. Thus, either $\|\varphi\|^2 \leq 8\|\psi\|^2$ or $\|\varphi\|^2 \geq \frac{3}{8\varepsilon^2}$. Now let us assume we are in the case $\|\varphi\|^2 \leq 8\|\psi\|^2$. Then

$$\big|\|\varphi\| - \|\psi\|\big| \leq \|\varphi - \psi\| = \varepsilon\|\mathcal{K}(\varphi)\| \leq \varepsilon\|\varphi\|^2 \leq 8\varepsilon\|\psi\|^2.$$

Thus,

$$\begin{aligned}\big|\|\varphi\|^2 - \|\psi\|^2\big| &= \big|\|\varphi\| + \|\psi\|\big|\big|\|\varphi\| - \|\psi\|\big|\\&\leq (1 + \sqrt{8})8\varepsilon\|\psi\|^3 \leq 32\varepsilon\|\psi\|^3.\end{aligned}$$

SM.8.4. *Proof of Lemma B.4*

Let $\varepsilon_T = \frac{C}{\sqrt{\lambda_T}}$. Define the set

$$\Omega = \left\{\varepsilon_T\|\Delta\hat{\psi}\| \leq 1/8 \wedge \|\Delta\hat{\varphi}\|^2 \leq \frac{3}{8\varepsilon_T^2} \wedge \|\hat{\mathcal{K}}(\Delta\hat{\varphi})\| \leq \varepsilon_T\|\Delta\hat{\varphi}\|^2\right\}.$$

For large T, we have $\Omega \subset \{\|\Delta\hat{\varphi}\| < r\}$ for any $r > 0$. From Equation (A.4) and Lemma B.3, $|\|\Delta\hat{\varphi}\|^2 - \|\Delta\hat{\psi}\|^2| \leq 32\varepsilon_T\|\Delta\hat{\psi}\|^3$ if event Ω happens, while $|\|\Delta\hat{\varphi}\|^2 - \|\Delta\hat{\psi}\|^2| \leq 4c^2 + \|\Delta\hat{\psi}\|^2$ on Ω^c, where $c := \sup_{\varphi\in\Theta}\|\varphi\|$. Thus,

$$|\|\Delta\hat{\varphi}\|^2 - \|\Delta\hat{\psi}\|^2| \leq 32\varepsilon_T\|\Delta\hat{\psi}\|^3 1\{\Omega\} + (4c^2 + \|\Delta\hat{\psi}\|^2)1\{\Omega^c\}.$$

We deduce

$$\begin{aligned}
&\left|E[\|\Delta\hat{\varphi}\|^2] - E[\|\Delta\hat{\psi}\|^2]\right| \\
&\quad \leq E\big[|\|\Delta\hat{\varphi}\|^2 - \|\Delta\hat{\psi}\|^2|\big] \\
&\quad \leq 32\varepsilon_T E[\|\Delta\hat{\psi}\|^3] + 4c^2 P[\Omega^c] + E[\|\Delta\hat{\psi}\|^2 1\{\Omega^c\}].
\end{aligned}$$

From Cauchy–Schwarz inequality, $E[\|\Delta\hat{\psi}\|^2 1\{\Omega^c\}] \leq E[\|\Delta\hat{\psi}\|^4]^{1/2} P[\Omega^c]^{1/2}$. Further,

$$\begin{aligned}
P[\Omega^c] \leq{}& P\left[\|\Delta\hat{\psi}\|^2 > \frac{\lambda_T}{64C^2}\right] + P\left[\|\Delta\hat{\varphi}\|^2 > \frac{3\lambda_T}{8C^2}\right] \\
&+ P\left[\|\hat{\mathcal{K}}(\Delta\hat{\varphi})\| > \frac{C}{\sqrt{\lambda_T}}\|\Delta\hat{\varphi}\|^2\right].
\end{aligned}$$

From Lemma A.5, we get

$$\begin{aligned}
&\left|E[\|\Delta\hat{\varphi}\|^2] - E[\|\Delta\hat{\psi}\|^2]\right| \\
&\quad \leq 32C\frac{1}{\sqrt{\lambda_T}}E[\|\Delta\hat{\psi}\|^3] + 4c^2 P\left[\|\Delta\hat{\psi}\|^2 > \frac{\lambda_T}{64C^2}\right] \\
&\qquad + 4c^2 P\left[\|\Delta\hat{\varphi}\|^2 > \frac{3\lambda_T}{8C^2}\right] \\
&\qquad + E[\|\Delta\hat{\psi}\|^4]^{1/2} P\left[\|\Delta\hat{\psi}\|^2 > \frac{\lambda_T}{64C^2}\right]^{1/2} \\
&\qquad + E[\|\Delta\hat{\psi}\|^4]^{1/2} P\left[\|\Delta\hat{\varphi}\|^2 > \frac{3\lambda_T}{8C^2}\right]^{1/2} \\
&\qquad + O(T^{-\bar{b}})
\end{aligned}$$

for any $\bar{b} > 0$. Using $E[\|\Delta\hat{\psi}\|^4] \leq \frac{1}{\sqrt{\lambda_T}}E[\|\Delta\hat{\psi}\|^3]$ and $\sqrt{ab} \leq a + b$ for $a, b \geq 0$, the conclusion follows.

SM.8.5. *Proof of Lemma B.5*

To simplify the notation, we temporarily absorb the factor $\tau(1-\tau)$ in the definition of m and $\hat{m}$. We have

$$
\begin{aligned}
\text{(SM.34)}\quad & P[\|\hat{\varphi}-\varphi_0\| \geq \varepsilon_T] \\
&\leq P\Big[\inf_{\varphi\in\Theta:\|\varphi-\varphi_0\|\geq\varepsilon_T} Q_\infty(\varphi)+\lambda_T\|\varphi\|_H^2 \leq Q_\infty(\hat{\varphi})+\lambda_T\|\hat{\varphi}\|_H^2\Big] \\
&\leq P[Q_\infty(\hat{\varphi})+\lambda_T\|\hat{\varphi}\|_H^2-\lambda_T\|\varphi_0\|_H^2 \geq C(\varepsilon_T,\lambda_T)].
\end{aligned}
$$

Let us now bound $Q_\infty(\hat{\varphi})+\lambda_T\|\hat{\varphi}\|_H^2$ in probability. Since

$$
Q_\infty(\hat{\varphi}) = \frac{1}{T}\sum_{t=1}^T m(\hat{\varphi},Z_t)^2 - \left(\frac{1}{T}\sum_{t=1}^T m(\hat{\varphi},Z_t)^2 - Q_\infty(\hat{\varphi})\right)
$$

and $\hat{\varphi}\in\Theta$, we have, for any $\eta>0$,

$$
\begin{aligned}
& P[Q_\infty(\hat{\varphi})+\lambda_T\|\hat{\varphi}\|_H^2 \geq \eta+\lambda_T\|\varphi_0\|_H^2] \\
&\leq P\left[\frac{1}{T}\sum_{t=1}^T m(\hat{\varphi},Z_t)^2+\lambda_T\|\hat{\varphi}\|_H^2 \geq \eta/2+\lambda_T\|\varphi_0\|_H^2\right] \\
&\quad + P\left[\sup_{\varphi\in\Theta}\left|\frac{1}{T}\sum_{t=1}^T m(\varphi,Z_t)^2 - Q_\infty(\varphi)\right| \geq \eta/2\right] \\
&=: P_1+P_2.
\end{aligned}
$$

To bound probability P_1, we denote $\langle m_1,m_2\rangle_T := \frac{1}{T}\sum_{t=1}^T m_1(Z_t)m_2(Z_t)$ and $\|m\|_T := \langle m,m\rangle_T^{1/2}$. Then $\frac{1}{T}\sum_{t=1}^T m(\hat{\varphi},Z_t)^2 = \|m(\hat{\varphi},\cdot)\|_T^2$ and $Q_T(\hat{\varphi}) = \|\hat{m}(\hat{\varphi},\cdot)\|_T^2$. From the condition $Q_T(\hat{\varphi})+\lambda_T\|\hat{\varphi}\|_H^2 \leq Q_T(\varphi_0)+\lambda_T\|\varphi_0\|_H^2$, we deduce $\|m(\hat{\varphi},\cdot)\|_T^2+\lambda_T\|\hat{\varphi}\|_H^2+2\langle m(\hat{\varphi},\cdot),\ \Delta\hat{m}(\hat{\varphi},\cdot)\rangle_T+\|\Delta\hat{m}(\hat{\varphi},\cdot)\|_T^2 - \|\Delta\hat{m}(\varphi_0,\cdot)\|_T^2-\lambda_T\|\varphi_0\|_H^2 \leq 0$. Using the Cauchy–Schwarz inequality, we deduce that $\delta_T := \sqrt{\|m(\hat{\varphi},\cdot)\|_T^2+\lambda_T\|\hat{\varphi}\|_H^2}$ satisfies $\delta_T^2-2d_{1,T}\delta_T+d_{2,T}\leq 0$, where $d_{1,T} := \|\Delta\hat{m}(\hat{\varphi},\cdot)\|_T$ and $d_{2,T} := \|\Delta\hat{m}(\hat{\varphi},\cdot)\|_T^2-\|\Delta\hat{m}(\varphi_0,\cdot)\|_T^2-\lambda_T\|\varphi_0\|_H^2$. Then

$$
\begin{aligned}
\delta_T &\leq d_{1,T}+\sqrt{d_{1,T}^2-d_{2,T}} \\
&\leq \sqrt{\sup_{\varphi\in\Theta}\frac{1}{T}\sum_{t=1}^T \Delta\hat{m}(\varphi,Z_t)^2} + \sqrt{\sup_{\varphi\in\Theta}\frac{1}{T}\sum_{t=1}^T \Delta\hat{m}(\varphi,Z_t)^2+\lambda_T\|\varphi_0\|_H^2} \\
&\leq D_T+\sqrt{D_T^2+\lambda_T\|\varphi_0\|_H^2},
\end{aligned}
$$

where $D_T := \sup_{\varphi\in\Theta}\sup_{z\in\mathcal{Z}}|\Delta\hat{m}(\varphi, z)|$. Furthermore,

$$\begin{aligned}\delta_T^2 &\le 2D_T^2 + 2D_T\sqrt{D_T^2 + \lambda_T\|\varphi_0\|_H^2} + \lambda_T\|\varphi_0\|_H^2 \\ &\le 4D_T^2 + 2D_T\|\varphi_0\|_H\sqrt{\lambda_T} + \lambda_T\|\varphi_0\|_H^2.\end{aligned}$$

We get

$$P_1 = P[\delta_T^2 \ge \eta/2 + \lambda_T\|\varphi_0\|_H^2] \le P[2D_T^2 + D_T\|\varphi_0\|_H\sqrt{\lambda_T} \ge \eta/4].$$

Since the parabola $q(x) = 2x^2 + x\|\varphi_0\|_H\sqrt{\lambda_T} - \eta/4$ has roots

$$x_{1,2} = \frac{-\|\varphi_0\|_H\sqrt{\lambda_T} \pm \sqrt{\lambda_T\|\varphi_0\|_H^2 + 2\eta}}{4},$$

we deduce that

$$P_1 \le P\left[D_T \ge \frac{\sqrt{\lambda_T\|\varphi_0\|_H^2 + 2\eta} - \sqrt{\lambda_T\|\varphi_0\|_H^2}}{4}\right].$$

Thus,

$$\begin{aligned}&P[Q_\infty(\hat{\varphi}) + \lambda_T\|\hat{\varphi}\|_H^2 \ge \eta + \lambda_T\|\varphi_0\|_H^2] \\ &\quad\le P\left[D_T \ge \frac{\sqrt{\lambda_T\|\varphi_0\|_H^2 + 2\eta} - \sqrt{\lambda_T\|\varphi_0\|_H^2}}{4}\right] \\ &\qquad + P\left[\sup_{\varphi\in\Theta}\left|\frac{1}{T}\sum_{t=1}^{T} m(\varphi, Z_t)^2 - Q_\infty(\varphi)\right| \ge \eta/2\right].\end{aligned} \tag{SM.35}$$

From (SM.34) and (SM.35), the conclusion follows.

SM.8.6. *Proof of Lemma B.6*

Let $\varphi_{\lambda,A} := (\lambda + A^*A)^{-1}A^*A\varphi_0$ be the Tikhonov solution of the linearized problem $A\varphi = r$. Below we show in part (a) that $\|\varphi_\lambda - \varphi_0\|_H = O(\lambda^\delta)$ under Assumption 4(i) and (ii) before showing in part (b) that the statement of the lemma holds.

Part (a)

We follow the idea of the proof of Theorem 10.7 in Engl, Hanke, and Neubauer (2000) that consists in comparing φ_λ and $\varphi_{\lambda,A}$. By Lemma A.16 in Section SM.3 we have

$$\mathcal{A}(\varphi_\lambda) - \tau = A(\varphi_\lambda - \varphi_0) + R(\varphi_\lambda, \varphi_0) \tag{SM.36}$$

and

$$\mathcal{A}(\varphi_{\lambda,A}) - \tau = A(\varphi_{\lambda,A} - \varphi_0) + R(\varphi_{\lambda,A}, \varphi_0), \tag{SM.37}$$

where

$$\begin{aligned} &\|R(\varphi_\lambda, \varphi_0)\|_{L^2(F_Z)} \le \frac{1}{2}c\|\varphi_\lambda - \varphi_0\|^2, \\ &\|R(\varphi_{\lambda,A}, \varphi_0)\|_{L^2(F_Z)} \le \frac{1}{2}c\|\varphi_{\lambda,A} - \varphi_0\|^2. \end{aligned} \tag{SM.38}$$

From the definition of φ_λ, we have

$$Q_\infty(\varphi_\lambda) + \lambda\|\varphi_\lambda\|_H^2 \le Q_\infty(\varphi_{\lambda,A}) + \lambda\|\varphi_{\lambda,A}\|_H^2.$$

We deduce

$$\begin{aligned} \|\varphi_\lambda - \varphi_0\|_H^2 = \|\varphi_\lambda\|_H^2 + \|\varphi_\lambda - \varphi_0\|_H^2 - \|\varphi_\lambda\|_H^2 \\ \le \frac{1}{\lambda}(Q_\infty(\varphi_{\lambda,A}) - Q_\infty(\varphi_\lambda)) \\ + \|\varphi_{\lambda,A}\|_H^2 + \|\varphi_\lambda - \varphi_0\|_H^2 - \|\varphi_\lambda\|_H^2 \\ = \frac{1}{\lambda}\big(\|\mathcal{A}(\varphi_{\lambda,A}) - \tau\|_{L^2(F_Z,\tau)}^2 - \|\mathcal{A}(\varphi_\lambda) - \tau\|_{L^2(F_Z,\tau)}^2\big) \\ + \|\varphi_{\lambda,A} - \varphi_0\|_H^2 + 2\langle\varphi_{\lambda,A} - \varphi_0, \varphi_0\rangle_H \\ - 2\langle\varphi_\lambda - \varphi_0, \varphi_0\rangle_H. \end{aligned} \tag{SM.39}$$

From (SM.37), we have

$$\begin{aligned} &\|\mathcal{A}(\varphi_{\lambda,A}) - \tau\|_{L^2(F_Z,\tau)}^2 \\ &\quad = \|A(\varphi_{\lambda,A} - \varphi_0)\|_{L^2(F_Z,\tau)}^2 + \|R(\varphi_{\lambda,A}, \varphi_0)\|_{L^2(F_Z,\tau)}^2 \\ &\qquad + 2\langle A(\varphi_{\lambda,A} - \varphi_0), R(\varphi_{\lambda,A}, \varphi_0)\rangle_{L^2(F_Z,\tau)}. \end{aligned} \tag{SM.40}$$

Then we get

$$\begin{aligned} \|\varphi_\lambda - \varphi_0\|_H^2 \le \frac{1}{\lambda}\|A(\varphi_{\lambda,A} - \varphi_0)\|_{L^2(F_Z,\tau)}^2 \\ + \frac{1}{\lambda}\|R(\varphi_{\lambda,A}, \varphi_0)\|_{L^2(F_Z,\tau)}^2 \\ + \frac{2}{\lambda}\langle A(\varphi_{\lambda,A} - \varphi_0), R(\varphi_{\lambda,A}, \varphi_0)\rangle_{L^2(F_Z,\tau)} \end{aligned} \tag{SM.41}$$

$$
\begin{aligned}
&- \frac{1}{\lambda} \|\mathcal{A}(\varphi_\lambda) - \tau\|^2_{L^2(F_Z,\tau)} \\
&+ \|\varphi_{\lambda,A} - \varphi_0\|^2_H + 2\langle \varphi_{\lambda,A} - \varphi_0, \varphi_0 \rangle_H \\
&- 2\langle \varphi_\lambda - \varphi_0, \varphi_0 \rangle_H .
\end{aligned}
$$

Now, from Assumption 4(ii) there exists $\psi_0 \in L^2(F_Z,\tau)$ such that $\varphi_0 = A^*\psi_0$ and $\|\psi_0\|_{L^2(F_Z,\tau)} < 1/c$. Indeed, $\psi_0 = \sum_{j=1}^\infty \frac{\langle \phi_j, \varphi_0\rangle_H}{\omega_j}\psi_j$, where $\{\phi_j, \psi_j, \omega_j : j = 1, 2, \ldots\}$ is called the singular system of operator A with $\omega_j = \sqrt{\nu_j}$ (Kress (1999, p. 278)), and $\|\psi_0\|^2_{L^2(F_Z,\tau)} = \sum_{j=1}^\infty \frac{\langle \phi_j, \varphi_0\rangle_H^2}{\nu_j}$. Then, from (SM.36), we have

$$
\begin{aligned}
\text{(SM.42)}\quad &\|\mathcal{A}(\varphi_\lambda) - \tau\|^2_{L^2(F_Z,\tau)} \\
&= \|\mathcal{A}(\varphi_\lambda) - \tau + \lambda\psi_0\|^2_{L^2(F_Z,\tau)} - \lambda^2 \|\psi_0\|^2_{L^2(F_Z,\tau)} \\
&\quad - 2\lambda\langle \mathcal{A}(\varphi_\lambda) - \tau, \psi_0\rangle_{L^2(F_Z,\tau)} \\
&= \|\mathcal{A}(\varphi_\lambda) - \tau + \lambda\psi_0\|^2_{L^2(F_Z,\tau)} - \lambda^2 \|\psi_0\|^2_{L^2(F_Z,\tau)} \\
&\quad - 2\lambda\langle \varphi_\lambda - \varphi_0, \varphi_0\rangle_H - 2\lambda\langle R(\varphi_\lambda, \varphi_0), \psi_0\rangle_{L^2(F_Z,\tau)}.
\end{aligned}
$$

By replacing (SM.42) into (SM.41) and using $\langle \varphi_{\lambda,A} - \varphi_0, \varphi_0\rangle_H = \langle A(\varphi_{\lambda,A} - \varphi_0), \psi_0\rangle_{L^2(F_Z,\tau)}$, we get

$$
\begin{aligned}
\text{(SM.43)}\quad \|\varphi_\lambda - \varphi_0\|^2_H \le\ &\frac{1}{\lambda}\|A(\varphi_{\lambda,A} - \varphi_0)\|^2_{L^2(F_Z,\tau)} \\
&+ \frac{1}{\lambda}\|R(\varphi_{\lambda,A}, \varphi_0)\|^2_{L^2(F_Z,\tau)} \\
&+ \frac{2}{\lambda}\langle A(\varphi_{\lambda,A} - \varphi_0), R(\varphi_{\lambda,A}, \varphi_0)\rangle_{L^2(F_Z,\tau)} \\
&- \frac{1}{\lambda}\|\mathcal{A}(\varphi_\lambda) - \tau + \lambda\psi_0\|^2_{L^2(F_Z,\tau)} \\
&+ 2\langle R(\varphi_\lambda, \varphi_0), \psi_0\rangle_{L^2(F_Z,\tau)} \\
&+ \langle 2A(\varphi_{\lambda,A} - \varphi_0) + \lambda\psi_0, \psi_0\rangle_{L^2(F_Z,\tau)} + \|\varphi_{\lambda,A} - \varphi_0\|^2_H.
\end{aligned}
$$

From (SM.38), we have the inequalities

$$
\begin{aligned}
&2\langle R(\varphi_\lambda, \varphi_0), \psi_0\rangle_{L^2(F_Z,\tau)} \le c\|\psi_0\|_{L^2(F_Z,\tau)}\|\varphi_\lambda - \varphi_0\|^2, \\
&\frac{1}{\lambda}\|R(\varphi_{\lambda,A}, \varphi_0)\|^2_{L^2(F_Z,\tau)} \le \frac{1}{4\lambda}c^2\|\varphi_{\lambda,A} - \varphi_0\|^4.
\end{aligned}
$$

Moreover, by using (SM.38), $\varphi_{\lambda,A} - \varphi_0 = -\lambda(\lambda + A^*A)^{-1}\varphi_0$, and $(\lambda + A^*A)^{-1}\varphi_0 = A^*(\lambda + AA^*)^{-1}\psi_0$, we have

$$\begin{aligned}
&\frac{2}{\lambda}\langle A(\varphi_{\lambda,A} - \varphi_0), R(\varphi_{\lambda,A}, \varphi_0)\rangle_{L^2(F_Z,\tau)} \\
&\quad \leq \frac{1}{\lambda} c \|A(\varphi_{\lambda,A} - \varphi_0)\|_{L^2(F_Z,\tau)} \|\varphi_{\lambda,A} - \varphi_0\|^2 \\
&\quad \leq c\|\psi_0\|_{L^2(F_Z,\tau)} \|\varphi_{\lambda,A} - \varphi_0\|^2
\end{aligned}$$

and

$$\begin{aligned}
&\frac{1}{\lambda}\|A(\varphi_{\lambda,A} - \varphi_0)\|^2_{L^2(F_Z,\tau)} + \langle 2A(\varphi_{\lambda,A} - \varphi_0) + \lambda\psi_0, \psi_0\rangle_{L^2(F_Z,\tau)} \\
&\quad = \lambda^3 \|(\lambda + AA^*)^{-1}\psi_0\|^2_{L^2(F_Z,\tau)}.
\end{aligned}$$

Thus, from (SM.43), we deduce

$$\begin{aligned}
\|\varphi_\lambda - \varphi_0\|^2_H \leq{}& c\|\psi_0\|_{L^2(F_Z,\tau)} \|\varphi_\lambda - \varphi_0\|^2 \\
&+ \big(1 + c\|\psi_0\|_{L^2(F_Z,\tau)}\big) \|\varphi_{\lambda,A} - \varphi_0\|^2_H \\
&+ \frac{1}{4\lambda} c^2 \|\varphi_{\lambda,A} - \varphi_0\|^4 + \lambda^3 \|(\lambda + AA^*)^{-1}\psi_0\|^2_{L^2(F_Z,\tau)}.
\end{aligned}$$

Then, since $c\|\psi_0\|_{L^2(F_Z,\tau)} < 1$, we get

$$\begin{aligned}
\|\varphi_\lambda - \varphi_0\|^2_H = O\bigg(&\|\varphi_{\lambda,A} - \varphi_0\|^2_H + \frac{1}{\lambda}\|\varphi_{\lambda,A} - \varphi_0\|^4 \\
&+ \lambda^3 \|(\lambda + AA^*)^{-1}\psi_0\|^2_{L^2(F_Z,\tau)}\bigg).
\end{aligned}$$

Now, from Assumption 4(i),

$$\|\varphi_{\lambda,A} - \varphi_0\|^2_H = \lambda^2 \sum_{j=1}^{\infty} \frac{\langle \phi_j, \varphi_0\rangle^2_H}{(\lambda + \nu_j)^2} = O(\lambda^{2\delta}),$$

$$\|(\lambda + AA^*)^{-1}\psi_0\|^2_{L^2(F_Z,\tau)} = \lambda^2 \sum_{j=1}^{\infty} \frac{\langle \phi_j, \varphi_0\rangle^2_H}{(\lambda + \nu_j)^2 \nu_j} = O(\lambda^{2\delta - 1}),$$

and we conclude that $\|\varphi_\lambda - \varphi_0\|_H = O(\lambda^\delta)$.

Part (b)

From part (a) and Assumption 4(i), it follows that $\|\varphi_\lambda - \varphi_0\| = o(\sqrt{\lambda})$ and $\|\varphi_{\lambda,A} - \varphi_0\| = o(\sqrt{\lambda})$. For $\eta := d^2 - c > 0$, let λ be such that $\|\varphi_\lambda - \varphi_{\lambda,A}\| \leq$

$\varepsilon\eta/2$ and $\|\varphi_\lambda - \varphi_0\| \leq \varepsilon\eta/2$, where $\varepsilon := d\sqrt{\lambda}$. Then, by Assumption 2, we have

$$\inf_{\substack{\varphi\in\Theta:\\ \|\varphi-\varphi_0\|\geq d\sqrt{\lambda}}} Q_\infty(\varphi) + \lambda\|\varphi\|_H^2 = \inf_{\substack{\varphi\in\Theta:\\ \|\varphi-\varphi_0\|\geq \varepsilon}} Q_\infty(\varphi) + \lambda\|\varphi\|_H^2$$
$$\geq \inf_{\substack{\varphi\in\Theta:\\ \|\varphi-\varphi_\lambda\|\geq \varepsilon(1-\eta/2)}} Q_\infty(\varphi) + \lambda\|\varphi\|_H^2$$
$$= \inf_{\substack{\varphi\in\Theta:\\ \|\varphi-\varphi_\lambda\|= \varepsilon(1-\eta/2)}} Q_\infty(\varphi) + \lambda\|\varphi\|_H^2.$$

The last equality comes from the fact that the minimum is taken on the boundary of set $\{\varphi \in \Theta : \|\varphi - \varphi_\lambda\| \geq \varepsilon(1-\eta/2)\}$ (otherwise the function $Q_\infty(\varphi) + \lambda\|\varphi\|_H^2$ admits a local minimum). Moreover, $\|\varphi - \varphi_\lambda\| = \varepsilon(1-\eta/2)$ implies $\varepsilon \geq \|\varphi - \varphi_{\lambda,A}\| \geq \varepsilon(1-\eta)$. Thus

$$\inf_{\substack{\varphi\in\Theta:\\ \|\varphi-\varphi_\lambda\|= \varepsilon(1-\eta/2)}} Q_\infty(\varphi) + \lambda\|\varphi\|_H^2 \geq \inf_{\substack{\varphi\in\Theta:\\ \varepsilon\geq\|\varphi-\varphi_{\lambda,A}\|\geq \varepsilon(1-\eta)}} Q_\infty(\varphi) + \lambda\|\varphi\|_H^2.$$

We get

$$\inf_{\substack{\varphi\in\Theta:\\ \|\varphi-\varphi_0\|\geq d\sqrt{\lambda}}} Q_\infty(\varphi) + \lambda\|\varphi\|_H^2 \geq \inf_{\substack{\varphi\in\Theta:\\ d_2\sqrt{\lambda}\geq\|\varphi-\varphi_{\lambda,A}\|\geq d_1\sqrt{\lambda}}} Q_\infty(\varphi) + \lambda\|\varphi\|_H^2,$$

where $d_1 = d(1-\eta)$ and $d_2 = d$.

Now, we prove that for $c < d_1^2$,

$$\inf_{\substack{\varphi\in\Theta:\\ d_2\sqrt{\lambda}\geq\|\varphi-\varphi_{\lambda,A}\|\geq d_1\sqrt{\lambda}}} Q_\infty(\varphi) + \lambda\|\varphi\|_H^2 \geq \lambda\|\varphi_0\|_H^2 + c\lambda\Gamma(\lambda) \tag{SM.44}$$

for small $\lambda > 0$. Since $\eta > 0$ can be chosen arbitrarily small, the conclusion follows.

The proof is by contradiction. Suppose that (SM.44) is not true for $c < d_1^2$. Then there exist sequences (λ_n), $\lambda_n \to 0$, and $(\varphi_n) \subset \Theta$ such that $d_1\sqrt{\lambda_n} \leq \|\varphi_n - \varphi_{\lambda_n,A}\| \leq d_2\sqrt{\lambda_n}$ and

$$Q_\infty(\varphi_n) + \lambda_n\|\varphi_n\|_H^2 \leq \lambda_n\|\varphi_0\|_H^2 + c_1\lambda_n\Gamma(\lambda_n)$$

for $c < c_1 < d_1^2$. By using the linearization $\mathcal{A}(\varphi) - \tau = A\Delta\varphi + R(\varphi, \varphi_0)$, Assumption A.4(ii), and Lemma A.16, we get

$$Q_\infty(\varphi_n) = \langle\Delta\varphi_n, A^*A\Delta\varphi_n\rangle_H + 2\langle A\Delta\varphi_n, R(\varphi_n,\varphi_0)\rangle_{L^2(F_Z,\tau)}$$
$$+ \|R(\varphi_n,\varphi_0)\|_{L^2(F_Z,\tau)}^2$$
$$\geq \langle\Delta\varphi_n, A^*A\Delta\varphi_n\rangle_H - 2C\|A\|\|\Delta\varphi_n\|^3.$$

Further, $\|\Delta\varphi_n\| \leq \|\varphi_n - \varphi_{\lambda_n,A}\| + \|\varphi_{\lambda_n,A} - \varphi_0\| = O(\sqrt{\lambda_n})$ and we get

$$\langle \Delta\varphi_n, A^*A\Delta\varphi_n\rangle_H + \lambda_n\|\varphi_n\|_H^2 \leq \lambda_n\|\varphi_0\|_H^2 + c_1\lambda_n\Gamma(\lambda_n) + O\big(\lambda_n^{3/2}\big).$$

By Assumption 4(iii), we deduce for large n that

$$\langle \Delta\varphi_n, A^*A\Delta\varphi_n\rangle_H + \lambda_n\|\varphi_n\|_H^2 \leq \lambda_n\|\varphi_0\|_H^2 + c_2\lambda_n\Gamma(\lambda_n) \tag{SM.45}$$

for $c_1 < c_2 < d_1^2$. Let us now show that the latter inequality cannot hold. For this purpose, we derive an explicit expression for

$$\tilde{C}(\varepsilon, \lambda) := \inf_{\varphi:\|\varphi-\varphi_{\lambda,A}\|\geq\varepsilon} \langle \Delta\varphi, A^*A\Delta\varphi\rangle_H + \lambda\|\varphi\|_H^2.$$

From GS, proof of Proposition 2, the penalized linearized criterion can be rewritten as

$$\begin{aligned}
&\langle \Delta\varphi, A^*A\Delta\varphi\rangle_H + \lambda\|\varphi\|_H^2 \\
&\quad = \langle \varphi - \varphi_{\lambda,A}, (A^*A+\lambda)(\varphi - \varphi_{\lambda,A})\rangle_H + \lambda\langle\varphi_{\lambda,A}, \varphi_0\rangle_H.
\end{aligned}$$

Thus

$$\begin{aligned}
\tilde{C}(\varepsilon, \lambda) &= \inf_{\varphi:\|\varphi-\varphi_{\lambda,A}\|\geq\varepsilon} \langle \varphi - \varphi_{\lambda,A}, (A^*A+\lambda)(\varphi - \varphi_{\lambda,A})\rangle_H \\
&\quad + \lambda\langle\varphi_{\lambda,A}, \varphi_0\rangle_H \\
&= \inf_{\psi:\|\psi\|\geq\varepsilon} \langle\psi, (A^*A+\lambda)\psi\rangle_H + \lambda\langle\varphi_{\lambda,A}, \varphi_0\rangle_H \\
&= \varepsilon^2 \inf_{\psi:\|\psi\|=1} \langle\psi, (A^*A+\lambda)\psi\rangle_H + \lambda\langle\varphi_{\lambda,A}, \varphi_0\rangle_H \\
&= \varepsilon^2\Gamma(\lambda) + \lambda\langle\varphi_{\lambda,A}, \varphi_0\rangle_H.
\end{aligned}$$

Moreover, $\langle\varphi_{\lambda,A}, \varphi_0\rangle_H = \|\varphi_0\|_H^2 + \langle\varphi_{\lambda,A} - \varphi_0, \varphi_0\rangle_H$. Thus, we get $\tilde{C}(\varepsilon, \lambda) = \varepsilon^2\Gamma(\lambda) + \lambda\|\varphi_0\|_H^2 + \lambda\langle\varphi_{\lambda,A} - \varphi_0, \varphi_0\rangle_H$. From Assumption 4(i),

$$\tilde{C}(\varepsilon, \lambda) = \varepsilon^2\Gamma(\lambda) + \lambda\|\varphi_0\|_H^2 + O\big(\lambda^{3/2}\big).$$

Now

$$\begin{aligned}
&\langle \Delta\varphi_n, A^*A\Delta\varphi_n\rangle_H + \lambda_n\|\varphi_n\|_H^2 \\
&\quad \geq \tilde{C}(d_1\sqrt{\lambda_n}, \lambda_n) = d_1^2\lambda_n\Gamma(\lambda_n) + \lambda_n\|\varphi_0\|_H^2 + O\big(\lambda_n^{3/2}\big).
\end{aligned}$$

By Assumption 4(iii) and $c_2 < d_1^2$, this is incompatible with (SM.45).

SM.8.7. *Proof of Lemma B.7*

SM.8.7.1. *Proof of* $P[\|\Delta\hat{\varphi}\|^2 \geq d^2\lambda_T] = O(T^{-\bar{b}})$

Let $T^* = T$ to ease notation. We have to bound the two probabilities in the RHS of inequality (SM.18). Let us consider the first one. From the proofs of Lemmas A.3 and B.1, we know that

$$\sup_{\varphi\in\Theta}\sup_{z\in\mathcal{Z}}|\Delta\hat{m}(\varphi,z)| \leq \sup_{x\in[0,1],y\in\mathbb{R},z\in\mathcal{Z}}\left|\hat{f}_{X|Z}(x|z)\hat{F}_{Y|X,Z}(y|x,z) - f_{X|Z}(x|z)F_{Y|X,Z}(y|x,z)\right|$$

and

$$\hat{f}_{X|Z}(x|z)\hat{F}_{Y|X,Z}(y|x,z) - f_{X|Z}(x|z)F_{Y|X,Z}(y|x,z) = \frac{1}{f_{Z^*}(z)}\frac{\hat{\Psi}(x,y,z)-\Psi(x,y,z)}{1+\dfrac{\hat{f}_{Z^*}(z)-f_{Z^*}(z)}{f_{Z^*}(z)}} - \left(\frac{\Psi(x,y,z)}{f_{Z^*}(z)}\right)\frac{1}{f_{Z^*}(z)}\frac{\hat{f}_{Z^*}(z)-f_{Z^*}(z)}{1+\dfrac{\hat{f}_{Z^*}(z)-f_{Z^*}(z)}{f_{Z^*}(z)}},$$

where

$$\hat{\Psi}(x,y,z) := \int_{-\infty}^{y}\hat{f}_{X,Y,Z^*}(x,v,z)\,dv = \frac{1}{Th_T^{1+d_Z}}\sum_{t=1}^{T} IK\left(\frac{y-Y_t}{h_T}\right)K\left(\frac{x-X_t}{h_T}\right)K\left(\frac{z-Z_t^*}{h_T}\right)$$

and $\Psi(x,y,z) := \int_{-\infty}^{y} f_{X,Y,Z^*}(x,v,z)\,dv$. Then using that $\frac{\Psi(x,y,z)}{f_{Z^*}(z)}$ is bounded (Assumptions A.1(i) and A.4(ii)) and $\inf_{z\in\mathcal{Z}} f_{Z^*}(z) > 0$ (see Assumption A.1(iii)), for $a_T := \sqrt{c\lambda_T\Gamma(\lambda_T)^2}$ we get

$$P\Big[\sup_{\varphi\in\Theta}\sup_{z\in\mathcal{Z}}|\Delta\hat{m}(\varphi,z)| \geq a_T\Big] \leq P\left[\sup_{x\in[0,1],y\in\mathbb{R},z\in\mathcal{Z}}\left|\frac{\hat{\Psi}(x,y,z)-\Psi(x,y,z)}{1+\dfrac{\hat{f}_{Z^*}(z)-f_{Z^*}(z)}{f_{Z^*}(z)}}\right| \geq \frac{a_T}{\log T}\right] \tag{SM.46}$$

$$
\begin{aligned}
&+ P\left[\sup_{x\in[0,1],y\in\mathbb{R},z\in\mathcal{Z}}\left|\frac{\hat{f}_{Z^*}(z)-f_{Z^*}(z)}{1+\dfrac{\hat{f}_{Z^*}(z)-f_{Z^*}(z)}{f_{Z^*}(z)}}\right| \geq \frac{a_T}{\log T}\right] \\
&\leq P\left[\sup_{x\in[0,1],y\in\mathbb{R},z\in\mathcal{Z}}|\hat{\Psi}(x,y,z)-\Psi(x,y,z)| \geq \frac{a_T}{2\log T}\right] \\
&\quad + P\left[\sup_{z\in\mathcal{Z}}|\hat{f}_{Z^*}(z)-f_{Z^*}(z)| \geq \frac{a_T}{2\log T}\right] \\
&\quad + 2P\left[\sup_{z\in\mathcal{Z}}\left|\frac{\hat{f}_{Z^*}(z)-f_{Z^*}(z)}{f_{Z^*}(z)}\right| \geq \frac{1}{2}\right].
\end{aligned}
$$

To bound the first probability on the RHS, as in the proof of Lemma B.1, we use that $\hat{\Psi}(x,y,z)-\Psi(x,y,z)=0$ for y outside a compact interval $[-\bar{\varepsilon},1+\bar{\varepsilon}]$, because of the assumptions of compact support for the kernel (Assumption A.2) and for Y. Further, using that $E[\hat{\Psi}(x,y,z)]-\Psi(x,y,z)=O(h_T^m)$ uniformly in z, and $h_T^m = o(\frac{a_T}{\log T})$ from Assumption 4(i) and $\gamma < \frac{2m\eta}{1+2a}$, we have

$$
\begin{aligned}
&P\left[\sup_{x\in[0,1],y\in\mathbb{R},z\in\mathcal{Z}}|\hat{\Psi}(x,y,z)-\Psi(x,y,z)| \geq \frac{a_T}{2\log T}\right] \\
&\quad \leq P\left[\sup_{x\in[0,1],y\in[-\bar{\varepsilon},1+\bar{\varepsilon}],z\in\mathcal{Z}}\left|\hat{\Psi}(x,y,z)-E[\hat{\Psi}(x,y,z)]\right| \geq \frac{a_T}{4\log T}\right].
\end{aligned}
$$

Then, using a covering argument and Bernstein's inequality as in the proof of Lemma B.1, we have

$$
\begin{aligned}
&P\left[\sup_{x\in[0,1],y\in[-\bar{\varepsilon},1+\bar{\varepsilon}],z\in\mathcal{Z}}\left|\hat{\Psi}(x,y,z)-E[\hat{\Psi}(x,y,z)]\right| \geq \frac{a_T}{4\log T}\right] \\
&\quad \leq c_1 n_T \exp\left(-c_2\frac{T h_T^{d_Z+1}a_T^2}{(\log T)^2}\right)
\end{aligned}
$$

for some constants c_1 and c_2, where n_T is such that $n_T = O(T^c)$ for some $c>0$ and $\frac{1}{h_T^{d_Z} n_T} = o(\frac{a_T}{\log T})$. Then, $P[\sup_{x\in[0,1],y\in\mathbb{R},z\in\mathcal{Z}}|\hat{\Psi}(x,y,z)-\Psi(x,y,z)| \geq \frac{a_T}{2\log T}] = O(T^{-\bar{b}})$ for any $\bar{b}>0$, since $\frac{T h_T^{d_Z+1}a_T^2}{(\log T)^2}$ diverges as a power of T. The second probability in the RHS of (SM.46) is bounded in a similar vein, and the third probability is dominated by the second.

Finally, let us consider the second probability in the RHS of inequality (SM.18). From the proof of Lemma A.3(ii), we know that

$$\sup_{\varphi\in\Theta}\left|\frac{1}{T\tau(1-\tau)}\sum_{t=1}^{T} m(\varphi, Z_t)^2 - Q_\infty(\varphi)\right| \leq \sup_{\varrho\in[0,1]^4}\left|\frac{1}{T\tau(1-\tau)}\sum_{t=1}^{T}\big(a(Z_t,\varrho) - E[a(Z,\varrho)]\big)\right|,$$

where $a(z,\varrho) := f_{X|Z}(x|z)f_{X|Z}(\xi|z)[F_{U|X,Z}(u|x,z) - \tau][F_{U|X,Z}(v|\xi,z) - \tau]$, $\varrho := (x,\xi,u,v) \in [0,1]^4$. Then, for $a_T := c\tau(1-\tau)\lambda_T \Gamma(\lambda_T)$,

$$\text{(SM.47)}\quad P\left[\sup_{\varphi\in\Theta}\left|\frac{1}{T\tau(1-\tau)}\sum_{t=1}^{T} m(\varphi, Z_t)^2 - Q_\infty(\varphi)\right| \geq c\frac{\lambda_T}{\log(1/\lambda_T)}\right] \leq P\left[\sup_{\varrho\in[0,1]^4}\left|\frac{1}{T}\sum_{t=1}^{T}\big(a(Z_t,\varrho) - E[a(Z,\varrho)]\big)\right| \geq a_T\right].$$

Let us bound the probability in the RHS of (SM.47). Let us introduce a covering of $[0,1]^4$ with n_T balls $B_{j,T} := \{\varrho \in \mathbb{R}^4 : |\varrho - \varrho_{j,T}| \leq \frac{C}{n_T}\}$, $j = 1,\dots,n_T$, where C is a constant, $\varrho_{j,T} \in [0,1]^4$, and n_T is such that $n_T = O(T^{\bar{c}})$ for some $\bar{c} > 0$ and $\frac{1}{n_T} = o(\frac{\lambda_T}{\log(1/\lambda_T)})$. Since function $a(z,\varrho)$ is Lipschitz w.r.t. $\varrho \in [0,1]^4$ (Assumption A.3), we have, for $\hat{\Phi}(\varrho) := \frac{1}{T}\sum_{t=1}^{T} a(Z_t,\varrho)$, that $|\hat{\Phi}(\varrho) - \hat{\Phi}(\varrho_{j,T})| \leq \frac{C_1}{n_T}$ and $|E[\hat{\Phi}(\varrho)] - E[\hat{\Phi}(\varrho_{j,T})]| \leq \frac{C_1}{n_T}$ if $\varrho \in B_{j,T}$, for a constant C_1 and $j = 1,\dots,n_T$. Then

$$P\left[\sup_{\varrho\in[0,1]^4}\left|\frac{1}{T}\sum_{t=1}^{T}\big(a(Z_t,\varrho) - E[a(Z,\varrho)]\big)\right| \geq a_T\right] \leq n_T \sup_{\varrho\in[0,1]^4} P\left[\left|\frac{1}{T}\sum_{t=1}^{T}\big(a(Z_t,\varrho) - E[a(Z,\varrho)]\big)\right| \geq \frac{1}{2}a_T\right].$$

To bound the latter probability, by noting that function a is bounded by $\sup_{z,\varrho}|a(z,\varrho)| \leq 4\sup_{x,z} f_{X|Z}(x|z)^2 =: C_2$, we get from Hoeffding's inequality (e.g., Bosq (1998, Theorem 1.2(1))) that

$$P\left[\left|\frac{1}{T}\sum_{t=1}^{T}\big(a(Z_t,\varrho) - E[a(Z,\varrho)]\big)\right| \geq \frac{1}{2}a_T\right] \leq 2\exp\left(-\frac{1}{8C_2}Ta_T^2\right).$$

Then we get $P[\sup_{\varrho\in[0,1]^4}|\frac{1}{T}\sum_{t=1}^{T}(a(Z_t,\varrho)-E[a(Z,\varrho)])|\geq a_T]=O(T^{-\bar{b}})$ for any $\bar{b}>0$, since Ta_T^2 diverges as a power of T.

SM.8.7.2. *Proof of* $P[\|\Delta\hat{\psi}\|^2\geq d^2\lambda_T]=O(T^{-\bar{b}})$

We have from (A.4) that on set $\{\|\Delta\hat{\varphi}\|<r\wedge\|\hat{\mathcal{K}}(\Delta\hat{\varphi})\|\leq\frac{C}{\sqrt{\lambda_T}}\|\Delta\hat{\varphi}\|^2\}$ it holds that

$$\|\Delta\hat{\psi}\|=\|\Delta\hat{\varphi}-\mathcal{K}_T(\Delta\hat{\varphi})\|\leq\|\Delta\hat{\varphi}\|+\frac{C}{\sqrt{\lambda_T}}\|\Delta\hat{\varphi}\|^2,$$

where $C>\frac{1}{2\sqrt{\tau(1-\tau)}}\sup_{x,y,z}|\nabla_y f_{X,Y|Z}(x,y|z)|$. Then we get

$$\begin{aligned}\text{(SM.48)}\quad & P[\|\Delta\hat{\psi}\|^2\geq d^2\lambda_T]\\ &\leq P[\|\Delta\hat{\varphi}\|\geq r]+P\left[\|\hat{\mathcal{K}}(\Delta\hat{\varphi})\|>\frac{C}{\sqrt{\lambda_T}}\|\Delta\hat{\varphi}\|^2\right]\\ &\quad+P\left[\|\Delta\hat{\varphi}\|\geq\frac{1}{2}d\sqrt{\lambda_T}\right]+P\left[\frac{C}{\sqrt{\lambda_T}}\|\Delta\hat{\varphi}\|^2\geq\frac{1}{2}d\sqrt{\lambda_T}\right].\end{aligned}$$

The second term in the RHS is $O(T^{-\bar{b}})$ for any $\bar{b}>0$ by Lemma A.5. The first, third, and fourth terms in the RHS are $O(T^{-\bar{b}})$ for any $\bar{b}>0$ by part (i).

SM.8.8. *Proof of Lemma B.8*

SM.8.8.1. *Proof of Part (i)*

The proof of part (i) is similar to GS, proof of Lemma A.3. The main modification concerns the nondifferentiability of the moment function w.r.t. data (see Lemmas C.2 and C.3 below).

Let us first expand the function $(\lambda_T+A^*A)^{-1}A^*(\hat{\zeta}-E\hat{\zeta})$ w.r.t. the basis of eigenfunctions $\{\phi_j\}$ of operator A^*A, with eigenvalues ν_j. We have

$$\begin{aligned}&(\lambda_T+A^*A)^{-1}A^*(\hat{\zeta}-E\hat{\zeta})\\ &=\sum_{j=1}^{\infty}\langle\phi_j,(\lambda_T+A^*A)^{-1}A^*(\hat{\zeta}-E\hat{\zeta})\rangle_H\phi_j\\ &=\sum_{j=1}^{\infty}\langle(\lambda_T+A^*A)^{-1}\phi_j,A^*(\hat{\zeta}-E\hat{\zeta})\rangle_H\phi_j\\ &=\sum_{j=1}^{\infty}\frac{1}{\lambda_T+\nu_j}\langle\phi_j,A^*(\hat{\zeta}-E\hat{\zeta})\rangle_H\phi_j.\end{aligned}$$

Define the variables for $j \in \mathbb{N}$,

$$Z_{j,T} := \frac{1}{\sqrt{\nu_j}} \langle \phi_j, \sqrt{T} A^*(\hat{\zeta} - E\hat{\zeta}) \rangle_H \tag{SM.49}$$

$$= \frac{1}{\sqrt{\nu_j}} \langle A\phi_j, \sqrt{T}(\hat{\zeta} - E\hat{\zeta}) \rangle_{L^2(F_Z,\tau)}$$

$$= -\frac{\sqrt{T}}{\tau(1-\tau)} \int\int \frac{1}{\sqrt{\nu_j}} (A\phi_j)(z) 1_{\varphi_0}(w) \times [\hat{f}_{X,Y,Z}(w,z) - E\hat{f}_{X,Y,Z}(w,z)]\, dw\, dz, \tag{SM.50}$$

where $1_{\varphi_0}(w) := 1\{y \le \varphi_0(x)\} - \tau$ and $w := (x, y)$. Then we can write $(\lambda_T + A^*A)^{-1} A^*(\hat{\zeta} - E\hat{\zeta}) = \frac{1}{\sqrt{T}} \sum_{j=1}^{\infty} \frac{\sqrt{\nu_j}}{\lambda_T + \nu_j} Z_{j,T} \phi_j$. We deduce

$$E\big[\|(\lambda_T + A^*A)^{-1} A^*(\hat{\zeta} - E\hat{\zeta})\|^2\big] = \frac{1}{T} \sum_{j,l=1}^{\infty} \frac{\sqrt{\nu_j}}{\lambda_T + \nu_j} \frac{\sqrt{\nu_l}}{\lambda_T + \nu_l} \langle \phi_j, \phi_l \rangle E[Z_{j,T} Z_{l,T}]. \tag{SM.51}$$

To derive the asymptotic behavior of the RHS, we need the following lemmas.

LEMMA C.2: *Let* $g(r) = \psi(z)(1\{y \le \varphi_0(x)\} - \tau)$, *where* $r := (x, y, z) = (w, z) \in \mathbb{R}^2 \times \mathbb{R}^{d_Z} = \mathbb{R}^d$. *Let* $F_{X,Y,Z}$ *be a distribution on* $\mathbb{R}^d$ *with margins* F_W *on* $\mathbb{R}^2$ *and* F_Z *on* $\mathbb{R}^{d_Z}$, *and let* $\psi, \nabla\psi \in L^2(F_Z)$. *Let* k *be a bounded product kernel on* $\mathbb{R}^d$ *with compact support. Define the function* $\Delta g_h(r) := \int \frac{1}{h^d} k(\frac{v}{h}) |g(r - v) - g(r)|\, dv$, $r \in \mathbb{R}^d$, *for* $h > 0$. *Then there exists a constant* $c > 0$ (*independent of* h) *such that*

$$\|\Delta g_h\|^2_{L^2(F)} \le ch^2 \|\nabla\psi\|^2_{L^2(F_Z)} + ch^2 \int |\nabla\psi(z)|^2 \sup_{\zeta \in B_{ch}(0)} |f_Z(z + \zeta) - f_Z(z)|\, dz + c\|\psi\|^2_{L^2(F_Z)} \int 1\big\{|y - \varphi_0(x)| \le c\sqrt{h}\big\} f_W(w)\, dw.$$

LEMMA C.3: *Let* $\{R_t = (X_t, Y_t, Z_t) : t = 1, \ldots, T\}$ *be i.i.d. variables with value in a convex set* $\mathcal{S} \subset \mathbb{R}^d$ *and density* $f_{X,Y,Z}$ *satisfying Assumptions* A.1(ii) *and* A.4(i), *and such that* $E[1\{Y \le \varphi_0(X)\} - \tau \mid Z] = 0$. *Let* $\hat{f}_{X,Y,Z}$ *denote the kernel estimator of* $f_{X,Y,Z}$, *with kernel* K *satisfying Assumption* A.2 *and band-*

width $h_T \to 0$. Let $\mathcal{G}$ denote the set of functions

$$\mathcal{G} = \big\{g(r) = \psi_g(z)\big(1\{y \le \varphi_0(x)\} - \tau\big) : \|\psi_g\|_{H^1(F_Z)} < \infty, \\ \|\psi_g\|_{H^1(\bar{F}_Z)} < \infty\big\},$$

where $\|\cdot\|_{H^1(\bar{F}_Z)}$ denotes the Sobolev norm w.r.t. the density $\bar{f}_Z(z) := \frac{\int q(w,z)\,dw}{\int q(w,z)\,dw\,dz}$, with q as in Assumption A.4(i), *that is, $\|\psi\|^2_{H^1(\bar{F}_Z)} = \int \psi^2 \bar{f}_Z + \int (\nabla\psi)^2 \bar{f}_Z$. Further, for $g \in \mathcal{G}$ and $h > 0$, denote $\rho(g,h)^2 := \int g(r)^2 1\{r \in \partial\mathcal{S}(h)\} f_{X,Y,Z}(r)\,dr$, where $\partial\mathcal{S}(h) = \{r \in \mathcal{S} : \mathrm{dist}(r, \mathcal{S}^c) \le h\}$. Define $V_T(g) := \sqrt{T}\int g(r)[\hat{f}_{X,Y,Z}(r) - E\hat{f}_{X,Y,Z}(r)]\,dr$, $T \in \mathbb{N}$, for $g \in \mathcal{G}$. Then*

$$\begin{aligned} E[V_T(g)V_T(e)] &= \mathrm{Cov}[g(R), e(R)] + O(\rho(g,\kappa h_T)\rho(e,\kappa h_T)) \\ &\quad + O\big(h_T^{1/4}\big(\|\psi_g\|_{H^1(F_Z)} + \|\psi_g\|_{H^1(\bar{F}_Z)}\big) \\ &\quad \times \big(\|\psi_e\|_{H^1(F_Z)} + \|\psi_e\|_{H^1(\bar{F}_Z)}\big)\big) \end{aligned}$$

uniformly in $g, e \in \mathcal{G}$, for a constant $\kappa > 0$.

LEMMA C.4: *Under Assumptions* A.1, A.2, A.4(i), *and* A.5(ii) *and* (iii), *$E[Z^2_{j,T}] = 1 + O(h_T^{1/4} j^2)$ uniformly in $j \in \mathbb{N}$.*

LEMMA C.5: *Let $\{Z_j : j = 1, 2, \ldots\}$ be a sequence of zero mean random variables (r.v.'s) and let $(\alpha_{j,l})$, $j, l = 1, 2, \ldots$, be an array of positive numbers. Denote the correlation $\rho_{j,l} = \mathrm{corr}(Z_j, Z_l)$. Then*

$$\left|\sum_{j,l=1}^{\infty} \alpha_{j,l} E[Z_j Z_l] - \sum_{j=1}^{\infty} \alpha_{j,j} E[Z_j^2]\right| \\ \le \left(\sum_{j=1}^{\infty} \alpha_{j,j} E[Z_j^2]\right)\left(\sum_{j,l=1, j\ne l}^{\infty} \rho_{jl}^2 \frac{\alpha_{j,l}^2}{\alpha_{j,j}\alpha_{l,l}}\right)^{1/2}.$$

Let us now conclude the proof of Lemma B.8(i). We apply Lemma C.5 to sequence $Z_j = Z_{j,T}$ in (SM.49) with $\alpha_{j,l} := \frac{1}{T}\frac{\sqrt{\nu_j}}{\lambda_T + \nu_j}\frac{\sqrt{\nu_l}}{\lambda_T + \nu_l}\langle\phi_j, \phi_l\rangle$. It follows from (SM.51) that

$$\text{(SM.52)} \quad E\big[\|(\lambda_T + A^*A)^{-1}A^*(\hat{\psi} - E\hat{\psi})\|^2\big] \\ = \left(\frac{1}{T}\sum_{j=1}^{\infty} \frac{\nu_j}{(\lambda_T + \nu_j)^2}\|\phi_j\|^2 E[Z^2_{j,T}]\right)(1 + R_{1,T}),$$

where $|R_{1,T}| \leq (\sum_{j,l=1,j\neq l}^{\infty} \rho_{jl,T}^2 \frac{\langle\phi_j,\phi_l\rangle^2}{\|\phi_j\|^2\|\phi_l\|^2})^{1/2}$ and $\rho_{jl,T} = \mathrm{corr}(Z_{j,T}, Z_{l,T})$. By using $|\rho_{jl,T}| \leq 1$ and Assumption A.5(i), it follows that $R_{1,T} = O(1)$. Furthermore, from Lemma C.4,

$$\frac{1}{T}\sum_{j=1}^{\infty}\frac{\nu_j}{(\lambda_T+\nu_j)^2}\|\phi_j\|^2 E[Z_{j,T}^2] = \frac{1}{T}\sum_{j=1}^{\infty}\frac{\nu_j}{(\lambda_T+\nu_j)^2}\|\phi_j\|^2 + O\left(h_T^{1/4}\frac{1}{T}\sum_{j=1}^{\infty}\frac{\nu_j}{(\lambda_T+\nu_j)^2}\|\phi_j\|^2 j^2\right) = V_T(\lambda_T)\left(1+O\left(h_T^{1/4}\frac{V_T(\lambda_T;2)}{V_T(\lambda_T)}\right)\right). \tag{SM.53}$$

From (SM.52) and (SM.53), the conclusion follows.

SM.8.8.2. *Proof of Part (ii)*

The proof follows from the next lemma and Lemma B.8(i).

LEMMA C.6: *Under Assumptions* A.1, A.2, A.4(i), *and* A.5, $E[\|(\lambda_T + A^*A)^{-1}A^*(\hat{\zeta} - E\hat{\zeta})\|^4] = O(E[\|(\lambda_T + A^*A)^{-1}A^*(\hat{\zeta} - E\hat{\zeta})\|^2]^2)$.

SM.8.9. *Proof of Lemma B.9*

The proof is omitted since it is analogous to that of Lemma A.4 in GS. The reason is that $-\hat{\zeta}$ is an integral of the moment function $1\{y \leq \varphi_0(x)\} - \tau$ w.r.t. the kernel density estimator, and the proof of Lemma A.4 in GS allows for nondifferentiable moment functions.

SM.8.10. *Proof of Lemma B.10*

We separate the proof of part (i) into parts (a) and (b).

(a) First, we show that if $\gamma < \frac{1}{2}\min\{1-(d_Z+1)\eta, m\eta\}$, $E[\|(1+S(\lambda_T)\hat{U})^{-1} \times S(\lambda_T)\hat{U}\|^8] = O(1)$ and $E[\|S(\lambda_T)\hat{U}\|^8] = o(1)$, where m is the order of the kernel K, d_Z is the dimension of Z, $S(\lambda_T) := (\lambda_T + A^*A)^{-1}$, and $\hat{U} := \hat{A}_0^*\hat{A}_0 - A^*A$, then under Assumptions A.1, A.2, A.3(ii), A.4, and A.5, we have $E[\|\mathcal{R}_T\|^2] = o(E[\|\mathcal{V}_T + \mathcal{B}_T\|^2])$.

(b) Second, we show that if $\gamma < \frac{1}{2}\min\{1-2\eta, 2m\eta\}$ and $\eta < \frac{1}{2(d_Z+1)}$, then under Assumptions A.1, A.2, A.4, and A.5, we have $E[\|(1+S(\lambda_T)\hat{U})^{-1}S(\lambda_T) \times \hat{U}\|^8] = O(1)$ and $E[\|S(\lambda_T)\hat{U}\|^8] = o(1)$.

Combining (a) and (b), and using that $E[\|\mathcal{V}_T + \mathcal{B}_T\|^2] = O(M_T(\lambda_T))$ from (SM.21) and Lemmas B.8 and B.9, part (i) follows. The proof of part (ii) follows the same lines and is therefore omitted.

SM.8.10.1. *Proof of Part (i)(a)*

The proof of part (i)(a) follows the body of the proof of part (i) of Lemma A.5 in GS. The main modifications concern the proof of Lemma C.7 below and the bound of term $\mathcal{R}_{5,T}$ below. Let us write

$$
\begin{aligned}
\text{(SM.54)}\quad \mathcal{R}_T &= [(\lambda_T + \hat{A}_0^*\hat{A}_0)^{-1}\hat{A}_0^*\hat{A}_0 - (\lambda_T + A^*A)^{-1}A^*A]\varphi_0 \\
&\quad + [(\lambda_T + \hat{A}_0^*\hat{A}_0)^{-1} - (\lambda_T + A^*A)^{-1}]A^*E\hat{\zeta} \\
&\quad + [(\lambda_T + \hat{A}_0^*\hat{A}_0)^{-1} - (\lambda_T + A^*A)^{-1}]A^*(\hat{\zeta} - E\hat{\zeta}) \\
&\quad + (\lambda_T + \hat{A}_0^*\hat{A}_0)^{-1}(\hat{A}_0^*(\hat{\zeta} - \hat{q}) - A^*\hat{\zeta}) \\
&\quad - (\lambda_T + \hat{A}_0^*\hat{A}_0)^{-1}(\hat{A}^* - \hat{A}_0^*)(\hat{\mathcal{A}}(\hat{\varphi}) - \tau) \\
&=: \mathcal{R}_{1,T} + \mathcal{R}_{2,T} + \mathcal{R}_{3,T} + \mathcal{R}_{4,T} + \mathcal{R}_{5,T}.
\end{aligned}
$$

We bound the terms $\mathcal{R}_{i,T}$ separately.

• *Bound of* $E[\|\mathcal{R}_{1,T}\|^2]$. We can write

$$
\begin{aligned}
&(\lambda_T + \hat{A}_0^*\hat{A}_0)^{-1}\hat{A}_0^*\hat{A}_0 - (\lambda_T + A^*A)^{-1}A^*A \\
&\quad = (\lambda_T + \hat{A}_0^*\hat{A}_0)^{-1}(\hat{A}_0^*\hat{A}_0 - A^*A) \\
&\qquad + [(\lambda_T + \hat{A}_0^*\hat{A}_0)^{-1} - (\lambda_T + A^*A)^{-1}]A^*A \\
&\quad = (\lambda_T + \hat{A}_0^*\hat{A}_0)^{-1}(\hat{A}_0^*\hat{A}_0 - A^*A) \\
&\qquad - (\lambda_T + \hat{A}_0^*\hat{A}_0)^{-1}(\hat{A}_0^*\hat{A}_0 - A^*A)(\lambda_T + A^*A)^{-1}A^*A \\
&\quad = -(\lambda_T + \hat{A}_0^*\hat{A}_0)^{-1}(\hat{A}_0^*\hat{A}_0 - A^*A)[(\lambda_T + A^*A)^{-1}A^*A - 1].
\end{aligned}
$$

Thus, we get $\mathcal{R}_{1,T} = -\hat{S}(\lambda_T)\hat{U}\mathcal{B}_T$, where $\hat{S}(\lambda_T) := (\lambda_T + \hat{A}_0^*\hat{A}_0)^{-1}$. Moreover, using $\hat{S}(\lambda_T) - S(\lambda_T) = -(1 + S(\lambda_T)\hat{U})^{-1}S(\lambda_T)\hat{U}S(\lambda_T)$, we get $\mathcal{R}_{1,T} = [(1 + S(\lambda_T)\hat{U})^{-1}S(\lambda_T)\hat{U} - 1]S(\lambda_T)\hat{U}\mathcal{B}_T$. Thus

$$
\begin{aligned}
\|\mathcal{R}_{1,T}\| &\le \big\|(1 + S(\lambda_T)\hat{U})^{-1}S(\lambda_T)\hat{U} - 1\big\| \|S(\lambda_T)\hat{U}\| \|\mathcal{B}_T\| \\
&\le \big(\big\|(1 + S(\lambda_T)\hat{U})^{-1}S(\lambda_T)\hat{U}\big\| + 1\big)\|S(\lambda_T)\hat{U}\| b(\lambda_T),
\end{aligned}
$$

where

$$
b(\lambda_T)^2 := \|\mathcal{B}_T\|^2 = \int \mathcal{B}_T(x)^2\,dx.
$$

We conclude, by Cauchy–Schwarz inequality, that

$$E[\|\mathcal{R}_{1,T}\|^2] \leq E\big[\big(\big\|(1+S(\lambda_T)\hat{U})^{-1}S(\lambda_T)\hat{U}\big\|+1\big)^4\big]^{1/2} \times E\big[\|S(\lambda_T)\hat{U}\|^4\big]^{1/2} b(\lambda_T)^2.$$

Since the first term of the RHS is $O(1)$ and the second is $o(1)$, it follows that

$$E[\|\mathcal{R}_{1,T}\|^2]^{1/2} = o(b(\lambda_T)). \tag{SM.55}$$

• *Bound of* $E[\|\mathcal{R}_{2,T}\|^2]$. Similarly to previous lines, we have

$$\begin{aligned}(\lambda_T + \hat{A}_0^*\hat{A}_0)^{-1} - (\lambda_T + A^*A)^{-1} \\ = -(\lambda_T + \hat{A}_0^*\hat{A}_0)^{-1}(\hat{A}_0^*\hat{A}_0 - A^*A)(\lambda_T + A^*A)^{-1}.\end{aligned} \tag{SM.56}$$

Thus, we get

$$\begin{aligned}\mathcal{R}_{2,T} &= -\hat{S}(\lambda_T)\hat{U}S(\lambda_T)A^*E\hat{\zeta} \\ &= \big[(1+S(\lambda_T)\hat{U})^{-1}S(\lambda_T)\hat{U} - 1\big]S(\lambda_T)\hat{U}S(\lambda_T)A^*E\hat{\zeta}\end{aligned}$$

from the arguments in the first point above. Thus, by Cauchy–Schwarz inequality,

$$E[\|\mathcal{R}_{2,T}\|^2]^{1/2} \leq E\big[\big(\big\|(1+S(\lambda_T)\hat{U})^{-1}S(\lambda_T)\hat{U}\big\|+1\big)^4\big]^{1/4} \times E\big[\|S(\lambda_T)\hat{U}\|^4\big]^{1/4}\|S(\lambda_T)A^*E\hat{\zeta}\|.$$

Using $\|S(\lambda_T)A^*E\hat{\zeta}\| = O(\frac{1}{\lambda_T}\|E\hat{\zeta}\|_{L^2(F_Z,\tau)}) = O(\frac{1}{\lambda_T}h_T^m)$, $h_T^m = \lambda_T^2$, and $\lambda_T = o(b(\lambda_T))$, we get $\|S(\lambda_T)A^*E\hat{\zeta}\| = o(b(\lambda_T))$. It follows that

$$E[\|\mathcal{R}_{2,T}\|^2]^{1/2} = o(b(\lambda_T)). \tag{SM.57}$$

• *Bound of* $E[\|\mathcal{R}_{3,T}\|^2]$. From (SM.56) and the arguments in the first point above, we have

$$\begin{aligned}\mathcal{R}_{3,T} &= -(\lambda_T + \hat{A}_0^*\hat{A}_0)^{-1}(\hat{A}_0^*\hat{A}_0 - A^*A)(\lambda_T + A^*A)^{-1}A^*(\hat{\zeta} - E\hat{\zeta}) \\ &= \big[(1+S(\lambda_T)\hat{U})^{-1}S(\lambda_T)\hat{U} - 1\big]S(\lambda_T)\hat{U}S(\lambda_T)A^*(\hat{\zeta} - E\hat{\zeta}).\end{aligned}$$

Thus we get

$$E[\|\mathcal{R}_{3,T}\|^2]^{1/2} \leq E\big[\big(\big\|(1+S(\lambda_T)\hat{U})^{-1}S(\lambda_T)\hat{U}\big\|+1\big)^2 \times \|S(\lambda_T)\hat{U}\|^2\|S(\lambda_T)A^*(\hat{\zeta} - E\hat{\zeta})\|^2\big]^{1/2}$$

$$\leq E\big[\big(\big\|(1+S(\lambda_T)\hat{U})^{-1}S(\lambda_T)\hat{U}\big\|+1\big)^4\|S(\lambda_T)\hat{U}\|^4\big]^{1/4}$$
$$\times E\big[\|S(\lambda_T)A^*(\hat{\zeta}-E\hat{\zeta})\|^4\big]^{1/4}$$
$$\leq E\big[\big(\big\|(1+S(\lambda_T)\hat{U})^{-1}S(\lambda_T)\hat{U}\big\|+1\big)^8\big]^{1/8}$$
$$\times E\big[\|S(\lambda_T)\hat{U}\|^8\big]^{1/8}E\big[\|S(\lambda_T)A^*(\hat{\zeta}-E\hat{\zeta})\|^4\big]^{1/4}.$$

The first two terms of the RHS are $O(1)$ and $o(1)$, respectively. Then it follows that

(SM.58) $$E[\|\mathcal{R}_{3,T}\|^2]^{1/2}=o\big(E\big[\|(\lambda_T+A^*A)^{-1}A^*(\hat{\zeta}-E\hat{\zeta})\|^2\big]^{1/2}\big)$$

from Lemma C.6.

• *Bound of* $E[\|\mathcal{R}_{4,T}\|^2]$. We have $\|\mathcal{R}_{4,T}\|\leq\|\mathcal{R}_{4,T}\|_H\leq\|\hat{S}(\lambda_T)\|_H\|\hat{A}_0^*(\hat{\zeta}-\hat{q})-A^*\hat{\zeta}\|_H$. Moreover, $\|\hat{S}(\lambda_T)\|_H\leq 1/\lambda_T$ P-a.s. Now we use that $\hat{A}_0^*=\mathcal{E}\mathcal{D}^{-1}\tilde{\hat{A}}_0$, where

$$\tilde{\hat{A}}_0\psi(x):=\frac{1}{T\tau(1-\tau)}\sum_{t=1}^{T}\hat{f}_{X,Y|Z}(x,\varphi_0(x)|Z_t)\psi(Z_t).$$

From Lemmas A.17(iii) and A.18(ii) in Section SM.3, the operator $\mathcal{E}\mathcal{D}^{-1}$ from $L^2[0,1]$ to $H^l[0,1]$ is continuous. We get $\|\hat{A}_0^*(\hat{\zeta}-\hat{q})-A^*\hat{\zeta}\|_H=O(\|\tilde{\hat{A}}_0(\hat{\zeta}-\hat{q})-\tilde{A}\hat{\zeta}\|)$. We get

(SM.59) $$E[\|\mathcal{R}_{4,T}\|^2]=o(b(\lambda_T)^2),$$

using the next lemma.

LEMMA C.7: *Under Assumptions* A.1, A.2, *and* A.5, *and* $\gamma<\frac{1}{2}\min\{1-(d_Z+1)\eta, m\eta\}$, *we have* $E[\|\tilde{\hat{A}}_0(\hat{\zeta}-\hat{q})-\tilde{A}\hat{\zeta}\|^2]^{1/2}=o(\lambda_T b(\lambda_T))$.

• *Bound of* $E[\|\mathcal{R}_{5,T}\|^2]$. We have

$$\|\mathcal{R}_{5,T}\|\leq\|\mathcal{R}_{5,T}\|_H$$
$$\leq(1/\lambda_T)\|\mathcal{E}\mathcal{D}^{-1}\|_{HL}\|\tilde{\hat{A}}-\tilde{\hat{A}}_0\|_{\mathcal{L}(L^2(\hat{F}_{Z,\tau}),L^2[0,1])}$$
$$\times\|\hat{\mathcal{A}}(\hat{\varphi})-\tau\|_{L^2(\hat{F}_{Z,\tau})},$$

where $\|\mathcal{E}\mathcal{D}^{-1}\|_{HL}:=\|\mathcal{E}\mathcal{D}^{-1}\|_{\mathcal{L}(L^2[0,1],H^l[0,1])}<\infty$ (Lemmas A.17(iii) and A.18(ii)) and

$$\tilde{\hat{A}}\psi(x):=\frac{1}{T\tau(1-\tau)}\sum_{t=1}^{T}\hat{f}_{X,Y|Z}(x,\hat{\varphi}(x)|Z_t)\psi(Z_t).$$

Using the triangular inequality and the Cauchy–Schwarz inequality, for $\psi \in L^2(\hat{F}_Z, \tau)$,

$$\begin{aligned}
&\|(\tilde{\hat{A}} - \tilde{\hat{A}}_0)\psi\| \\
&\quad \le \frac{1}{\tau(1-\tau)T} \sum_{t=1}^{T} |\psi(Z_t)| \\
&\qquad \times \left(\int \left[\hat{f}_{X,Y|Z}(x, \hat{\varphi}(x)|Z_t) - \hat{f}_{X,Y|Z}(x, \varphi_0(x)|Z_t) \right]^2 dx \right)^{1/2} \\
&\quad \le \frac{1}{\sqrt{\tau(1-\tau)}} \|\psi\|_{L^2(\hat{F}_Z,\tau)} \\
&\qquad \times \sup_{z \in \mathcal{Z}} \left(\int \left[\hat{f}_{X,Y|Z}(x, \hat{\varphi}(x)|z) - \hat{f}_{X,Y|Z}(x, \varphi_0(x)|z) \right]^2 dx \right)^{1/2}.
\end{aligned}$$

Further, from the mean value theorem, for $z \in \mathcal{Z}$, we have

$$\int \left[\hat{f}_{X,Y|Z}(x, \hat{\varphi}(x)|z) - \hat{f}_{X,Y|Z}(x, \varphi_0(x)|z) \right]^2 dx \le \hat{\sigma}^2 \|\Delta\hat{\varphi}\|^2,$$

where $\hat{\sigma} := \sup_{x \in [0,1], y \in \mathbb{R}, z \in \mathcal{Z}} |\nabla_y \hat{f}_{X,Y|Z}(x, y, z)|$. Thus, we get

$$\|\tilde{\hat{A}} - \tilde{\hat{A}}_0\|_{\mathcal{L}(L^2(\hat{F}_Z,\tau), L^2[0,1])} \le C\hat{\sigma}\|\Delta\hat{\varphi}\|$$

and

$$E[\|\mathcal{R}_{5,T}\|^2] \le \frac{C}{\lambda_T^2} E\big[\hat{\sigma}^2 \|\Delta\hat{\varphi}\|^2 \|\hat{\mathcal{A}}(\hat{\varphi}) - \tau\|^2_{L^2(\hat{F}_Z,\tau)}\big].$$

Now, since $\|\Delta\hat{\varphi}\|^2 \le 4 \sup_{\varphi \in \Theta} \|\varphi\|^2$, $\|\hat{\mathcal{A}}(\hat{\varphi}) - \tau\|^2_{L^2(\hat{F}_Z,\tau)} \le \frac{4}{\tau(1-\tau)}$, and $\hat{\sigma} \le C h_T^{-3-d_Z}$, we have $E[\hat{\sigma}^2 \|\Delta\hat{\varphi}\|^2 \|\hat{\mathcal{A}}(\hat{\varphi}) - \tau\|^2_{L^2(\hat{F}_Z,\tau)}] \le C(\log T)^2 [\|\Delta\hat{\varphi}\|^2 \|\hat{\mathcal{A}}(\hat{\varphi}) - \tau\|^2_{L^2(\hat{F}_Z,\tau)} 1\{\hat{\sigma} \le c \log T\}] + C h_T^{-6-2d_Z} E[1\{\hat{\sigma} > c \log T\}]$. For large c, the second term is $O(T^{-\bar{b}})$ for any $\bar{b} > 0$. We deduce that $E[\|\mathcal{R}_{5,T}\|^2] = O(\frac{(\log T)^2}{\lambda_T^2} E[\|\Delta\hat{\varphi}\|^2 \times \|\hat{\mathcal{A}}(\hat{\varphi}) - \tau\|^2_{L^2(\hat{F}_Z,\tau)}])$. By a similar argument as above, using that $\|\hat{\mathcal{A}}(\hat{\varphi}) - \tau\|^2_{L^2(\hat{F}_Z,\tau)}$ is bounded and is of the order $\frac{1}{T h_T^{d_Z}} + h_T^{2m}$, and a large deviation bound argument, it follows that

$$\begin{aligned}
\text{(SM.60)} \quad E[\|\mathcal{R}_{5,T}\|^2] &= O\left(E[\|\Delta\hat{\varphi}\|^2] \frac{(\log T)^2}{\lambda_T^2} \left(\frac{1}{T h_T^{d_Z}} + h_T^{2m} \right) \right) \\
&= o\big(E[\|\Delta\hat{\varphi}\|^2]\big).
\end{aligned}$$

Now, using Lemmas B.4 and B.7, and $E[\frac{1}{\sqrt{\lambda_T}}\|\Delta\hat{\psi}\|^3] \le E[\|\Delta\hat{\psi}\|^2 d1\{\|\Delta\hat{\psi}\| \le d\sqrt{\lambda_T}\}] + E[\frac{1}{\sqrt{\lambda_T}}\|\Delta\hat{\psi}\|^3 1\{\|\Delta\hat{\psi}\| \ge d\sqrt{\lambda_T}\}] = O(E[\|\Delta\hat{\psi}\|^2])$ for $d > 0$, we get $E[\|\Delta\hat{\varphi}\|^2] = O(E[\|\Delta\hat{\psi}\|^2])$. Thus, from (SM.55), (SM.57), (SM.58), (SM.59), and (SM.60), and triangular inequality, we deduce

$$E[\|\mathcal{R}_T\|^2] = o\big(b(\lambda_T)^2 + E\big[\|(\lambda_T + A^*A)^{-1}A^*(\hat{\zeta} - E\hat{\zeta})\|^2\big] + E[\|\Delta\hat{\psi}\|^2]\big). \tag{SM.61}$$

From (SM.19) and Lemmas B.8 and B.9, we get $E[\|\Delta\hat{\psi}\|^2] = O(b(\lambda_T)^2 + E[\|(\lambda_T + A^*A)^{-1}A^*(\hat{\zeta} - E\hat{\zeta})\|^2])$. From (SM.61), we get

$$E[\|\mathcal{R}_T\|^2] = o\big(b(\lambda_T)^2 + E\big[\|(\lambda_T + A^*A)^{-1}A^*(\hat{\zeta} - E\hat{\zeta})\|^2\big]\big). \tag{SM.62}$$

Further, using that (see (SM.21) and Lemma B.9) $E[\|\mathcal{V}_T + \mathcal{B}_T\|^2] = b(\lambda_T)^2 + E[\|(\lambda_T + A^*A)^{-1}A^*(\hat{\zeta} - E\hat{\zeta})\|^2]$, up to terms that are negligible asymptotically w.r.t. the RHS, the conclusion follows.

SM.8.10.2. *Proof of Part (i)(b)*

We have $\|S(\lambda_T)\hat{U}\| \le \|S(\lambda_T)\hat{U}\|_{HL} \le \|S(\lambda_T)\|_H \|\hat{U}\|_{HL}$, where $\|S(\lambda_T)\hat{U}\| := \|S(\lambda_T)\hat{U}\|_{\mathcal{L}(L^2[0,1],L^2[0,1])}$ denotes operator norm in $L^2[0,1]$. Using $\|S(\lambda_T)\|_H \le 1/\lambda_T$ and $\|\hat{U}\|_{HL} \le \|\mathcal{E}\mathcal{D}^{-1}\|_{HL}\|\tilde{\hat{A}}_0\hat{A}_0 - \tilde{A}A\|$, P-a.s., we deduce

$$\|S(\lambda_T)\hat{U}\| \le \|S(\lambda_T)\hat{U}\|_{HL} \le \frac{\|\mathcal{E}\mathcal{D}^{-1}\|_{HL}}{\lambda_T}\|\tilde{\hat{A}}_0\hat{A}_0 - \tilde{A}A\|, \quad P\text{-a.s.} \tag{SM.63}$$

Moreover,

$$\begin{aligned}
&\big\|(1 + S(\lambda_T)\hat{U})^{-1}S(\lambda_T)\hat{U}\big\| \\
&\quad \le \big\|(1 + S(\lambda_T)\hat{U})^{-1}S(\lambda_T)\hat{U}\big\|_{HL} \\
&\quad \le \big\|(1 + S(\lambda_T)\hat{U})^{-1}\big\|_H \|S(\lambda_T)\hat{U}\|_{HL}, \quad P\text{-a.s.,}
\end{aligned}$$

which implies

$$\begin{aligned}
&E\big[\big\|(1 + S(\lambda_T)\hat{U})^{-1}S(\lambda_T)\hat{U}\big\|^8\big] \\
&\quad \le E\big[\big\|(1 + S(\lambda_T)\hat{U})^{-1}\big\|_H^{16}\big]^{1/2} E\big[\|S(\lambda_T)\hat{U}\|_{HL}^{16}\big]^{1/2},
\end{aligned} \tag{SM.64}$$

using Cauchy–Schwarz inequality. From the argument of Hall and Horowitz (2005, p. 2925), we have $\|(1 + S(\lambda_T)\hat{U})^{-1}\|_H \le C(1 + \frac{1}{\lambda_T}1\{\|S(\lambda_T)\hat{U}\|_H \ge \frac{1}{2}\}) \le C(1 + \frac{1}{\lambda_T}1\{\|S(\lambda_T)\hat{U}\|_{HL} \ge \frac{1}{2}\})$, P-a.s., where we used that $\|S(\lambda_T)\hat{U}\|_H \le$

$\|S(\lambda_T)\hat{U}\|_{HL}$ in the second inequality. As in the argument of Hall and Horowitz (2005) in their inequality (6.27), from Markov inequality it follows that

$$E\big[\big\|(1+S(\lambda_T)\hat{U})^{-1}\big\|_H^{16}\big] \le C\bigg(1+\frac{1}{\lambda_T^{16}}E\big[\|S(\lambda_T)\hat{U}\|_{HL}^{2l}\big]\bigg) \tag{SM.65}$$

for any $l \in \mathbb{N}$, for a constant C depending on l but not on T. From (SM.63)–(SM.65), and using $\frac{1}{Th_T^2} + h_T^{2m} = O(\lambda_T^{2+\varepsilon})$, $\varepsilon > 0$, the conclusion follows from the next lemma.

LEMMA C.8: *Under Assumptions* A.1, A.2, *and* A.5, *and* $\eta < \frac{1}{2(d_Z+1)}$, *we have* $E[\|\tilde{\hat{A}}_0\hat{A}_0 - \tilde{A}A\|^{2\zeta}] = O(a_T^\zeta)$ *for any* $\zeta \in \mathbb{N}$, *where* $a_T := \frac{1}{Th_T^2} + h_T^{2m}$.

SM.8.11. *Proof of Lemma B.11*

We have $v_j = -(j, \log j)'$. Then $v_j - v_{n_T} = -(j - n_T, \log(j/n_T))'$. We have

$$V'_{(T)}V_{(T)} = \begin{pmatrix} \sum_j (j-n_T)^2 & \sum_j (j-n_T)\log(j/n_T) \\ & \sum_j \log(j/n_T)^2 \end{pmatrix},$$

where $\sum_j$ denotes summation over $j = n_T/T, \dots, n_T - 1$ and

$$D_{(T)} = \begin{pmatrix} \Big[\sum_j (j-n_T)^2\Big]^{-1/2} & \\ & \Big[\sum_j \log(j/n_T)^2\Big]^{-1/2} \end{pmatrix}.$$

Thus

$$D_{(T)}\big(V'_{(T)}V_{(T)}\big)D_{(T)} = \begin{pmatrix} 1 & \dfrac{\sum_j (j-n_T)\log(j/n_T)}{\sqrt{\sum_j (j-n_T)^2}\sqrt{\sum_j \log(j/n_T)^2}} \\ & 1 \end{pmatrix}.$$

Moreover,

$$m_{(T)} = -\begin{pmatrix} \sum_j (j-n_T) \\ \sum_j \log(j/n_T) \end{pmatrix}.$$

To find the asymptotic behavior of $D_{(T)}(V_{(T)}'V_{(T)})D_{(T)}$ and $m_{(T)}$, we use the next lemma.

LEMMA C.9: *We have* (i) $\sum_j (j-n_T) = -\frac{n_T^2}{8}(1+o(1))$, (ii) $\sum_j \log(j/n_T) = \frac{n_T}{2}(\log 2 - 1)\cdot(1+o(1))$, (iii) $\sum_j (j-n_T)^2 = \frac{n_T^3}{24}(1+o(1))$, (iv) $\sum_j \log(j/n_T)^2 = n_T(1-\frac{1}{2}(\log 2)^2 - \log 2)\cdot(1+o(1))$, *and* (v) $\sum_j (j-n_T)\log(j/n_T) = \frac{n_T^2}{16}(5 - 6\log 2)(1+o(1))$.

Thus, we get

$$D_{(T)}\big(V_{(T)}'V_{(T)}\big)D_{(T)} \to \begin{pmatrix} 1 & \dfrac{\sqrt{3}}{4}\dfrac{5-6\log 2}{\sqrt{2-2\log 2-(\log 2)^2}} \\ & 1 \end{pmatrix},$$

which shows part (i). Moreover,

$$d_{(T)} = \begin{pmatrix} O\big(n_T^{-3/2}\big) \\ O\big(n_T^{-1/2}\big) \end{pmatrix}, \quad m_{(T)} = \begin{pmatrix} O(n_T^2) \\ O(n_T) \end{pmatrix},$$

and

$$\zeta_{(T)} = \begin{pmatrix} N_T - n_T \\ \log(N_T/n_T) \end{pmatrix} = \begin{pmatrix} O(n_T) \\ O(1) \end{pmatrix},$$

since $N_T = O(n_T)$. Then part (ii) follows.

SM.8.12. *Proof of Lemma B.12*

From Assumption 5,

$$\begin{aligned} \phi_{S,j}(x)^2 &= \sum_{k=0}^{S-1} c_{2,j-k}\exp(w_{j-k}\beta)\chi_{j-k} \\ &= c_{2,j}\exp(w_j\beta)\sum_{k=0}^{S-1}\frac{c_{2,j-k}}{c_{2,j}}\exp((w_{j-k}-w_j)\beta)\chi_{j-k}. \end{aligned}$$

Then

$$\begin{aligned} &\sum_{k=0}^{S-1}\frac{c_{2,j-k}}{c_{2,j}}\exp((w_{j-k}-w_j)\beta)\chi_{j-k} \\ &\quad = \sum_{k=0}^{S-1}\chi_{j-k} + \sum_{k=0}^{S-1}\left[\frac{c_{2,j-k}}{c_{2,j}}\exp((w_{j-k}-w_j)\beta) - 1\right]\chi_{j-k} \\ &\quad = 1 + o(1) \end{aligned}$$

as $j \to \infty$, since $c_{2,j}$ converges and $w_{j-k} - w_j \to 0$ as $j \to \infty$, for any k. The conclusion follows.

SM.8.13. *Proof of Lemma B.13*

Write

$$\hat{\chi}_j - \chi_j = \frac{2S}{n_T} \sum_{k:n_{T/2}\le k<n_T, k=j \bmod S} \left(\frac{\bar{\phi}_k(x)^2}{\bar{\phi}_{S,k}(x)^2} - \frac{\phi_k(x)^2}{\phi_{S,k}(x)^2} \right) + \frac{2S}{n_T} \sum_{k:n_{T/2}\le k<n_T, k=j \bmod S} \left(\frac{\xi_k^*}{\phi_{S,k}(x)^2} - 1 \right) \chi_k + o(1),$$

where $\xi_k^* = c_{2,k} \exp(w_k \beta)$. Thus, from Lemma B.12, we get

$$|\hat{\chi}_j - \chi_j| \le \sup_{1\le k\le n_T} \left| \frac{\bar{\phi}_k(x)^2}{\bar{\phi}_{S,k}(x)^2} - \frac{\phi_k(x)^2}{\phi_{S,k}(x)^2} \right| + C \sup_{n_T/2\le k} \left| \frac{c_{2,k}}{c_{2,k}^*} - 1 \right| + o(1)$$

for all j and a constant C. By using $\sup_{1\le j\le n_T} \frac{|\bar{\phi}_j(x)^2 - \phi_j(x)^2|}{\xi_j^*} = o_p(1)$ from Appendix A.5, $\sup_{1\le j\le n_T} \frac{|\bar{\phi}_{S,j}(x)^2 - \phi_{S,j}(x)^2|}{\xi_j^*} = o_p(1)$, $\phi_{S,j}(x)^2/\xi_j^* = 1 + o(1)$ from Lemma B.12, and $c_{2,j}^*/c_{2,j} = 1 + o(1)$, the conclusion follows.

SM.8.14. *Proof of Lemma B.14*

The proof is along the lines of the proof of Lemma B.8(i). We use the next lemma, which is a modification of Lemma C.4.

LEMMA C.10: *Under Assumptions* A.1, A.2, A.4(i), *and* A.5(ii)–(iii′), (i) $E[Z_{j,T}^2] = 1 + o(1)$ *uniformly in* $j \in \mathbb{N}$ *and* (ii) $E[Z_{j,T} Z_{l,T}] = o(1)$ *uniformly in* $j, l \in \mathbb{N}$, $j \ne l$.

From Lemma C.10(ii), we get $\text{corr}(Z_{j,T}, Z_{l,T}) = o(1)$ uniformly in $j, l \in \mathbb{N}$, $j \ne l$. Then the conclusion follows from equation (SM.52), Assumption A.5(i), and Lemma C.10(i).

SM.8.15. *Proof of Lemma B.15*

When $\alpha_4 > 2\alpha_1$, the quantity $\sum_{j=1}^{\infty} \frac{a_j}{(\lambda+\nu_j)^2}$ converges to $\sum_{j=1}^{\infty} a_j \nu_j^{-2} < \infty$ as $\lambda \to 0$. Let us now consider the case $\alpha_4 < 2\alpha_1$. Without loss of generality, let $\nu_j = j^{-\alpha_2} e^{-\alpha_1 j}$, $a_j = j^{-\alpha_3} e^{-\alpha_4 j} \chi_j$, and $\chi_j \le 1$. Then

$$\sum_{j=1}^{\infty} \frac{a_j}{(\lambda + \nu_j)^2} = \sum_{j=1}^{\infty} \frac{a_j \nu_j^{-2}}{(\lambda w_j + 1)^2} = \sum_{j=1}^{\infty} \frac{w_j^{\delta}}{(\lambda w_j + 1)^2} j^{\rho} \chi_j,$$

where $w_j := 1/\nu_j$, $\rho := \frac{\alpha_2\alpha_4}{\alpha_1} - \alpha_3$, and $\delta := \frac{2\alpha_1-\alpha_4}{\alpha_1} \in (0,2)$.

Define

$$J(\lambda) := \lambda^\delta n_\lambda^{-\rho} \sum_{j=1}^{\infty} \frac{a_j}{(\lambda+\nu_j)^2} = \sum_{j=1}^{\infty} \frac{(\lambda w_j)^\delta}{(\lambda w_j+1)^2}\left(\frac{j}{n_\lambda}\right)^\rho \chi_j.$$

The conclusion follows if we show that $J(\lambda) \asymp 1$. We give the proof when $\rho \geq 0$ (similar arguments apply when $\rho < 0$). We split $J(\lambda)$ as

$$\begin{aligned} J(\lambda) &= \sum_{j=1}^{N_1(\lambda)} \frac{(\lambda w_j)^\delta}{(\lambda w_j+1)^2}\left(\frac{j}{n_\lambda}\right)^\rho \chi_j + \sum_{j=N_1(\lambda)+1}^{N_2(\lambda)} \frac{(\lambda w_j)^\delta}{(\lambda w_j+1)^2}\left(\frac{j}{n_\lambda}\right)^\rho \chi_j \\ &\quad + \sum_{j=N_2(\lambda)+1}^{\infty} \frac{(\lambda w_j)^\delta}{(\lambda w_j+1)^2}\left(\frac{j}{n_\lambda}\right)^\rho \chi_j \\ &=: J_1(\lambda) + J_2(\lambda) + J_3(\lambda), \end{aligned}$$

where $N_1(\lambda), N_2(\lambda) \in \mathbb{N}$ are such that $N_1(\lambda) < n_\lambda < N_2(\lambda)$, $\frac{\nu_{n_\lambda}}{\nu_{N_1(\lambda)}} = o(1)$, $(\frac{N_2(\lambda)}{n_\lambda})^\rho (\frac{\nu_{N_2(\lambda)}}{\nu_{n_\lambda}})^{2-\delta} = o(1)$, and $r_\lambda := \max\{\frac{n_\lambda - N_1(\lambda)}{n_\lambda}, \frac{N_2(\lambda)-n_\lambda}{n_\lambda}\} = o(1)$. First, we show that $J_i(\lambda) = o(1)$ for $i = 1, 3$. We have

$$\begin{aligned} J_1(\lambda) &\leq \sum_{j=1}^{N_1(\lambda)} \frac{(\lambda w_j)^\delta}{(\lambda w_j+1)^2} \leq \lambda^\delta \sum_{j=1}^{N_1(\lambda)} w_j^\delta \asymp \frac{1}{w_{n_\lambda}^\delta} w_{N_1(\lambda)}^\delta = \left(\frac{\nu_{n_\lambda}}{\nu_{N_1(\lambda)}}\right)^\delta, \\ &= o(1), \end{aligned} \tag{SM.66}$$

where we used $\sum_{j=1}^n w_j^\delta = O(w_n^\delta)$ as $n \to \infty$. Similarly,

$$\begin{aligned} J_3(\lambda) &= \sum_{j=N_2(\lambda)+1}^{\infty} \frac{(\lambda w_j)^\delta}{(\lambda w_j+1)^2}\left(\frac{j}{n_\lambda}\right)^\rho \chi_j \\ &\leq \frac{\lambda^{-2+\delta}}{n_\lambda^\rho} \sum_{j=N_2(\lambda)+1}^{\infty} \nu_j^{2-\delta} j^\rho \\ &\leq \frac{\lambda^{-2+\delta}}{n_\lambda^\rho} \sum_{j=N_2(\lambda)+1}^{\infty} e^{-\zeta j} j^m, \end{aligned} \tag{SM.67}$$

where $\zeta = (2-\delta)\alpha_1 > 0$ and $m = \lceil \rho - \alpha_2(2-\delta) \rceil \leq \rho$. Now, by using

$$\sum_{j=n}^{\infty} e^{-\zeta j} j^m = \left(-\frac{d}{d\tau}\right)^m \sum_{j=n}^{\infty} e^{-\zeta j} = \left(-\frac{d}{d\tau}\right)^m \frac{e^{-\zeta n}}{1-e^{-\zeta}} = O(n^m e^{-\zeta n})$$

as $n \to \infty$, we get

$$J_3(\lambda) = O\left(\frac{\lambda^{-2+\delta}}{n_\lambda^\rho} N_2(\lambda)^\rho \nu_{N_2(\lambda)}^{2-\delta}\right)$$
$$= O\left(\left(\frac{N_2(\lambda)}{n_\lambda}\right)^\rho \left(\frac{\nu_{N_2(\lambda)}}{\nu_{n_\lambda}}\right)^{2-\delta}\right) = o(1).$$

Second, we can write $J_2(\lambda)$ as

$$J_2(\lambda) = \sum_{j=N_1(\lambda)+1}^{N_2(\lambda)} \frac{(\lambda w_j)^\delta}{(\lambda w_j + 1)^2} \chi_j + \sum_{j=N_1(\lambda)+1}^{N_2(\lambda)} \frac{(\lambda w_j)^\delta}{(\lambda w_j + 1)^2} \left[\left(1 + \frac{j - n_\lambda}{n_\lambda}\right)^\rho - 1\right] \chi_j \tag{SM.68}$$

and bound the second term in the RHS. For $N_1(\lambda) + 1 \leq j \leq N_2(\lambda)$, the variable $x := \frac{j-n_\lambda}{n_\lambda}$ is such that $|x| \leq r_\lambda$. Thus, $|(1 + \frac{j-n_\lambda}{n_\lambda})^\rho - 1| = |(1 + x)^\rho - 1| = O(r_\lambda)$, since the function $x \longmapsto (1 + x)^\rho$ has bounded derivative around 0. We deduce that the second term in the RHS of (SM.68) is $o(\sum_{j=N_1(\lambda)+1}^{N_2(\lambda)} \frac{(\lambda w_j)^\delta}{(\lambda w_j+1)^2} \chi_j)$. Hence we get that the sum is such that $J(\lambda) = (\sum_{j=N_1(\lambda)+1}^{N_2(\lambda)} \frac{(\lambda w_j)^\delta}{(\lambda w_j+1)^2} \chi_j)[1 + o(1)] + o(1)$. Moreover, note that

$$\sum_{j=N_1(\lambda)+1}^{N_2(\lambda)} \frac{(\lambda w_j)^\delta}{(\lambda w_j + 1)^2} \chi_j = \sum_{j=1}^{\infty} \frac{(\lambda w_j)^\delta}{(\lambda w_j + 1)^2} \chi_j - \sum_{j=1}^{N_1(\lambda)} \frac{(\lambda w_j)^\delta}{(\lambda w_j + 1)^2} \chi_j - \sum_{j=N_2(\lambda)+1}^{\infty} \frac{(\lambda w_j)^\delta}{(\lambda w_j + 1)^2} \chi_j$$
$$= \sum_{j=1}^{\infty} \frac{(\lambda w_j)^\delta}{(\lambda w_j + 1)^2} \chi_j + o(1)$$

from similar arguments as above and as in (SM.66) and (SM.67). Thus, we have proved that

$$J(\lambda) = \sum_{j=1}^{\infty} \frac{(\lambda w_j)^\delta}{(\lambda w_j + 1)^2} \chi_j [1 + o(1)] + o(1). \tag{SM.69}$$

The conclusion follows if we show that

$$\sum_{j=1}^{\infty} \frac{(\lambda w_j)^\delta}{(\lambda w_j + 1)^2} \chi_j \asymp 1.$$

By the assumptions on sequence χ_j, we can find $\tilde{n}_\lambda$ such that $\nu_{\tilde{n}_\lambda} \asymp \lambda$ and $\chi_{\tilde{n}_\lambda} \asymp 1$. Then the conclusion follows by using that

$$\sum_{j=1}^{\infty} \frac{(\lambda w_j)^\delta}{(\lambda w_j + 1)^2} \chi_j = \sum_{j=1}^{\tilde{n}_\lambda - 1} \frac{(\lambda w_j)^\delta}{(\lambda w_j + 1)^2} \chi_j + \frac{(\lambda w_{\tilde{n}_\lambda})^\delta}{(\lambda w_{\tilde{n}_\lambda} + 1)^2} \chi_{\tilde{n}_\lambda} + \sum_{j=\tilde{n}_\lambda + 1}^{\infty} \frac{(\lambda w_j)^\delta}{(\lambda w_j + 1)^2} \chi_j$$

and

$$\frac{(\lambda w_{\tilde{n}_\lambda})^\delta}{(\lambda w_{\tilde{n}_\lambda} + 1)^2} \chi_{\tilde{n}_\lambda} \asymp 1,$$

$$0 < \sum_{j=1}^{\tilde{n}_\lambda - 1} \frac{(\lambda w_j)^\delta}{(\lambda w_j + 1)^2} \chi_j \leq \sum_{j=1}^{\tilde{n}_\lambda - 1} (\lambda w_j)^\delta = O\big(\big(\lambda w_{\tilde{n}_\lambda}\big)^\delta\big) = O(1),$$

$$0 < \sum_{j=\tilde{n}_\lambda + 1}^{\infty} \frac{(\lambda w_j)^\delta}{(\lambda w_j + 1)^2} \chi_j \leq \lambda^{-2+\delta} \sum_{j=\tilde{n}_\lambda + 1}^{\infty} \nu_j^{2-\delta} = \left(\frac{\nu_{\tilde{n}_\lambda}}{\lambda}\right)^{2-\delta} = O(1).$$

SM.8.16. *Proof of Lemma C.1*

Let $a \in [0, 1]$. By the Cauchy–Schwarz inequality, we have

$$|\varphi(y) - \varphi(x)| = \left| \int_x^y \nabla\varphi(\xi)\, d\xi \right| \leq \sqrt{y - x} \|\nabla\varphi\| \leq \|\nabla\varphi\|$$

for any $x, y \in [0, 1]$. Then

$$\begin{aligned} \|\varphi\|^2 &= \int [\varphi(x) - \varphi(a) + \varphi(a)]^2\, dx \\ &= \varphi(a)^2 + 2\varphi(a) \int [\varphi(x) - \varphi(a)]\, dx + \int [\varphi(x) - \varphi(a)]^2\, dx \\ &\geq \varphi(a)^2 - 2|\varphi(a)| \|\nabla\varphi\|. \end{aligned}$$

Thus $z = |\varphi(a)|$ satisfies the inequality $z^2 - 2\|\nabla\varphi\| z - \|\varphi\|^2 \leq 0$. We get

$$|\varphi(a)| \leq \frac{2\|\nabla\varphi\| + \sqrt{4\|\nabla\varphi\|^2 + 4\|\varphi\|^2}}{2} \leq 2\|\varphi\|_{H^1}.$$

Since this bound holds for any $a \in [0, 1]$, the conclusion follows.

SM.8.17. *Proof of Lemma C.2*

Let $r := (x, y, z) =: (w, z)$ and $v := (x_1, y_1, z_1) =: (w_1, z_1)$. Using $|g(r-v) - g(r)| \leq 2|\psi(z-z_1) - \psi(z)| + |\psi(z)||1_{\varphi_0}(w-w_1) - 1_{\varphi_0}(w)|$, where $1_{\varphi_0}(w) := 1\{y \leq \varphi_0(x)\}$, we have $|\Delta g_h(r)| \leq 2\int \frac{1}{h^{d_Z}} k(\frac{z_1}{h})|\psi(z-z_1) - \psi(z)|\, dz_1 + |\psi(z)| \times \int \frac{1}{h^2} k(\frac{w_1}{h})|1_{\varphi_0}(w-w_1) - 1_{\varphi_0}(w)|\, dw_1$, where, for expository purposes, we denote by k any part k_w or k_z of product kernel $k(r) = k_w(w)k_z(z)$ and we assume $k \geq 0$. Then, by the triangular inequality,

$$\|\Delta g_h\|_{L^2(F)} \leq 2\|\Delta\psi_h\|_{L^2(F_Z)} + \|\psi\|_{L^2(F_Z)}\|\chi\|_{L^2(F_W)}, \tag{SM.70}$$

where $\Delta\psi_h(z) := \int \frac{1}{h^{d_Z}} k(\frac{z_1}{h})|\psi(z-z_1) - \psi(z)|\, dz_1$ and

$$\chi(w) := \int \frac{1}{h^2} k\left(\frac{w_1}{h}\right)\left|1_{\varphi_0}(w-w_1) - 1_{\varphi_0}(w)\right| dw_1.$$

Let us first consider $\|\Delta\psi_h\|_{L^2(F_Z)}$. Write $\Delta\psi_h(z) = \int k(\zeta)|\psi(z-h\zeta) - \psi(z)|\, d\zeta$ and use $\psi(z-h\zeta) - \psi(z) = -\int_0^h (\nabla\psi(z-t\zeta)\cdot\zeta)\, dt$. Thus, we get

$$\begin{aligned}
|\psi(z-h\zeta) - \psi(z)| &\leq \int_0^h |\nabla\psi(z-t\zeta)||\zeta|\, dt \\
&\leq |\zeta|\left(\int_0^h |\nabla\psi(z-t\zeta)|^2\, dt\right)^{1/2}\left(\int_0^h dt\right)^{1/2} \\
&= |\zeta|\sqrt{h}\left(\int_0^h |\nabla\psi(z-t\zeta)|^2\, dt\right)^{1/2}.
\end{aligned}$$

We deduce from the Cauchy–Schwarz inequality that

$$\begin{aligned}
\Delta\psi_h(z) &\leq \sqrt{h}\int k(\zeta)|\zeta|\left(\int_0^h |\nabla\psi(z-t\zeta)|^2\, dt\right)^{1/2} d\zeta \\
&\leq \sqrt{h}w_2^{1/2}\left(\int k(\zeta)\left(\int_0^h |\nabla\psi(z-t\zeta)|^2\, dt\right) d\zeta\right)^{1/2},
\end{aligned}$$

where $w_2 := \int k(\zeta)|\zeta|^2\, d\zeta$. Thus, we get

$$\begin{aligned}
&\int |\Delta\psi_h(z)|^2 f_Z(z)\, dz \\
&\quad \leq hw_2\int_0^h \int k(\zeta)\left(\int |\nabla\psi(z-t\zeta)|^2 f_Z(z)\, dz\right) d\zeta\, dt.
\end{aligned}$$

Now,

$$\int |\nabla\psi(z-t\zeta)|^2 f_Z(z)\,dz = \int |\nabla\psi(z)|^2 f(z+t\zeta)\,dz$$
$$\leq \int |\nabla\psi(z)|^2 f_Z(z)\,dz + \int |\nabla\psi(z)|^2 |f_Z(z+t\zeta)-f_Z(z)|\,dz.$$

Then

$$\int |\Delta\psi_h(z)|^2 f_Z(z)\,dz \leq h^2 w_2 \int |\nabla\psi(z)|^2 f_Z(z)\,dz + h w_2 \int_0^h \int k(\zeta) \int |\nabla\psi(z)|^2 |f_Z(z+t\zeta)-f_Z(z)|\,dz\,d\zeta\,dt.$$

Since k is bounded and has compact support, we get

$$\|\Delta\psi_h\|^2_{L^2(F_Z)} \leq h^2 w_2 \|\nabla\psi\|^2_{L^2(F_Z)} + c_2 h^2 w_2 \int |\nabla\psi(z)|^2 \sup_{\zeta\in B_{c_1 h}(0)} |f_Z(z+\zeta)-f_Z(z)|\,dz \tag{SM.71}$$

for some constants $c_1, c_2 > 0$.

Let us now consider $\|\chi\|_{L^2(F_W)}$. We use $|1_{\varphi_0}(w-w_1) - 1_{\varphi_0}(w)| \leq 1\{y > \varphi_0(x)\}1\{y \leq y_1 + \varphi_0(x-x_1)\} + 1\{y \leq \varphi_0(x)\}1\{y > y_1 + \varphi_0(x-x_1)\}$. Further, using $\varphi_0(x-x_1) \leq \varphi_0(x) + |\int_x^{x-x_1} \nabla\varphi_0(\xi)\,d\xi| \leq \varphi_0(x) + \|\varphi_0\|_H \sqrt{|x_1|}$, and similarly $\varphi_0(x-x_1) \geq \varphi_0(x) - \|\varphi_0\|_H\sqrt{|x_1|}$, we have $1\{y \leq y_1 + \varphi_0(x-x_1)\} \leq 1\{y - \varphi_0(x) \leq y_1 + \|\varphi_0\|_H\sqrt{|x_1|}\}$ and $1\{y > y_1 + \varphi_0(x-x_1)\} \leq 1\{y - \varphi_0(x) > y_1 - \|\varphi_0\|_H\sqrt{|x_1|}\}$. Since function k is bounded and has compact support, there exist constants $c_2, c_3 > 0$ such that $\chi(w) \leq c_3 1\{0 < y - \varphi_0(x) \leq c_2\sqrt{h}\} + c_3 1\{-c_2\sqrt{h} < y - \varphi_0(x) \leq 0\} \leq 2c_3 1\{|y-\varphi_0(x)| \leq c_2\sqrt{h}\}$ for small h. The conclusion follows from (SM.70) and (SM.71) for a suitable c.

SM.8.18. *Proof of Lemma C.3*

In this subsection, we suppress the subscript $_{X,Y,Z}$ in $\hat{f}$ and f. We have

$$E[V_T(g)V_T(e)] = T \int\int g(x)e(y)E\big[(\hat{f}(x) - E\hat{f}(x))(\hat{f}(y) - E\hat{f}(y))\big]\,dx\,dy.$$

Let us compute the cross-moment of the kernel estimator in the RHS:

$$\begin{aligned}
&E\big[(\hat{f}(x)-E\hat{f}(x))(\hat{f}(y)-E\hat{f}(y))\big] \\
&\quad = \frac{1}{Th_T^{2d}}\left\{E\left[K\left(\frac{R-x}{h_T}\right)K\left(\frac{R-y}{h_T}\right)\right]\right. \\
&\qquad \left. - E\left[K\left(\frac{R-x}{h_T}\right)\right]E\left[K\left(\frac{R-y}{h_T}\right)\right]\right\}.
\end{aligned}$$

Thus, we get

$$\begin{aligned}
\text{(SM.72)}\quad & E[V_T(g)V_T(e)] \\
&= \frac{1}{h_T^{2d}}\int\int g(x)e(y)E\left[K\left(\frac{R-x}{h_T}\right)K\left(\frac{R-y}{h_T}\right)\right]dx\,dy \\
&\quad - \left(\frac{1}{h_T^d}\int g(x)E\left[K\left(\frac{R-x}{h_T}\right)\right]dx\right) \\
&\quad \times \left(\frac{1}{h_T^d}\int e(x)E\left[K\left(\frac{R-x}{h_T}\right)\right]dx\right) \\
&=: A_T - B_T(g)B_T(e).
\end{aligned}$$

Let us derive the asymptotic expansions of these two terms.

SM.8.18.1. *Asymptotic Expansion of B_T*

Let us first consider the second term in (SM.72). We have

$$\begin{aligned}
B_T(g) &= \int\int g(x)\frac{1}{h_T^d}K\left(\frac{u-x}{h_T}\right)f(u)\,du\,dx \\
&= \int\int g(x)K(z)f(x+h_Tz)\,dz\,dx \\
&= \int\int g(x)K(z)[f(x+h_Tz)-f(x)]\,dz\,dx.
\end{aligned}$$

This term can be bounded by $|B_T(g)| \leq \int|g(x)|(\int|K(z)||f(x+h_Tz)-f(x)|\,dz)\,dx$. Since K has a bounded support (Assumption A.2(ii)), and by the mean-value theorem, $\int|K(z)||f(x+h_Tz)-f(x)|\,dz \leq h_T(\int|K(z)||z|\,dz)q(x)$ for large T, where q is defined in Assumption A.4(i), then, by Cauchy–Schwarz inequality and Assumption A.4(i),

$$
\begin{aligned}
|B_T(g)| &\le c h_T \int |g(x)| q(x)\,dx \\
&\le c h_T \left(\int |g(x)|^2 q(x)\,dx \right)^{1/2} \left(\int q(x)\,dx \right)^{1/2}
\end{aligned}
$$

for a constant c. Thus, using that $1\{y \le \varphi_0(x)\} - \tau$ is bounded gives

$$
B_T(g) = O\big(h_T \|\psi_g\|_{L^2(\bar{F}_Z)}\big). \tag{SM.73}
$$

SM.8.18.2. *Asymptotic Expansion of A_T*

Let us now consider the first term in (SM.72). We have

$$
\begin{aligned}
&\frac{1}{h_T^{2d}} E\left[K\left(\frac{R-x}{h_T} \right) K\left(\frac{R-y}{h_T} \right) \right] \\
&\quad = \frac{1}{h_T^d} \int K(u) K\left(u + \frac{x-y}{h_T} \right) f(x + h_T u)\,du.
\end{aligned}
$$

We get

$$
\begin{aligned}
A_T \;=\; & \frac{1}{h_T^d} \int \int \int g(x) e(y) K(u) K\left(u + \frac{x-y}{h_T} \right) f(x)\,du\,dx\,dy \\
& + \frac{1}{h_T^d} \int \int \int g(x) e(y) K(u) K\left(u + \frac{x-y}{h_T} \right) \\
& \times [f(x + h_T u) - f(x)]\,du\,dx\,dy \\
=:\; & \tilde{A}_{1T} + \tilde{A}_{2T}.
\end{aligned}
$$

To rewrite these terms, we have

$$
\begin{aligned}
&\int \frac{1}{h_T^d} K\left(u + \frac{x-y}{h_T} \right) e(y)\,dy \\
&\quad = \int K(u+z) e(x - h_T z) 1_S(x - h_T z)\,dz \\
&\quad = \left(\int K(u+z) 1_S(x - h_T z)\,dz \right) e(x) \\
&\qquad + \int K(u+z)[e(x - h_T z) - e(x)] 1_S(x - h_T z)\,dz.
\end{aligned}
$$

Thus, we get

$$
\begin{aligned}
\tilde{A}_{1T} = &\int g(x)e(x)f(x) \\
&\times \left(\int\int K(u)K(u+z)1_{\mathcal{S}}(x-h_T z)\,dz\,du\right)dx \\
&+ \int\int g(x)f(x)K(u) \\
&\times \int K(u+z)[e(x-h_T z)-e(x)]1_{\mathcal{S}}(x-h_T z)\,dz\,du\,dx \\
= &\operatorname{Cov}[g(R), e(R)] \\
&- \int g(x)e(x)f(x) \\
&\times \left(\int\int K(u)K(u+z)1_{\mathcal{S}^c}(x-h_T z)\,dz\,du\right)dx \\
&+ \int\int g(x)f(x)K(u) \\
&\times \left(\int K(u+z)[e(x-h_T z)-e(x)]1_{\mathcal{S}}(x-h_T z)\,dz\right)du\,dx.
\end{aligned}
$$

Similarly

$$
\begin{aligned}
\tilde{A}_{2T} = &\int g(x)e(x)\int\int K(u)K(u+z)1_{\mathcal{S}}(x-h_T z) \\
&\times [f(x+h_T u)-f(x)]\,dz\,du\,dx \\
&+ \int\int g(x)K(u) \\
&\times \left(\int K(u+z)[e(x-h_T z)-e(x)]1_{\mathcal{S}}(x-h_T z)\,dz\right) \\
&\times [f(x+h_T u)-f(x)]\,du\,dx.
\end{aligned}
$$

We conclude that

$$
\begin{aligned}
A_T = &\operatorname{Cov}[g(R), e(R)] \\
&- \int g(x)e(x)f(x) \\
&\times \left(\int\int K(u)K(u+z)1_{\mathcal{S}^c}(x-h_T z)\,dz\,du\right)dx
\end{aligned}
\tag{SM.74}
$$

$$
\begin{aligned}
&+\int g(x)f(x)\Big(\int\int K(u)K(u+z)[e(x-h_Tz)-e(x)] \\
&\times 1_{\mathcal{S}}(x-h_Tz)\,dz\,du\Big)\,dx \\
&+\int g(x)e(x)\int\int K(u)K(u+z)1_{\mathcal{S}}(x-h_Tz) \\
&\times [f(x+h_Tu)-f(x)]\,dz\,du\,dx \\
&+\int\int g(x)K(u) \\
&\times \Big(\int K(u+z)[e(x-h_Tz)-e(x)]1_{\mathcal{S}}(x-h_Tz)\,dz\Big) \\
&\times [f(x+h_Tu)-f(x)]\,du\,dx
\end{aligned}
$$

$$
=: \operatorname{Cov}[g(R), e(R)] + A_{1,T} + A_{2,T} + A_{3,T} + A_{4,T}. \tag{SM.75}
$$

Let us now bound terms $A_{1,T}$–$A_{4,T}$, separately.

• *Bound of* $A_{1,T}$. We have

$$
|A_{1,T}| \le \int |g(x)||e(x)|f(x) \times \Big(\int \overline{K}(z)1_{\mathcal{S}}(x)1_{\mathcal{S}^c}(x-h_Tz)\,dz\Big)\,dx,
$$

where $\overline{K}(z) := \int |K(u)K(u+z)|\,du$. By Assumption A.2(ii), $\overline{K}$ has bounded support included in $B_\kappa(0)$, $\kappa = 2\sup_{z\in \operatorname{supp}(K)}|z|$. Then $\int \overline{K}(z)1_{\mathcal{S}}(x)1_{\mathcal{S}^c}(x-h_Tz)\,dz \le c1\{x \in \partial\mathcal{S}(\kappa h_T)\}$ for large T, for a constant c. Thus,

$$
\begin{aligned}
|A_{1,T}| &\le c\int |g(x)||e(x)|1\{x\in\partial\mathcal{S}(\kappa h_T)\}f(x)\,dx \\
&\le c\rho(g,\kappa h_T)\rho(e,\kappa h_T)
\end{aligned} \tag{SM.76}
$$

by Cauchy–Schwarz inequality.

• *Bound of* $A_{2,T}$. We have $A_{2,T} = \int g(x)(\int k(z)[e(x-h_Tz)-e(x)]1_{\mathcal{S}}(x-h_Tz)\,dz)f(x)\,dx$, thus $|A_{2,T}| \le \int |g(x)|\Delta e_{h_T}(x)f(x)\,dx$, where $\Delta e_{h_T}(x) := \int |k(z)||e(x-h_Tz)-e(x)|\,dz$ and $k(z) := \int K(u)K(u+z)\,du$. By the Cauchy–Schwarz inequality, we have

$$
|A_{2,T}| \le \|g\|_{L^2(F)}\big\|\Delta e_{h_T}\big\|_{L^2(F)}. \tag{SM.77}
$$

To bound the term $\|\Delta e_{h_T}\|_{L^2(F)}$, we can apply Lemma C.2. Function k is bounded and has compact support by Assumption A.2(i) and (ii). We get

$$\begin{aligned}\|\Delta e_{h_T}\|^2_{L^2(F)} &\le ch_T^2\|\nabla\psi_e\|^2_{L^2(F_Z)} \\ &\quad + ch_T^2\int|\nabla\psi_e(z)|^2\sup_{\zeta\in B_{ch_T}(0)}|f_Z(z+\zeta)-f_Z(z)|\,dz \\ &\quad + c\|\psi_e\|^2_{L^2(F_Z)}\int 1\{|y-\varphi_0(x)|\le c\sqrt{h_T}\}f_W(w)\,dw.\end{aligned}$$

Term $\int 1\{|y-\varphi_0(x)|\le c\sqrt{h_T}\}f_W(w)\,dw$ is $O(\sqrt{h_T})$. Further, from Assumption A.4(i) and using $\sup_{\zeta\in B_{ch_T}(0)}|\nabla_z f_Z(z+\zeta)|\le\int\sup_{\zeta\in B_{ch_T}(0)}|\nabla f_{X,Y,Z}(w,z+\zeta)|\,dw\le\int q(w,z)\,dw$, we deduce $\sup_{\zeta\in B_{ch_T}(0)}|f_Z(z+\zeta)-f_Z(z)|\le ch_T\int q(w,z)\,dw$. Thus, $\|\Delta e_{h_T}\|^2_{L^2(F)}=O(h_T^2\|\nabla\psi_e\|^2_{L^2(F_Z)}+h_T^3\|\nabla\psi_e\|^2_{L^2(\bar F_Z)}+h_T^{1/2}\|\psi_e\|^2_{L^2(F_Z)})$. We conclude from (SM.77) that

$$\begin{aligned}A_{2,T}&=O\big(h_T\|\psi_g\|_{L^2(F_Z)}\|\nabla\psi_e\|_{L^2(F_Z)}+h_T^{3/2}\|\psi_g\|_{L^2(F_Z)}\|\nabla\psi_e\|_{L^2(\bar F_Z)} \\ &\quad + h_T^{1/4}\|\psi_g\|_{L^2(F_Z)}\|\psi_e\|_{L^2(F_Z)}\big).\end{aligned}\tag{SM.78}$$

• *Bound of* $A_{3,T}$. We have

$$\begin{aligned}|A_{3,T}|&\le\left(\int|K(z)|\,dz\right)\int|g(x)||e(x)| \\ &\quad\times\left(\int|K(u)||f(x+h_Tu)-f(x)|\,du\right)dx.\end{aligned}$$

Again, by Assumptions A.2(i) and (ii) and A.4(i), $\int|K(u)||f(x+h_Tu)-f(x)|\,du\le ch_Tq(x)$ for a constant c and large T. Thus,

$$\begin{aligned}|A_{3,T}|&\le\tilde c h_T\int|g(x)||e(x)|q(x)\,dx \\ &\le\tilde c h_T\left(\int g(x)^2q(x)\,dx\right)^{1/2}\left(\int e(x)^2q(x)\,dx\right)^{1/2}\end{aligned}$$

from Cauchy–Schwarz inequality, and using that $1\{y\le\varphi_0(x)\}-\tau$ is bounded, we have

$$A_{3,T}=O\big(h_T\|\psi_g\|_{L^2(\bar F_Z)}\|\psi_e\|_{L^2(\bar F_Z)}\big).\tag{SM.79}$$

• *Bound of* $A_{4,T}$. We have $A_{4,T}=\int K(u)(\int g(x)[\Delta e_{u,h_T}(x)][f(x+h_Tu)-f(x)]\,dx)\,du$, where $\Delta e_{u,h_T}(x)=\int k(z;u)[e(x-h_Tz)-e(x)]1_{\mathcal S}(x-h_Tz)\,dz$

and $k(z; u) = K(u + z)$. Since K has bounded support, then

$$|A_{4,T}| \leq \int |K(u)| \left(\int |g(x)| \left| \Delta e_{h_T}(x) \right| |f(x + h_T u) - f(x)| \, dx \right) du$$
$$= \int |g(x)| \Delta e_{h_T}(x) \pi_T(x) \, dx,$$

where $\Delta e_{h_T}(x) := \int 1\{|z| < c_1\} |e(x - h_T z) - e(x)| \, dz$ and $\pi_T(x) := \int |K(u)| \times |f(x + h_T u) - f(x)| \, du$ for a constant c_1. By Cauchy–Schwarz inequality,

$$|A_{4,T}| \leq \left(\int |g(x)|^2 \pi_T(x) \, dx \right)^{1/2} \left(\int \left| \Delta e_{h_T}(x) \right|^2 \pi_T(x) \, dx \right)^{1/2}. \tag{SM.80}$$

Since $\pi_T(x) \leq c_2 h_T q(x)$ for a constant c_2, we get

$$\int |g(x)|^2 \pi_T(x) \, dx = O\left(h_T \| \psi_g \|^2_{L^2(\bar{F}_Z)} \right). \tag{SM.81}$$

To bound $\int |\Delta e_{h_T}(x)|^2 \pi_T(x) \, dx$, we apply the argument in the proof of Lemma C.2, with $k(z) = 1(|z| < c_1)$. Then

$$\begin{aligned}
&\int \left| \Delta e_{h_T}(u) \right|^2 \pi_T(u) \, du \\
&\quad \leq c h_T^2 \int |\nabla \psi_e(z)|^2 \pi_{T,Z}(z) \, dz \\
&\qquad + c h_T^2 \int |\nabla \psi_e(z)|^2 \sup_{\zeta \in B_{c h_T}(0)} |\pi_{T,Z}(z + \zeta) - \pi_{T,Z}(z)| \, dz \\
&\qquad + c \left(\int |\psi_e(z)|^2 \pi_{T,Z}(z) \, dz \right) \\
&\qquad \times \int 1\left\{ |y - \varphi_0(x)| \leq c \sqrt{h_T} \right\} \pi_{T,W}(w) \, dw,
\end{aligned}$$

where $\pi_{T,Z}(z) := \int \pi_T(w, z) \, dw$ and $\pi_{T,W}(w) := \int \pi_T(w, z) \, dz$. Using that for large T, $\sup_{\zeta \in B_{c h_T}(0)} \pi_{T,Z}(z + \zeta) \leq c_2 h_T \int q(w, z) \, dw$ for some constant c_2, we deduce that

$$\int \left| \Delta e_{h_T}(x) \right|^2 \pi_T(x) \, dx = O\left(h_T^3 \| \nabla \psi_e \|^2_{L^2(\bar{F}_Z)} + h_T^{3/2} \| \psi_e \|^2_{L^2(\bar{F}_Z)} \right). \tag{SM.82}$$

From (SM.80), (SM.81), and (SM.82), we get

$$A_{4,T} = O\left(h_T^2 \| \psi_g \|_{L^2(\bar{F}_Z)} \| \nabla \psi_e \|_{L^2(\bar{F}_Z)} + h_T^{5/4} \| \psi_g \|_{L^2(\bar{F}_Z)} \| \psi_e \|_{L^2(\bar{F}_Z)} \right). \tag{SM.83}$$

From (SM.72), (SM.73), (SM.74), (SM.76), (SM.78), (SM.79), and (SM.83), we derive the asymptotic expansion of $E[V_T(g)V_T(e)]$:

$$\begin{aligned}
&E[V_T(g)V_T(e)] \\
&\quad = \mathrm{Cov}[g(R), e(R)] + O(\rho(g, \kappa h_T)\rho(e, \kappa h_T)) \\
&\qquad + O\big(h_T\|\psi_g\|_{L^2(F_Z)}\|\nabla\psi_e\|_{L^2(F_Z)} + h_T\|\psi_g\|_{L^2(\bar{F}_Z)}\|\psi_e\|_{L^2(\bar{F}_Z)} \\
&\qquad + h_T^{3/2}\|\psi_g\|_{L^2(F_Z)}\|\nabla\psi_e\|_{L^2(\bar{F}_Z)} + h_T^2\|\psi_g\|_{L^2(\bar{F}_Z)}\|\nabla\psi_e\|_{L^2(\bar{F}_Z)} \\
&\qquad + h_T^{1/4}\|\psi_g\|_{L^2(F_Z)}\|\psi_e\|_{L^2(F_Z)} + h_T^{5/4}\|\psi_g\|_{L^2(\bar{F}_Z)}\|\psi_e\|_{L^2(\bar{F}_Z)}\big).
\end{aligned}$$

SM.8.19. *Proof of Lemma C.4*

Let us apply Lemma C.3 with $u = (w, z)$ and $g(u) = e(u) = (1/\sqrt{\nu_j})(A\phi_j) \times (z)\frac{1\{y\le\varphi_0(x)\}-\tau}{\tau(1-\tau)} =: g_j(u)$ for any $j \in \mathbb{N}$. We have $V[g_j(R)] = \frac{1}{\nu_j}\frac{1}{[\tau(1-\tau)]^2}E[(A\phi_j) \times (Z)^2 E[(1\{Y \le \varphi_0(X)\} - \tau)^2 \mid Z]] = \frac{1}{\nu_j}\frac{1}{\tau(1-\tau)}E[(A\phi_j)(Z)^2] = \frac{1}{\nu_j}\langle\phi_j, A^*A \times \phi_j\rangle_H = 1$. Thus, we get

$$E[Z_{j,T}^2] = 1 + O(\rho(g_j, \kappa h_T)^2) + O\big(h_T^{1/4}\big(\|\psi_j\|_{H^2(F_Z)} + \|\psi_j\|_{H^2(\bar{F}_Z)}\big)^2\big).$$

Let us now bound the terms in the RHS. We have $\rho(g_j, \kappa h_T)^2 = E[g_j(R)^2 1\{R \in \partial\mathcal{S}(\kappa h_T)\}] \le \frac{1}{[\tau(1-\tau)]^2}E[|\psi_j(Z)|^{\bar{s}}]^{2/\bar{s}}P[R \in \partial\mathcal{S}(\kappa h_T)]^{1-2/\bar{s}} = O(h_T^{1-2/\bar{s}})$ from Assumption A.5(ii). Using $\bar{s} \ge 4$, we get $\rho(g_j, \kappa h_T)^2 = O(h_T^{1/2})$. Moreover, $\|\psi_j\|^2_{L^2(\bar{F}_Z)} = \frac{\int\int\psi_j(z)^2 q(w,z)\,dw\,dz}{\int\int q(w,z)\,dw\,dz}$ and by the Cauchy–Schwarz inequality,

$$\int\int\psi_j(z)^2 q(w, z)\,dw\,dz \le E\big[|\psi_j(z)|^4\big]^{1/2}\bigg(\int\frac{q(s)^2}{f_{X,Y,Z}(s)}\,ds\bigg)^{1/2}.$$

From Assumptions A.4(i) and A.5(ii), it follows that $\|\psi_j\|^2_{L^2(\bar{F}_Z)} = O(1)$ uniformly in $j \in \mathbb{N}$. Similarly, $\|\nabla\psi_j\|^2_{L^2(\bar{F}_Z)} \le \frac{E[|\nabla\psi_j(Z)|^4]^{1/2}}{\int\int q(w,z)\,dw\,dz}(\int\frac{q(s)^2}{f_{X,Y,Z}(s)}\,ds)^{1/2} = O(j^2)$ by Assumptions A.4(i) and A.5(iii). The conclusion follows.

SM.8.20. *Proof of Lemma C.5*

The proof of Lemma C.5 is given in the technical report of Gagliardini and Scaillet (2012, GS)).

SM.8.21. *Proof of Lemma C.6*

From the proof of Lemma B.8(i), we have $(\lambda_T + A^*A)^{-1}A^*(\hat{\zeta} - E\hat{\zeta}) = \sum_{j=1}^{\infty} c_{j,T}Z_{j,T}\phi_j$, where $c_{j,T} := \frac{1}{\sqrt{T}}\frac{\sqrt{\nu_j}}{\lambda_T+\nu_j}$ and $Z_{j,T}$ are defined in (SM.49) and

(SM.50). Using $\|(\lambda_T + A^*A)^{-1}A^*(\hat{\psi} - E\hat{\psi})\|^2 = \sum_{j,l=1}^{\infty} c_{j,T}c_{l,T}Z_{j,T}Z_{l,T}\langle\phi_j, \phi_l\rangle$, we get

$$E\big[\|(\lambda_T + A^*A)^{-1}A^*(\hat{\psi} - E\hat{\psi})\|^4\big] = \sum_{j,l,m,n=1}^{\infty} c_{j,T}c_{l,T}c_{m,T}c_{n,T}E[Z_{j,T}Z_{l,T}Z_{m,T}Z_{n,T}]\langle\phi_j, \phi_l\rangle\langle\phi_m, \phi_n\rangle.$$

Let us now bound the expectation terms. By twice applying the Cauchy–Schwarz inequality, we get $|E[Z_{j,T}Z_{l,T}Z_{m,T}Z_{n,T}]| \le E[Z_{j,T}^4]^{1/4}E[Z_{l,T}^4]^{1/4} \times E[Z_{m,T}^4]^{1/4}E[Z_{n,T}^4]^{1/4}$. Let $\bar{c}_{j,T} := c_{j,T}E[Z_{j,T}^4]^{1/4}$. Then we get

$$\begin{aligned}
&E\big[\|(\lambda_T + A^*A)^{-1}A^*(\hat{\psi} - E\hat{\psi})\|^4\big] \\
&\le \sum_{j,l,m,n=1}^{\infty} \bar{c}_{j,T}\bar{c}_{l,T}\bar{c}_{m,T}\bar{c}_{n,T}|\langle\phi_j, \phi_l\rangle||\langle\phi_m, \phi_n\rangle| \\
&= \sum_{j\in\mathbb{N}} \bar{c}_{j,T}^4\|\phi_j\|^4 + 4\sum_{(j,l)\in\mathbb{D}^2} \bar{c}_{j,T}^3\bar{c}_{l,T}\|\phi_j\|^2|\langle\phi_j, \phi_l\rangle| \\
&\quad + \sum_{(j,l)\in\mathbb{D}^2} \bar{c}_{j,T}^2\bar{c}_{l,T}^2\big(\|\phi_j\|^2\|\phi_l\|^2 + 2|\langle\phi_j, \phi_l\rangle|^2\big) \\
&\quad + 2\sum_{(j,l,m)\in\mathbb{D}^3} \bar{c}_{j,T}^2\bar{c}_{l,T}\bar{c}_{m,T}\|\phi_j\|^2|\langle\phi_l, \phi_m\rangle| \\
&\quad + 4\sum_{(j,l,m)\in\mathbb{D}^3} \bar{c}_{j,T}^2\bar{c}_{l,T}\bar{c}_{m,T}|\langle\phi_j, \phi_l\rangle||\langle\phi_j, \phi_m\rangle| \\
&\quad + \sum_{(j,l,m,n)\in\mathbb{D}^4} \bar{c}_{j,T}\bar{c}_{l,T}\bar{c}_{m,T}\bar{c}_{n,T}|\langle\phi_j, \phi_l\rangle||\langle\phi_m, \phi_n\rangle| \\
&=: J_1 + 4J_2 + J_3 + 2J_4 + 4J_5 + J_6,
\end{aligned}$$

where $\mathbb{D}^d$ denotes the set of d-tuples, which consist of d different natural numbers. Let us now bound separately the different terms.

• *Bound of* J_1. We have

$$J_1 \le \sum_{j,l=1}^{\infty} \bar{c}_{j,T}^2\bar{c}_{l,T}^2\|\phi_j\|^2\|\phi_l\|^2 = \left(\sum_{j=1}^{\infty} \bar{c}_{j,T}^2\|\phi_j\|^2\right)^2 = \bar{q}_T^4,$$

where we denote $\bar{q}_T^2 := \sum_{j=1}^{\infty} \bar{c}_{j,T}^2\|\phi_j\|^2 = \frac{1}{T}\sum_{j=1}^{\infty} \frac{\nu_j}{(\lambda_T+\nu_j)^2}\|\phi_j\|^2 E[Z_{j,T}^4]^{1/2}$.

• *Bound of* J_2. Using $\bar{c}_{j,T}^2 \|\phi_j\|^2 \le \bar{q}_T^2$, $j \in \mathbb{N}$, we get

$$J_2 \le \bar{q}_T^2 \sum_{(j,l)\in\mathbb{D}^2} \bar{c}_{j,T}\bar{c}_{l,T}|\langle\phi_j, \phi_l\rangle|.$$

Let us consider the term

$$\sum_{(j,l)\in\mathbb{D}^2} \bar{c}_{j,T}\bar{c}_{l,T}|\langle\phi_j, \phi_l\rangle| = \sum_{j=1}^{\infty} \bar{c}_{j,T}\|\phi_j\|\left(\sum_{l:l\neq j} \bar{c}_{l,T}\|\phi_l\|\frac{|\langle\phi_j, \phi_l\rangle|}{\|\phi_j\|\|\phi_l\|}\right).$$

Using Cauchy–Schwarz inequality, for any j, we have

$$\begin{aligned}\sum_{l:l\neq j} \bar{c}_{l,T}\|\phi_l\|\frac{|\langle\phi_j, \phi_l\rangle|}{\|\phi_j\|\|\phi_l\|} &\le \left(\sum_{l:l\neq j}\bar{c}_{l,T}^2\|\phi_l\|^2\right)^{1/2}\left(\sum_{l:l\neq j}\frac{\langle\phi_j, \phi_l\rangle^2}{\|\phi_j\|^2\|\phi_l\|^2}\right)^{1/2} \\ &\le \bar{q}_T\left(\sum_{l:l\neq j}\frac{\langle\phi_j, \phi_l\rangle^2}{\|\phi_j\|^2\|\phi_l\|^2}\right)^{1/2}.\end{aligned}$$

Thus, we get again, by Cauchy–Schwarz inequality,

$$\begin{aligned}&\sum_{(j,l)\in\mathbb{D}^2} \bar{c}_{j,T}\bar{c}_{l,T}|\langle\phi_j, \phi_l\rangle| \\ &\quad\le \bar{q}_T\sum_{j=1}^{\infty}\bar{c}_{j,T}\|\phi_j\|\left(\sum_{l:l\neq j}\frac{\langle\phi_j, \phi_l\rangle^2}{\|\phi_j\|^2\|\phi_l\|^2}\right)^{1/2} \\ &\quad\le \bar{q}_T\left(\sum_{j=1}^{\infty}\bar{c}_{j,T}^2\|\phi_j\|^2\right)^{1/2}\left(\sum_{j=1}^{\infty}\sum_{l:l\neq j}\frac{\langle\phi_j, \phi_l\rangle^2}{\|\phi_j\|^2\|\phi_l\|^2}\right)^{1/2} = \bar{q}_T^2\bar{\rho},\end{aligned} \tag{SM.84}$$

where $\bar{\rho} := (\sum_{j,l=1:l\neq j}^{\infty}\frac{\langle\phi_j,\phi_l\rangle^2}{\|\phi_j\|^2\|\phi_l\|^2})^{1/2} < \infty$ by Assumption A.5(i). We deduce $J_2 \le \bar{q}_T^4\bar{\rho}$.

• *Bound of* J_3. We have $J_3 \le 3\sum_{(j,l)\in\mathbb{D}^2}\bar{c}_{j,T}^2\bar{c}_{l,T}^2\|\phi_j\|^2\|\phi_l\|^2 \le 3\bar{q}_T^4$.

• *Bound of* J_4. We have $J_4 \le (\sum_j \bar{c}_{j,T}^2\|\phi_j\|^2)\sum_{(l,m)\in\mathbb{D}^2}\bar{c}_{l,T}\bar{c}_{m,T}|\langle\phi_l, \phi_m\rangle| \le \bar{q}_T^4\bar{\rho}$, using (SM.84).

• *Bound of* J_5. We have

$$\begin{aligned}J_5 &\le \sum_j \bar{c}_{j,T}^2\|\phi_j\|^2\left(\sum_{l:l\neq j}\sum_{m:m\neq j}\bar{c}_{l,T}\|\phi_l\|\frac{|\langle\phi_j, \phi_l\rangle|}{\|\phi_j\|\|\phi_l\|}\bar{c}_{m,T}\|\phi_m\|\frac{|\langle\phi_j, \phi_m\rangle|}{\|\phi_j\|\|\phi_m\|}\right) \\ &= \sum_j \bar{c}_{j,T}^2\|\phi_j\|^2\left(\sum_{l:l\neq j}\bar{c}_{l,T}\|\phi_l\|\frac{|\langle\phi_j, \phi_l\rangle|}{\|\phi_j\|\|\phi_l\|}\right)^2.\end{aligned}$$

Using $\bar{c}_{j,T}^2 \|\phi_j\|^2 \le \bar{q}_T^2$ for all $j \in \mathbb{N}$ and

$$\sum_{l:l\neq j} \bar{c}_{l,T} \|\phi_l\| \frac{|\langle \phi_j, \phi_l \rangle|}{\|\phi_j\| \|\phi_l\|} \le \bar{q}_T \left(\sum_{l:l\neq j} \frac{\langle \phi_j, \phi_l \rangle^2}{\|\phi_j\|^2 \|\phi_l\|^2} \right)^{1/2},$$

we get $J_5 \le \bar{q}_T^4 \sum_j \sum_{l:l\neq j} \frac{\langle \phi_j, \phi_l \rangle^2}{\|\phi_j\|^2 \|\phi_l\|^2} = \bar{q}_T^4 \bar{\rho}^2$.

• *Bound of* J_6. Finally, $J_6 \le \sum_{(j,l)\in \mathbb{D}^2, (m,n)\in \mathbb{D}^2} \bar{c}_{j,T} \bar{c}_{l,T} \bar{c}_{m,T} \bar{c}_{n,T} |\langle \phi_j, \phi_l \rangle| |\langle \phi_m, \phi_n \rangle| = (\sum_{(j,l)\in \mathbb{D}^2} \bar{c}_{j,T} \bar{c}_{l,T} |\langle \phi_j, \phi_l \rangle|)^2 \le \bar{q}_T^4 \bar{\rho}^2$, using (SM.84).

To summarize, we have proved $E[\|(\lambda_T + A^* A)^{-1} A^* (\hat{\psi} - E\hat{\psi})\|^4] \le C \bar{q}_T^4$. Using a result similar to Lemma C.4, and Assumptions A.4(i), A.5(ii) and (iii), we can show that $\bar{q}_T^2 = O(\frac{1}{T} \sum_{j=1}^\infty \frac{\nu_j}{(\lambda_T + \nu_j)^2} \|\phi_j\|^2)$. Since $E[\|(\lambda_T + A^* A)^{-1} A^* (\hat{\psi} - E\hat{\psi})\|^2] = O(\frac{1}{T} \sum_{j=1}^\infty \frac{\nu_j}{(\lambda_T + \nu_j)^2} \|\phi_j\|^2)$ by Lemma B.8(i), the conclusion follows.

SM.8.22. *Proof of Lemma C.7*

Let $1_{\varphi_0}(w) = 1\{y \le \varphi_0(x)\} - \tau$. We have

$$\begin{aligned}
\text{(SM.85)}\quad & -(\tilde{\hat{A}}_0(\hat{\zeta} - \hat{q}) - \tilde{A}\hat{\zeta})(x) \\
&= \frac{1}{T\tau(1-\tau)} \sum_{t=1}^T \frac{\hat{f}_{X,Y,Z}(x, \varphi_0(x), Z_t)}{\hat{f}_Z(Z_t)^2} \int 1_{\varphi_0}(w) \hat{f}_{X,Y,Z}(w, Z_t)\, dw \\
&\quad - \frac{1}{\tau(1-\tau)} \int f_{X,Y|Z}(x, \varphi_0(x)|z) \int 1_{\varphi_0}(w) \hat{f}_{X,Y,Z}(w, z)\, dw\, dz \\
&= \frac{1}{T\tau(1-\tau)} \sum_{t=1}^T \Delta \hat{f}_{X,Y,Z}(x, \varphi_0(x), Z_t) \\
&\quad \times \frac{1}{\hat{f}_Z(Z_t)^2} \int 1_{\varphi_0}(w) \Delta \hat{f}_{X,Y,Z}(w, Z_t)\, dw \\
&\quad + \left\{ \frac{1}{T\tau(1-\tau)} \sum_{t=1}^T f_{X,Y|Z}(x, \varphi_0(x)|Z_t) \right. \\
&\quad \times \frac{1}{f_Z(Z_t)} \int 1_{\varphi_0}(w) \Delta \hat{f}_{X,Y,Z}(w, Z_t)\, dw \\
&\quad \left. - \frac{1}{\tau(1-\tau)} \int f_{X,Y|Z}(x, \varphi_0(x)|z) \int 1_{\varphi_0}(w) \Delta \hat{f}_{X,Y,Z}(w, z)\, dw\, dz \right\} \\
&\quad - \frac{1}{T\tau(1-\tau)} \sum_{t=1}^T f_{X,Y,Z}(x, \varphi_0(x), Z_t) \frac{\Delta \hat{f}_Z(Z_t)[\hat{f}_Z(Z_t) + f_Z(Z_t)]}{\hat{f}_Z(Z_t)^2 f_Z(Z_t)^2}
\end{aligned}$$

$$\times \int 1_{\varphi_0}(w)\Delta\hat{f}_{X,Y,Z}(w, Z_t)\,dw$$
$$=: I_1(x) + I_2(x) + I_3(x).$$

We consider in detail term $I_2(x)$; the analysis of the other terms is similar.

Write

$$\begin{aligned}
I_2(x) &= \frac{T-1}{T^2}\sum_{t=1}^{T}\Bigg\{f_{X,Y,Z}(x,\varphi_0(x),Z_t)\frac{\Sigma_0(Z_t)}{f_Z(Z_t)} \\
&\quad \times \int 1_{\varphi_0}(w)\bar{f}(w,Z_t)\,dw \\
&\quad - \int f_{X,Y,Z}(x,\varphi_0(x),z)\Sigma_0(z)\int 1_{\varphi_0}(w)\bar{f}(w,z)\,dw\,dz\Bigg\} \\
&\quad + \frac{T-1}{T^2}\sum_{t=1}^{T}\Bigg\{f_{X,Y,Z}(x,\varphi_0(x),Z_t)\frac{\Sigma_0(Z_t)}{f_Z(Z_t)} \\
&\quad \times \int 1_{\varphi_0}(w)b(w,Z_t)\,dw \\
&\quad - \int f_{X,Y,Z}(x,\varphi_0(x),z)\Sigma_0(z)\int 1_{\varphi_0}(w)b(w,z)\,dw\,dz\Bigg\} \\
&\quad + \frac{K(0)}{T^2h_T^{d_Z}}\sum_{t=1}^{T}\Bigg\{f_{X,Y,Z}(x,\varphi_0(x),Z_t)\frac{\Sigma_0(Z_t)}{f_Z(Z_t)} \\
&\quad - \int f_{X,Y,Z}(x,\varphi_0(x),z)\Sigma_0(z)\,dz\Bigg\} \\
&\quad \times \int 1_{\varphi_0}(w)K_h(w-W_t)\,dw \\
&=: I_{21}(x) + I_{22}(x) + I_{23}(x),
\end{aligned}$$

where $\Sigma_0(z) := 1/(f_Z(z)\tau(1-\tau))$, $\bar{f}(w, Z_t) := \hat{f}_{X,Y,Z,-t}(w, Z_t) - E[\hat{f}_{X,Y,Z,-t}(w, Z_t)|Z_t]$, $b(w, Z_t) := E[\hat{f}_{X,Y,Z,-t}(w, Z_t)|Z_t] - f_{X,Y,Z}(w, Z_t)$, and $\hat{f}_{X,Y,Z,-t}$ is the kernel density estimator based on the sample of $T-1$ observations excluding (W_t, Z_t). Note that $b(w, Z_t)$ is equal to the bias function $E[\Delta\hat{f}_{X,Y,Z}(w, z)]$ evaluated at $z = Z_t$. Let us first consider I_{21}. Using $\bar{f}(w, z) = \frac{1}{T-1}\sum_{s=1,s\neq t}^{T}\kappa_s(w, z)$, with $\kappa_s(w, z) := K_h(w - W_s)K_h(z - Z_s) - E[K_h(w - W)K_h(z - Z)]$, $K_h(u) :=$

$\frac{1}{h}K(\frac{u}{h})$, we have

$$I_{21}(x) = \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}\biggl\{a(x,\varphi_0(x),Z_t)\int \kappa_s(w,Z_t)1_{\varphi_0}(w)\,dw - \int a(x,\varphi_0(x),z)\int \kappa_s(w,z)1_{\varphi_0}(w)f_Z(z)\,dw\,dz\biggr\} =: \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T} U_{st}(x),$$

where $a(x,\varphi_0(x),z) := f_{X,Y|Z}(x,\varphi_0(x)|z)\Sigma_0(z)$. Variables $U_{st}(x)$ satisfy $E[U_{st}(x)|W_s,Z_s] = E[U_{st}(x)|W_t,Z_t] = 0$ for $s \neq t$. By symmetrization, we have

$$I_{21}(x) = \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s<t}\tilde{U}_{st}(x),$$

where $\tilde{U}_{st}(x) := U_{st}(x) + U_{ts}(x)$ and $E[\tilde{U}_{st}(x)|W_s,Z_s] = E[\tilde{U}_{st}(x)|W_t,Z_t] = 0$ for $s < t$. We get

$$E\biggl[\int I_{21}(x)^2\,dx\biggr] = \frac{1}{T^4}\sum_{t=1}^{T}\sum_{s<t}^{T}\int E[\tilde{U}_{st}(x)^2]\,dx \leq \frac{2}{T^2}\int E[U_{st}(x)^2]\,dx.$$

The term $E[(\int K_h(w - W_s)1_{\varphi_0}(w)\,dw)^2 K_h(Z_t - Z_s)^2 a(x,\varphi_0(x),Z_t)^2]$ in $E[U_{st}(x)^2]$ is dominant and is $O(h_T^{-d_Z})$ uniformly in $x \in [0,1]$ by Assumption A.4(ii). We get $E[\|I_{21}\|^2]^{1/2} = O(\frac{1}{Th_T^{d_Z/2}}) = o(\lambda_T b(\lambda_T))$ from the conditions on η and γ.

Let us now consider I_{22}. Since the observations are i.i.d. and the bias term $b(w,Z_t)$ is $O(h_T^m)$, we get $E[\|I_{22}\|^2]^{1/2} = O(\frac{1}{T^{1/2}}h_T^m) = o(\lambda_T b(\lambda_T))$.

Finally, to bound I_{23}, we use $|\int 1_{\varphi_0}(w)K_{h_T}(W_t - w)\,dw| \leq \int |1_{\varphi_0}(W_t - h_T v)||K(v)|\,dv \leq 2\|K\|_\infty$. Then we get

$$|I_{23}(x)| \leq \frac{C}{T^2 h_T^{d_Z}}\sum_{t=1}^{T}\bigl(\Sigma_0(Z_t) + E[\Sigma_0(Z_t)]\bigr)$$

by Assumption A.4(ii). By Cauchy–Schwarz inequality, it follows that

$$E[\|I_{23}\|^2]^{1/2} \leq \frac{C}{Th_T^{d_Z}}E[\Sigma_0(Z)^2]^{1/2}.$$

Since $\inf_{z\in\mathcal{Z}} f_Z(z) > 0$, we conclude that $E[\|I_{23}\|^2]^{1/2} = O(\frac{\log T}{T h_T^{d_Z}}) = o(\lambda_T b(\lambda_T))$.

SM.8.23. *Proof of Lemma C.8*

We have

$$
\begin{aligned}
&(\tilde{\hat{A}}_0 \hat{A}_0 \varphi)(x) \\
&\quad = \int \left(\frac{1}{T\tau(1-\tau)} \sum_{t=1}^{T} \frac{\hat{f}_{X,Y,Z}(x, \varphi_0(x), Z_t)\hat{f}_{X,Y,Z}(\xi, \varphi_0(\xi), Z_t)}{\hat{f}_Z(Z_t)^2} \right) \\
&\qquad \times \varphi(\xi)\, d\xi
\end{aligned}
$$

and

$$
\begin{aligned}
&(\tilde{A} A \varphi)(x) \\
&\quad = \frac{1}{\tau(1-\tau)} \int \left(\int \frac{f_{X,Y,Z}(x, \varphi_0(x), z) f_{X,Y,Z}(\xi, \varphi_0(\xi), z)}{f_Z(z)^2} f_Z(z)\, dz \right) \\
&\qquad \times \varphi(\xi)\, d\xi.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&\|\tilde{\hat{A}}_0 \hat{A}_0 - \tilde{A} A\|^2 \\
&\quad \le \frac{1}{[\tau(1-\tau)]^2} \\
&\qquad \times \int_{\mathcal{X}} \int_{\mathcal{X}} \left[\frac{1}{T} \sum_{t=1}^{T} \frac{\hat{f}_{X,Y,Z}(x, \varphi_0(x), Z_t)\hat{f}_{X,Y,Z}(\xi, \varphi_0(\xi), Z_t)}{\hat{f}_Z(Z_t)^2} \right. \\
&\qquad \left. - \int \frac{f_{X,Y,Z}(x, \varphi_0(x), z) f_{X,Y,Z}(\xi, \varphi_0(\xi), z)}{f_Z(z)^2} f_Z(z)\, dz \right]^2 dx\, d\xi \\
&\quad =: \int_{\mathcal{X}} \int_{\mathcal{X}} \hat{I}(x, \xi)^2\, dx\, d\xi =: \|\hat{I}\|^2_{L^2([0,1]^2)}, \qquad P\text{-a.s.}
\end{aligned}
$$

We decompose $\hat{I}(x, \xi)$ as

$$
\begin{aligned}
\hat{I}(x, \xi) \;=\; & \frac{1}{T\tau(1-\tau)} \sum_{t=1}^{T} \frac{\hat{f}_{X,Y,Z}(x, \varphi_0(x), Z_t)\hat{f}_{X,Y,Z}(\xi, \varphi_0(\xi), Z_t)}{\hat{f}_Z(Z_t)^2} \\
& - \int \hat{f}_{X,Y,Z}(x, \varphi_0(x), z)\hat{f}_{X,Y,Z}(\xi, \varphi_0(\xi), z) \Sigma_0(z)\, dz
\end{aligned}
$$

$$+\int \hat{f}_{X,Y,Z}(x,\varphi_0(x),z)\hat{f}_{X,Y,Z}(\xi,\varphi_0(\xi),z)\Sigma_0(z)\,dz$$
$$-\int f_{X,Y,Z}(x,\varphi_0(x),z)f_{X,Y,Z}(\xi,\varphi_0(\xi),z)\Sigma_0(z)\,dz$$
$$=: \hat{I}_1(x,\xi)+\hat{I}_2(x,\xi),$$

where Σ_0 is defined as in the proof of Lemma C.7. Then

$$\|\tilde{\hat{A}}_0\hat{A}_0-\tilde{A}A\| \le \sum_{j=1}^{2}\|\hat{I}_j\|_{L^2([0,1]^2)}, \quad P\text{-a.s.},$$

and

$$E[\|\tilde{\hat{A}}_0\hat{A}_0-\tilde{A}A\|^{2\zeta}]^{1/(2\zeta)} \le \sum_{j=1}^{2}E\big[\|\hat{I}_j\|^{2\zeta}_{L^2([0,1]^2)}\big]^{1/(2\zeta)}.$$

Moreover, for any j, $E[\|\hat{I}_j\|^{2\zeta}_{L^2([0,1]^2)}]=\int E[\hat{I}_j(r_1)^2\cdots\hat{I}_j(r_\zeta)^2]\,dr_1\cdots dr_\zeta$, where $r:=(x,\xi)$, and by repeatedly applying the Cauchy–Schwarz inequality, we get

$$\begin{aligned}
&E[\hat{I}_j(r_1)^2\hat{I}_j(r_2)^2\cdots\hat{I}_j(r_\zeta)^2]\\
&\quad\le E[\hat{I}_j(r_1)^4]^{1/2}E[\hat{I}_j(r_2)^4\cdots\hat{I}_j(r_\zeta)^4]^{1/2}\\
&\quad\le E[\hat{I}_j(r_1)^4]^{1/2}E[\hat{I}_j(r_2)^8]^{1/4}\cdots E\big[\hat{I}_j(r_\zeta)^{2^\zeta}\big]^{1/(2^{\zeta-1})}\\
&\quad\le \prod_{l=1}^{\zeta}E\big[\hat{I}_j(r_l)^{2^\zeta}\big]^{1/(2^{\zeta-1})}.
\end{aligned}$$

Thus, we get $E[\|\tilde{\hat{A}}_0\hat{A}_0-\tilde{A}A\|^{2\zeta}]\le(\sum_{j=1}^{2}(\int_{\mathcal{X}}\int_{\mathcal{X}}E[\hat{I}_j(x,\xi)^{2\bar{\zeta}}]^{1/\bar{\zeta}}\,dx\,d\xi)^{1/2})^{2\zeta}$, with $\bar{\zeta}=2^{\zeta-1}$. We deduce that the lemma is proved if we show that

$$\text{(SM.86)}\quad \int_{\mathcal{X}}\int_{\mathcal{X}}E[\hat{I}_j(x,\xi)^{2N}]\,dx\,d\xi=O(a_T^N)\quad\text{for any } N\in\mathbb{N}$$

and any $j=1,2$. This is implied by $E[\hat{I}_j(x,\xi)^{2N}]=O(a_T^N)$ uniformly in $x,\xi\in[0,1]$. Let us prove (SM.86) for the different terms.

• *Bound of* $\hat{I}_1$. Write

$$\hat{I}_1(x,\xi) = \frac{1}{T\tau(1-\tau)}\sum_{t=1}^{T}\frac{f_{X,Y,Z}(x,\varphi_0(x),Z_t)f_{X,Y,Z}(\xi,\varphi_0(\xi),Z_t)}{\hat{f}_Z(Z_t)^2}$$
$$-\int f_{X,Y,Z}(x,\varphi_0(x),z)f_{X,Y,Z}(\xi,\varphi_0(\xi),z)\Sigma_0(z)\,dz$$

$$
\begin{aligned}
&+\frac{1}{T\tau(1-\tau)} \\
&\times \sum_{t=1}^{T} \frac{\Delta\hat{f}_{X,Y,Z}(x,\varphi_0(x),Z_t) f_{X,Y,Z}(\xi,\varphi_0(\xi),Z_t)}{\hat{f}_Z(Z_t)^2} \\
&- \int \Delta\hat{f}_{X,Y,Z}(x,\varphi_0(x),z) f_{X,Y,Z}(\xi,\varphi_0(\xi),z)\Sigma_0(z)\,dz \\
&+\frac{1}{T\tau(1-\tau)} \\
&\times \sum_{t=1}^{T} \frac{f_{X,Y,Z}(x,\varphi_0(x),Z_t)\Delta\hat{f}_{X,Y,Z}(\xi,\varphi_0(\xi),Z_t)}{\hat{f}_Z(Z_t)^2} \\
&- \int f_{X,Y,Z}(x,\varphi_0(x),z)\Delta\hat{f}_{X,Y,Z}(\xi,\varphi_0(\xi),z)\Sigma_0(z)\,dz \\
&+\frac{1}{T\tau(1-\tau)} \\
&\times \sum_{t=1}^{T} \frac{\Delta\hat{f}_{X,Y,Z}(x,\varphi_0(x),Z_t)\Delta\hat{f}_{X,Y,Z}(\xi,\varphi_0(\xi),Z_t)}{\hat{f}_Z(Z_t)^2} \\
&- \int \Delta\hat{f}_{X,Y,Z}(x,\varphi_0(x),Z_t)\Delta\hat{f}_{X,Y,Z}(\xi,\varphi_0(\xi),Z_t)\Sigma_0(z)\,dz \\
=:& \sum_{j=1}^{4} \hat{I}_{1j}(x,\xi).
\end{aligned}
$$

We focus on the second term $\hat{I}_{12}$. By controlling term $\hat{f}_Z(Z_t)$ in the denominator, we see that the dominant component is

$$
\begin{aligned}
\hat{I}_{121}(x,\xi) :=& \frac{1}{T}\sum_{t=1}^{T} \frac{\Delta\hat{f}_{X,Y,Z}(x,\varphi_0(x),Z_t) f_{X,Y,Z}(\xi,\varphi_0(\xi),Z_t)}{f_Z(Z_t)}\Sigma_0(Z_t) \\
&- \int \Delta\hat{f}_{X,Y,Z}(x,\varphi_0(x),z) f_{X,Y,Z}(\xi,\varphi_0(\xi),z)\Sigma_0(z)\,dz \\
=& \frac{T-1}{T^2}\sum_{t=1}^{T}\Bigg(\frac{\bar{f}(x,\varphi_0(x),Z_t) f_{X,Y,Z}(\xi,\varphi_0(\xi),Z_t)}{f_Z(Z_t)}\Sigma_0(Z_t) \\
&- \int \bar{f}(x,\varphi_0(x),z) f_{X,Y,Z}(\xi,\varphi_0(\xi),z)\Sigma_0(z)\,dz\Bigg)
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{T-1}{T^2}\sum_{t=1}^{T}\biggl(\frac{b(x,\varphi_0(x),Z_t)f_{X,Y,Z}(\xi,\varphi_0(\xi),Z_t)}{f_Z(Z_t)}\Sigma_0(Z_t)\\
&-\int b(x,\varphi_0(x),z)f_{X,Y,Z}(\xi,\varphi_0(\xi),z)\Sigma_0(z)\,dz\biggr)\\
&+\frac{1}{Th_T^{d_Z}}K(0)\frac{1}{T}\sum_{t=1}^{T}K_h(x-X_t)K_h(\varphi_0(x)-Y_t)\\
&\times\biggl(\frac{f_{X,Y,Z}(\xi,\varphi_0(\xi),Z_t)}{f_Z(Z_t)}\Sigma_0(Z_t)\\
&-\int K_h(z-Z_t)f_{X,Y,Z}(\xi,\varphi_0(\xi),z)\Sigma_0(z)\,dz\biggr),
\end{aligned}
$$

with the notation introduced in the proof of Lemma C.7. We consider in detail the first term $\hat{I}_{1211}(x,\xi)$ (the second and the third terms can be bounded along the lines of terms I_{22} and I_{23} in the proof of Lemma C.7). Write $\bar{f}(x,\varphi_0(x),z)=\frac{1}{T-1}\sum_{s=1,s\neq t}^{T}\boldsymbol{\kappa}_s(x,z)$, where $\boldsymbol{\kappa}_s(x,z):=K_h(x-X_s)K_h(\varphi_0(x)-Y_s)K_h(z-Z_s)-E[K_h(x-X)K_h(\varphi_0(x)-Y)K_h(z-Z)]$. Then

$$
\begin{aligned}
&\hat{I}_{1211}(x,\xi)\\
&\quad=\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}\biggl[\boldsymbol{\kappa}_s(x,Z_t)a(\xi,Z_t)-\int\boldsymbol{\kappa}_s(x,z)a(\xi,z)f_Z(z)\,dz\biggr]\\
&\quad=:\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1,s\neq t}^{T}U_{st},
\end{aligned}
$$

where $a(\xi,z)=f_{X,Y|Z}(\xi,\varphi_0(\xi)|z)\Sigma_0(z)$. We get, for $N\in\mathbb{N}$,

$$
\begin{aligned}
&E[\hat{I}_{1211}(x,\xi)^{2N}]\\
&\quad=\frac{1}{T^{4N}}\sum_{t_1=1}^{T}\sum_{s_1=1,s_1\neq t_1}^{T}\cdots\sum_{t_{2N}=1}^{T}\sum_{s_{2N}=1,s_{2N}\neq t_{2N}}^{T}E\bigl[U_{s_1t_1}\cdots U_{s_{2N}t_{2N}}\bigr].
\end{aligned}
$$

Variables U_{st} satisfy $E[U_{st}|X_s,Z_s]=E[U_{st}|X_t,Z_t]=0$ for $s\neq t$ and any $x,\xi\in[0,1]$. Thus, the expectation $E[U_{s_1t_1}\cdots U_{s_{2N}t_{2N}}]$ can be different from zero only if the ordered $2N$-tuple $(s_1,t_1,\dots,s_{2N},t_{2N})$ is such that each element has multiplicity at least 2. Let us first bound the number $C(N,T)$ of such $2N$-tuples. We have $C(N,T)\leq\sum_{j=1}^{2N}C_j(N,T)$, where $C_j(N,T)$ is the number of $2N$-tuples with j different elements. Moreover, $C_j(N,T)=c_j(N)T(T-1)\cdots(T-j+1)$, where $c_j(N)$ is the number of ways in which we can build j groups using $4N$ different elements. Thus, $C(N,T)=O(T^{2N})$. Further, the dominant term in $E[U_{s_1t_1}\cdots U_{s_{2N}t_{2N}}]$ is $E[K_h(x-X_{s_1})K_h(\varphi_0(x)-Y_{s_1})K_h(Z_{t_1}-Z_{s_1})a(\xi,Z_{t_1})\cdots K_h(x-X_{s_{2N}})K_h(\varphi_0(x)-Y_{s_{2N}})K_h(Z_{t_{2N}}-Z_{s_{2N}})a(\xi,Z_{t_{2N}})]$. This is bounded by $O(\frac{(\log T)^{2N}}{h_T^{2N(2+d_Z)}})$. Thus, we get

$$E[\hat{I}_{1211}(x,\xi)^{2N}]=O\left(\frac{1}{T^{4N}}T^{2N}\frac{(\log T)^{2N}}{h_T^{2N(2+d_Z)}}\right)=O\left(\left(\frac{(\log T)^2}{T^2h_T^{2(2+d_Z)}}\right)^N\right)$$
$$=O\left(\left(\frac{1}{Th_T^2}\right)^N\right)=O(a_T^N),$$

where we used $\frac{(\log T)^2}{Th_T^{2(1+d_Z)}}=O(1)$.

• *Bound of* $\hat{I}_2$. Let us write

$$\begin{aligned}\hat{I}_2(x,\xi) &= \int \Delta\hat{f}_{X,Y,Z}(x,\varphi_0(x),z)f_{X,Y,Z}(\xi,\varphi_0(\xi),z)\Sigma_0(z)\,dz\\ &\quad+\int \Delta\hat{f}_{X,Y,Z}(\xi,\varphi_0(\xi),z)f_{X,Y,Z}(x,\varphi_0(x),z)\Sigma_0(z)\,dz\\ &\quad+\int \Delta\hat{f}_{X,Y,Z}(x,\varphi_0(x),z)\Delta\hat{f}_{X,Y,Z}(\xi,\varphi_0(\xi),z)\Sigma_0(z)\,dz\\ &=: \hat{I}_{21}(x,\xi)+\hat{I}_{22}(x,\xi)+\hat{I}_{23}(x,\xi).\end{aligned}$$

Let us start from term $\hat{I}_{21}(x,\xi)$. We have $E[\hat{I}_{21}(x,\xi)^{2\bar{\zeta}}]=\int E[\prod_l\Delta\hat{f}_{X,Y,Z}(x,\varphi_0(x),z_l)]\prod_l f_{X,Y,Z}(\xi,\varphi_0(\xi),z_l)\Sigma_0(z_l)\prod_l dz_l$, where $\prod_l$ denotes product from $l=1$ to $l=2\bar{\zeta}$. Let $\Delta\hat{f}_{X,Y,Z}(x,\varphi_0(x),z)=\bar{f}(x,\varphi_0(x),z)+\bar{b}(x,\varphi_0(x),z)$, with $\bar{f}(x,\varphi_0(x),z):=\hat{f}_{X,Y,Z}(x,\varphi_0(x),z)-E[\hat{f}_{X,Y,Z}(x,\varphi_0(x),z)]$ and $\bar{b}(x,\varphi_0(x),z):=E[\hat{f}_{X,Y,Z}(x,\varphi_0(x),z)]-f_{X,Y,Z}(x,\varphi_0(x),z)$. We get

$$\begin{aligned}&E[\hat{I}_{21}(x,\xi)^{2\bar{\zeta}}]\\ &\quad=\sum_{J=0}^{2\bar{\zeta}}\binom{2\bar{\zeta}}{J}\left(\int E\left[\prod_{l=1}^{J}\bar{f}(x,\varphi_0(x),z_l)\right]\right.\\ &\qquad\left.\times\prod_{l=1}^{J}f_{X,Y,Z}(\xi,\varphi_0(\xi),z_l)\Sigma_0(z_l)\prod_{l=1}^{J}dz_l\right)\\ &\qquad\times\left(\int\bar{b}(x,\varphi_0(x),z)f_{X,Y,Z}(\xi,\varphi_0(\xi),z)\Sigma_0(z)\,dz\right)^{2\bar{\zeta}-J}.\end{aligned}\tag{SM.87}$$

From Assumption A.4(ii) and (iii), we have $\int\bar{b}(x,\varphi_0(x),z)f_{X,Y,Z}(\xi,\varphi_0(\xi),z)\Sigma_0(z)\,dz=O(h_T^m)$ uniformly in $x,\xi\in[0,1]$. Furthermore, by writing $\bar{f}(x,$

$\varphi_0(x), z) = \frac{1}{T}\sum_{t=1}^{T}\kappa_t(x, \varphi_0(x), z)$ and using the independence of the observations, we can write $E[\prod_{l=1}^{J}\bar{f}(x, \varphi_0(x), z_l)] = \sum_{n=1}^{\lfloor J/2\rfloor}\frac{D_{T,n}}{T^J}J_n(x, z_1, \dots, z_J)$, where $J_n(x, z_1, \dots, z_J)$ is a term that is split in a sum of products of n expectations, $D_{T,n} := T(T-1)\cdots(T-n+1)$, and $\lfloor J/2\rfloor$ denotes the largest integer that is smaller than or equal to $J/2$. To derive the order of the term in $E[\hat{I}_{21}(x, \xi)^{2\bar{\zeta}}]$ that corresponds to $J_n(x, z_1, \dots, z_J)$, note that all the powers $h_T^{-d_Z}$ can be eliminated by a change of variable, while a power h_T^{-2} coming from variable X and Y can be eliminated for each expectation term contained in $J_n(x, z_1, \dots, z_J)$. Thus, $\int J_n(x, z_1, \dots, z_J)\prod_{l=1}^{J}f_{X,Y,Z}(\xi, \varphi_0(\xi), z_l)\Sigma_0(z_l)\prod_{l=1}^{J}dz_j = O(\frac{1}{h_T^{2(J-n)}})$ uniformly in $x, \xi \in [0, 1]$. This implies $\int E[\prod_{l=1}^{J}\bar{f}(x, z_l)]\prod_{l=1}^{J}f_{X,Y,Z}(\xi, \varphi_0(\xi), z_l)\Sigma_0(z_l)\prod_{l=1}^{J}dz_j = O(\sum_{n=1}^{\lfloor J/2\rfloor}\frac{1}{(Th_T^2)^{J-n}}) = O(\frac{1}{(Th_T^2)^{\lceil J/2\rceil}})$, since $(Th_T^2)^{-1} = o(1)$. From (SM.87), we get

$$E[\hat{I}_{21}(x, \xi)^{2\bar{\zeta}}] = O\left(\sum_{J=0}^{2\bar{\zeta}}\frac{1}{(Th_T^2)^{\lceil J/2\rceil}}(h_T^{2m})^{\bar{\zeta}-J/2}\right) = O\left(\sum_{k=0}^{\bar{\zeta}}\left(\frac{1}{Th_T^2}\right)^k(h_T^{2m})^{\bar{\zeta}-k}\right) = O(a_T^{\bar{\zeta}})$$

uniformly in $x, \xi \in [0, 1]$.

The bound for the term $E[\hat{I}_{22}(x, \xi)^{2\bar{\zeta}}]$ is obtained similarly, interchanging x and ξ. Finally, let us consider the term $E[\hat{I}_{23}(x, \xi)^{2\bar{\zeta}}]$. We have $E[\hat{I}_{23}(x, \xi)^{2\bar{\zeta}}] = \int E[\prod_l \Delta\hat{f}_{X,Y,Z}(x, \varphi_0(x), z_l)\Delta\hat{f}_{X,Y,Z}(\xi, \varphi_0(\xi), z_l)]\prod_l \Sigma_0(z_l)\prod_l dz_l$, where the product $\prod_l$ is from $l = 1$ to $l = 2\bar{\zeta}$. Let $\mathcal{C}_1$ and $\mathcal{C}_2$ denote subsets of $\{1, 2, \dots, 2\bar{\zeta}\}$, and let $\mathcal{C}_1^c$ and $\mathcal{C}_2^c$ be their complements. We can write

$$\begin{aligned}
\text{(SM.88)}\quad & E[\hat{I}_{23}(x, \xi)^{2\bar{\zeta}}] \\
&= \sum_{\mathcal{C}_1,\mathcal{C}_2}\int E\Big[\prod_{l\in\mathcal{C}_1}\bar{f}(x, \varphi_0(x), z_l)\prod_{p\in\mathcal{C}_2}\bar{f}(\xi, \varphi_0(\xi), z_p)\Big] \\
&\quad\times\Big(\prod_{l\in\mathcal{C}_1^c}\bar{b}(x, \varphi_0(x), z_l)\prod_{p\in\mathcal{C}_2^c}\bar{b}(\xi, \varphi_0(\xi), z_p)\Big)\prod_l \Sigma_0(z_l)\prod_l dz_l \\
&= O\Big(\sum_{\mathcal{C}_1,\mathcal{C}_2}(h_T^m)^{4\bar{\zeta}-|\mathcal{C}_1|-|\mathcal{C}_2|} \\
&\quad\times\int E\Big[\prod_{l\in\mathcal{C}_1}\bar{f}(x, \varphi_0(x), z_l)\prod_{p\in\mathcal{C}_2}\bar{f}(\xi, \varphi_0(\xi), z_p)\Big]\prod_{l\in\mathcal{C}_1\cup\mathcal{C}_2}dz_l\Big),
\end{aligned}$$

where $|\mathcal{C}|$ denotes the cardinality of set $\mathcal{C}$. Similarly to above, for any two subsets $\mathcal{C}_1$ and $\mathcal{C}_2$ such that $|\mathcal{C}_1| + |\mathcal{C}_2| = J$, we can write $E[\prod_{l\in\mathcal{C}_1}\bar{f}(x, \varphi_0(x), z_l) \times$

$\prod_{p\in\mathcal{C}_2}\bar{f}(\xi,\varphi_0(\xi),z_p)]=\sum_{n=1}^{\lfloor J/2\rfloor}\frac{D_{T,n}}{T^J}J_n(x,\xi,\{z_l,l\in\mathcal{C}_1\cup\mathcal{C}_2\})$. Moreover, $\int J_n(x,\xi,\{z_l,l\in\mathcal{C}_1\cup\mathcal{C}_2\})\prod_{l\in\mathcal{C}_1\cup\mathcal{C}_2}dz_l=O(\frac{1}{h_T^{2(J-n)}}\frac{1}{h_T^{d_Z|\mathcal{C}_1\cap\mathcal{C}_2|}})$ uniformly in $x,\xi\in[0,1]$. Thus, among the terms in the sum (SM.88) such that $|\mathcal{C}_1|+|\mathcal{C}_2|=J$, the dominant term is the one with $|\mathcal{C}_1|=|\mathcal{C}_2|$ if J is even or with $|\mathcal{C}_1|=|\mathcal{C}_2|\pm 1$ if J is odd. These dominant terms are of $O(\frac{1}{h_T^{2(J-n)}}\frac{1}{h_T^{d_Z\lceil J/2\rceil}})$. We deduce, for any $\mathcal{C}_1$ and $\mathcal{C}_2$ such $|\mathcal{C}_1|+|\mathcal{C}_2|=J$, that

$$\int E\Big[\prod_{l\in\mathcal{C}_1}\bar{f}(x,\varphi_0(x),z_l)\prod_{p\in\mathcal{C}_2}\bar{f}(\xi,\varphi_0(\xi),z_p)\Big]\prod_{l\in\mathcal{C}_1\cup\mathcal{C}_2}dz_l$$

$$=O\left(\sum_{n=1}^{\lfloor J/2\rfloor}\frac{1}{(Th_T^2)^{J-n}}\frac{1}{h_T^{d_Z\lceil J/2\rceil}}\right)=O\left(\frac{1}{(Th_T^2)^{\lceil J/2\rceil}}\frac{1}{h_T^{d_Z\lceil J/2\rceil}}\right).$$

Thus, we get

$$\begin{aligned}E[\hat{I}_{23}(x,\xi)^{2\bar{\zeta}}]&=O\left(\sum_{J=0}^{4\bar{\zeta}}(h_T^m)^{4\bar{\zeta}-J}\frac{1}{(Th_T^2)^{\lceil J/2\rceil}}\frac{1}{h_T^{d_Z\lceil J/2\rceil}}\right)\\&=O\left(\sum_{k=0}^{2\bar{\zeta}}\left(\frac{1}{Th_T^{2+d_Z}}\right)^k(h_T^{2m})^{2\bar{\zeta}-k}\right)\\&=O\left(\left(\frac{1}{Th_T^{2+d_Z}}+h_T^{2m}\right)^{2\bar{\zeta}}\right).\end{aligned}$$

Since $\frac{1}{Th_T^{2+2d_Z}}=O(1)$ implies $(\frac{1}{Th_T^{2+d_Z}}+h_T^{2m})^2=O(\frac{1}{Th_T^2}+h_T^{2m})$, we get $E[\hat{I}_{23}(x,\xi)^{2\bar{\zeta}}]=O(a_T^{\bar{\zeta}})$ uniformly in $x,\xi\in[0,1]$.

SM.8.24. *Proof of Lemma C.9*

We have

$$\sum_j(j-n_T)=-\sum_{j=1}^{n_T/2}j=-\frac{n_T/2(n_T/2+1)}{2}=-\frac{n_T^2}{8}(1+o(1)).$$

Moreover,

$$\begin{aligned}\sum_j\log(j/n_T)&=n_T\sum_j\frac{1}{n_T}\log(j/n_T)=n_T\left(\int_{1/2}^1\log(y)\,dy+o(1)\right)\\&=\frac{n_T}{2}(\log 2-1)(1+o(1)),\end{aligned}$$

where we use the property of a Riemann sum and

$$\int_{1/2}^{1} \log(y)\,dy = [y\log(y) - y]|_{1/2}^{1} = \frac{1}{2}(\log 2 - 1).$$

Similarly,

$$\begin{aligned}
\sum_j (j - n_T)^2 &= \sum_{j=1}^{n_T/2} j^2 = \frac{n_T/2(n_T/2+1)(n_T+1)}{6} = \frac{n_T^3}{24}(1+o(1)),\\
\sum_j \log(j/n_T)^2 &= n_T\left(\int_{1/2}^{1} \log(y)^2\,dy + o(1)\right)\\
&= n_T\left(1 - \frac{1}{2}(\log 2)^2 - \log 2\right)(1+o(1)),\\
\sum_j (j - n_T)\log(j/n_T) &= n_T^2 \sum_j \frac{1}{n_T}\left(\frac{j}{n_T} - 1\right)\log(j/n_T)\\
&= n_T^2\left(\int_{1/2}^{1} (y-1)\log(y)\,dy + o(1)\right)\\
&= \frac{n_T^2}{16}(5 - 6\log 2)(1+o(1)),
\end{aligned}$$

where we use

$$\begin{aligned}
\int_{1/2}^{1} \log(y)^2\,dy &= [y\log(y)^2]|_{1/2}^{1} - 2\int_{1/2}^{1} \log(y)\,dy\\
&= -\frac{1}{2}(\log 2)^2 - \log 2 + 1,\\
\int_{1/2}^{1} (y-1)\log(y)\,dy &= \int_{1/2}^{1} y\log(y)\,dy - \int_{1/2}^{1} \log(y)\,dy\\
&= \frac{y^2}{2}\log(y)\Big|_{1/2}^{1} - \frac{1}{2}\int_{1/2}^{1} y\,dy - \frac{1}{2}(\log 2 - 1)\\
&= \frac{1}{16}(5 - 6\log 2).
\end{aligned}$$

SM.8.25. *Proof of Lemma C.10*

The proof is along the lines of the proof of Lemma C.4. We apply Lemma C.3 with $g(u) = g_j(u)$ and $e(u) = g_l(u)$ for any $j \neq l \in \mathbb{N}$. We have $\mathrm{Cov}[g_j(R), g_l(R)] = \frac{1}{\sqrt{\nu_j \nu_l}} \langle \phi_j, A^* A \phi_l \rangle_H = 0$ for $j \neq l$. Thus, we get $E[Z_{j,T} Z_{l,T}] = O(\rho(g_j, \kappa h_T)\rho(g_l, \kappa h_T)) + O(h_T^{1/4}(\|\psi_j\|_{H^2(F_Z)} + \|\psi_j\|_{H^2(\bar{F}_Z)}) \times (\|\psi_l\|_{H^2(F_Z)} + \|\psi_l\|_{H^2(\bar{F}_Z)}))$ uniformly in $j \neq l$. By similar arguments as in the proof of Lemma C.4, and using Assumptions A.4(i), A.5(ii) and (iii′), $\|\psi_j\|_{L^2(F_Z^*)}$, $\|\nabla \psi_j\|_{L^2(F_Z)}$, and $\|\nabla \psi_j\|_{L^2(F_Z^*)}$ are bounded and $\rho(g_j, \kappa h_T) = o(1)$ uniformly in $j \in \mathbb{N}$. The conclusion for part (ii) follows. The proof of part (i) is similar.

REFERENCES

ADAMS, R., AND J. FOURNIER (2003): *Sobolev Spaces*. Boston: Academic Press. [3,7]

ALT, H. (1992): *Linear Functional Analysis*. Berlin: Springer. [1]

ANDREWS, D. (1994): "Empirical Process Methods in Econometrics," in *Handbook of Econometrics*, Vol. 4, ed. by R. Engle and D. McFadden. Amsterdam: North-Holland, 2247–2294. [4,5]

BOSQ, D. (1998): *Nonparametric Statistics for Stochastic Processes*. Berlin: Springer. [32,33,45]

——— (2000): *Linear Processes in Function Spaces: Theory and Applications*. Berlin: Springer. [23]

DUBINSKIJ, J. (1986): *Sobolev Spaces of Infinite Order and Differential Equations*. Dordrecht: Kluwer Academic. [5]

ENGL, H., M. HANKE, AND A. NEUBAUER (2000): *Regularization of Inverse Problems*. Dordrecht: Kluwer Academic. [37]

GAGLIARDINI, P., AND O. SCAILLET (2012): "Tikhonov Regularization for Nonparametric Instrumental Variable Estimators," *Journal of Econometrics*, 167, 61–75. [14,69]

HALL, P., AND J. HOROWITZ (2005): "Nonparametric Methods for Inference in the Presence of Instrumental Variables," *The Annals of Statistics*, 33, 2904–2929. [28,54,55]

HANSEN, B. (2008): "Uniform Convergence Rates for Kernel Estimation With Dependent Data," *Econometric Theory*, 24, 726–748. [20,31,32]

HOROWITZ, J., AND S. LEE (2007): "Nonparametric Instrumental Variables Estimation of a Quantile Regression Model," *Econometrica*, 75, 1191–1208. [28]

KRESS, R. (1999): *Linear Integral Equations*. New York: Springer. [6,12,39]

REED, M., AND B. SIMON (1980): *Functional Analysis*. San Diego: Academic Press. [3]

WHITE, H., AND J. WOOLDRIDGE (1991): "Some Results on Sieve Estimation With Dependent Observations," in *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, ed. by W. Barnett, J. Powell, and G. Tauchen. New York: Cambridge University Press, 459–493. [14]

Faculty of Economics, University of Lugano, Via Buffi 13, 6900 Lugano, Switzerland and Swiss Finance Institute; patrick.gagliardini@usi.ch

and

GFRI, University of Geneva, Bd du Pont d'Arve 40, CH-1211 Genève 4, Suisse and Swiss Finance Institute; Olivier.Scaillet@unige.ch.

Manuscript received May, 2008; final revision received September, 2011.