

## Secondary Online Appendix to

# Optimal Sup-norm Rates and Uniform Inference on Nonlinear Functionals of Nonparametric IV Regression

Xiaohong Chen   Timothy M. Christensen

First version: August 2013. Revised: January 2017.

## G Supplementary Lemmas and Proofs

All the notation follow from the main text and the main online appendix. For a  $r \times c$  matrix  $A$  with  $r \leq c$  and full row rank  $r$  we let  $A_l^-$  denote its left pseudoinverse, namely  $(A'A)^- A'$  where  $'$  denotes transpose and  $^-$  denotes generalized inverse. We let  $s_{\min}(A)$  denote the minimum singular value of a rectangular matrix  $A$ .

Let  $s_{JK} = s_{\min}(G_b^{-1/2} S G_\psi^{-1/2})$ . Throughout the proofs in the appendix we use the identity

$$\begin{aligned} \psi^J(x)'(G_b^{-1/2} S)_l^- &= \psi^J(x)'(S'G_b^{-1}S)^- S'G_b^{-1/2} \\ &= \psi^J(x)'G_\psi^{-1/2}(G_\psi^{-1/2}S'G_b^{-1}S G_\psi^{-1/2})^- G_\psi^{-1/2}S'G_b^{-1/2} \\ &= \psi^J(x)'G_\psi^{-1/2}(G_b^{-1/2}S G_\psi^{-1/2})_l^- \end{aligned}$$

which implies that

$$\begin{aligned} \|\psi^J(x)'(G_b^{-1/2} S)_l^-\|_{\ell^2} &\leq \|\psi^J(x)'G_\psi^{-1/2}\|_{\ell^2} \|(G_b^{-1/2} S G_\psi^{-1/2})_l^-\|_{\ell^2} \\ &\leq \zeta_{\psi,J} \|(G_b^{-1/2} S G_\psi^{-1/2})_l^-\|_{\ell^2} \\ &\leq \zeta_{\psi,J} s_{JK}^{-1} \end{aligned} \tag{26}$$

by definition of  $\zeta_{\psi,J}$  and the fact that  $\|A_l^-\|_{\ell^2} \leq s_{\min}(A)^{-1}$  (see Lemma F.5).

### G.1 Proofs for Appendix A and Section 3.1

Since the proofs of results in Section 3.1 built upon those for results in Appendix A, we shall present the proofs for Appendix A first.

#### G.1.1 Proofs for Appendix A

*Proof of Lemma A.1.* First note that  $\tau_J > 0$  for all  $J$  by compactness and injectivity of  $T$ . Then:

$$s_{JK} = \inf_{h \in \Psi_J: \|h\|_{L^2(X)}=1} \|\Pi_K T h\|_{L^2(W)} \leq \inf_{h \in \Psi_J: \|h\|_{L^2(X)}=1} \|T h\|_{L^2(W)} = \tau_J^{-1}$$

holds uniformly in  $J$  because  $\Pi_K$  is a contraction, whence  $\tau_J \leq s_{JK}^{-1}$ . To derive a lower bound on  $\tau_J$ , the triangle inequality and Assumption 4(i) yield:

$$\begin{aligned} s_{JK} &= \inf_{h \in \Psi_{J,1}} \|\Pi_K T h\|_{L^2(W)} \\ &\geq \inf_{h \in \Psi_{J,1}} \|T h\|_{L^2(W)} - \sup_{h \in \Psi_{J,1}} \|(\Pi_K T - T)h\|_{L^2(W)} \\ &= (1 - o(1))\tau_J^{-1}. \end{aligned}$$

Therefore,  $s_{JK}^{-1} \leq (1 - o(1))^{-1}\tau_J$ . □

It is clear that Lemma A.2 is implied by the following lemma.

**Lemma G.1.** *Let Assumptions 1(iii) and 4(ii) hold. Then:*

- (1) (a)  $\|h_0 - \pi_J h_0\|_{L^2(X)} \asymp \|h_0 - \Pi_J h_0\|_{L^2(X)}$ ; and  
 (b)  $\tau_J \times \|T(h_0 - \pi_J h_0)\|_{L^2(W)} \leq \text{const} \times \|h_0 - \pi_J h_0\|_{L^2(X)}$ .
- (2) If Assumption 4(i) also holds, then: (a)  $\|Q_J h_0 - \pi_J h_0\|_{L^2(X)} \leq o(1) \times \|h_0 - \pi_J h_0\|_{L^2(X)}$ ; and  
 (b)  $\|h_0 - \Pi_J h_0\|_{L^2(X)} \asymp \|h_0 - Q_J h_0\|_{L^2(X)}$ .
- (3) If Assumption 4(iii') also holds, then:  $\|Q_J h_0 - \pi_J h_0\|_\infty \leq O(1) \times \|h_0 - \pi_J h_0\|_{L^2(X)}$ .
- (4) Further, if Condition (24) also holds, then Assumption 4(iii) is satisfied.

**Proof of Lemma G.1.** In what follows, “const” denotes a generic positive constant that may be different from line to line. Assumption 1(iii) guarantees  $\tau_J$  and  $\pi_J h_0$  are well defined. For part (1.a), we have:

$$\begin{aligned} \|h_0 - \Pi_J h_0\|_{L^2(X)} &\leq \|h_0 - \pi_J h_0\|_{L^2(X)} \\ &\leq \|h_0 - \Pi_J h_0\|_{L^2(X)} + \|\Pi_J h_0 - \pi_J h_0\|_{L^2(X)} \\ &\leq \|h_0 - \Pi_J h_0\|_{L^2(X)} + \tau_J \|T(\pi_J h_0 - \Pi_J h_0)\|_{L^2(W)} \\ &= \|h_0 - \Pi_J h_0\|_{L^2(X)} + \tau_J \|T\pi_J(h_0 - \Pi_J h_0)\|_{L^2(W)} \\ &\leq \|h_0 - \Pi_J h_0\|_{L^2(X)} + \tau_J \|T(h_0 - \Pi_J h_0)\|_{L^2(W)} \\ &= (1 + \text{const}) \times \|h_0 - \Pi_J h_0\|_{L^2(X)} \end{aligned}$$

where the third line is by definition of  $\tau_J$ , the fourth is because  $\pi_J h = h$  for all  $h \in \Psi_J$ , the final line is by Assumption 4(ii), and the fifth is because  $\pi_J$  is a weak contraction under the norm  $h \mapsto \|Th\|_{L^2(W)}$ . More precisely by the definition of  $\pi_J h_0$  we have:

$$\langle Th, T(h_0 - \pi_J h_0) \rangle_W = 0 \tag{27}$$

for all  $h \in \Psi_J$ , where  $\langle \cdot, \cdot \rangle_W$  denotes the  $L^2(W)$  inner product. With  $h = \pi_J h_0 - \Pi_J h_0 \in \Psi_J$  this implies

$$\langle T(\pi_J h_0 - \Pi_J h_0), T(h_0 - \pi_J h_0) \rangle_W = 0.$$

$$\langle T(\pi_J h_0 - \Pi_J h_0), T(h_0 - \Pi_J h_0) \rangle_W = \langle T(\pi_J h_0 - \Pi_J h_0), T(\pi_J h_0 - \Pi_J h_0) \rangle_W.$$

Thus  $\|T(\pi_J h_0 - \Pi_J h_0)\|_{L^2(W)} \leq \|T(h_0 - \Pi_J h_0)\|_{L^2(W)}$ .

For part (1.b):

$$\begin{aligned}
\tau_J \|T(h_0 - \pi_J h_0)\|_{L^2(W)} &\leq \tau_J \|T(h_0 - \Pi_J h_0)\|_{L^2(W)} \\
&\leq \text{const} \times \|h_0 - \Pi_J h_0\|_{L^2(X)} \\
&\leq \text{const} \times \|h_0 - \pi_J h_0\|_{L^2(X)}
\end{aligned}$$

where the first and final inequalities are by definition of  $\pi_J h_0$  and  $\Pi_J h_0$  and the second inequality is by Assumption 4(ii).

For part (2.a), Lemma A.1 guarantees that  $Q_J h_0$  is well defined and that  $s_{JK}^{-1} \leq 2\tau_J$  for all  $J$  sufficiently large. By definition of  $Q_J h_0$  we have:

$$\langle \Pi_K T h, T(h_0 - Q_J h_0) \rangle_W = 0 \quad (28)$$

for all  $h \in \Psi_J$ , where we use the fact that  $\langle \Pi_K f, g \rangle_W = \langle \Pi_K f, \Pi_K g \rangle_W$  holds for any  $f, g \in L^2(W)$  since  $\Pi_K$  is a projection). Substituting  $h = Q_J h_0 - \pi_J h_0 \in \Psi_J$  into the two equations (27) and (28) yields:

$$\langle (T - \Pi_K T)(Q_J h_0 - \pi_J h_0), T(h_0 - \pi_J h_0) \rangle_W + \langle \Pi_K T(Q_J h_0 - \pi_J h_0), T(h_0 - \pi_J h_0) \rangle_W = 0 \quad (29)$$

$$\langle \Pi_K T(Q_J h_0 - \pi_J h_0), T(h_0 - Q_J h_0) \rangle_W = 0. \quad (30)$$

By subtracting (30) from (29) we obtain

$$\langle (T - \Pi_K T)(Q_J h_0 - \pi_J h_0), T(h_0 - \pi_J h_0) \rangle_W + \|\Pi_K T(Q_J h_0 - \pi_J h_0)\|_{L^2(W)}^2 = 0$$

We have therefore proved

$$\|\Pi_K T(Q_J h_0 - \pi_J h_0)\|_{L^2(W)}^2 = |\langle (T - \Pi_K T)(Q_J h_0 - \pi_J h_0), T(h_0 - \pi_J h_0) \rangle_W|. \quad (31)$$

It follows from (31), the Cauchy-Schwarz inequality, and Assumption 4(i) that:

$$\begin{aligned}
s_{JK}^2 \|Q_J h_0 - \pi_J h_0\|_{L^2(X)}^2 &\leq \|\Pi_K T(Q_J h_0 - \pi_J h_0)\|_{L^2(W)}^2 \\
&\leq \|(T - \Pi_K T)(Q_J h_0 - \pi_J h_0)\|_{L^2(W)} \|T(h_0 - \pi_J h_0)\|_{L^2(W)} \\
&\leq o(\tau_J^{-1}) \|Q_J h_0 - \pi_J h_0\|_{L^2(X)} \|T(h_0 - \pi_J h_0)\|_{L^2(W)}.
\end{aligned} \quad (32)$$

$$\leq o(\tau_J^{-1}) \|Q_J h_0 - \pi_J h_0\|_{L^2(X)} \|T(h_0 - \pi_J h_0)\|_{L^2(W)}. \quad (33)$$

It follows by (33) and the relation  $s_{JK}^{-1} \leq 2\tau_J$  for all  $J$  large that:

$$\begin{aligned}
\|Q_J h_0 - \pi_J h_0\|_{L^2(X)} &\leq o(1) \times \tau_J \|T(h_0 - \pi_J h_0)\|_{L^2(W)} \\
&\leq o(1) \times \text{const} \times \|h_0 - \pi_J h_0\|_{L^2(X)}
\end{aligned}$$

where the final line is by part (1.b). For part (2.b), by definition of  $Q_J$ ,  $\Pi_J$  and results in part (1.a) and part (2.a), we have:

$$\begin{aligned}
\|h_0 - \Pi_J h_0\|_{L^2(X)} &\leq \|h_0 - Q_J h_0\|_{L^2(X)} \\
&\leq \|h_0 - \pi_J h_0\|_{L^2(X)} + \|\pi_J h_0 - Q_J h_0\|_{L^2(X)} \\
&\leq \|h_0 - \pi_J h_0\|_{L^2(X)} + o(1) \times \|h_0 - \pi_J h_0\|_{L^2(X)} \\
&= (1 + \text{const}) \times \|h_0 - \Pi_J h_0\|_{L^2(X)}.
\end{aligned}$$

This proves part (2.b).

For part (3), it follows from (32) and Assumption 4(iii') that

$$s_{JK}^2 \|Q_J h_0 - \pi_J h_0\|_{L^2(X)} \leq \text{const} \times (\zeta_{\psi, J} \tau_J)^{-1} \|T(h_0 - \pi_J h_0)\|_{L^2(W)}.$$

and hence

$$\begin{aligned} \|Q_J h_0 - \pi_J h_0\|_{L^2(X)} &\leq \text{const} \times \zeta_{\psi, J}^{-1} \times \tau_J \|T(h_0 - \pi_J h_0)\|_{L^2(W)} \\ &\leq \zeta_{\psi, J}^{-1} \times \text{const} \times \|h_0 - \pi_J h_0\|_{L^2(X)} \end{aligned} \quad (34)$$

by Part (1.b) and the fact that  $s_{JK}^{-1} \leq 2\tau_J$  for all  $J$  large. Therefore,

$$\begin{aligned} \|Q_J h_0 - \pi_J h_0\|_{\infty} &\leq \zeta_{\psi} \|Q_J h_0 - \pi_J h_0\|_{L^2(X)} \\ &\leq \text{const} \times \|h_0 - \pi_J h_0\|_{L^2(X)} \end{aligned}$$

where the last inequality is due to (34).

For part (4), by the triangle inequality, the results in part (1.a) and (3) and Condition (24) we have:

$$\begin{aligned} \|Q_J(h_0 - \Pi_J h_0)\|_{\infty} &\leq \|Q_J h_0 - \pi_J h_0\|_{\infty} + \|\pi_J h_0 - \Pi_J h_0\|_{\infty} \\ &\leq \text{const} \times \|h_0 - \Pi_J h_0\|_{L^2(X)} + \|h_0 - \Pi_J h_0\|_{\infty} \\ &\leq O(1) \times \|h_0 - \Pi_J h_0\|_{\infty} \end{aligned}$$

which completes the proof.  $\square$

Note that we may write  $\Pi_J h_0(x) = \psi^J(x)' c_J$  for some  $c_J$  in  $\mathbb{R}^J$ . We use this notation hereafter.

**Proof of Lemma A.3.** We first prove **Result (1)**. We begin by writing

$$\begin{aligned} \tilde{h}(x) - \Pi_J h_0(x) &= Q_J(h_0 - \Pi_J h_0)(x) \\ &\quad + \psi^J(x)' (G_b^{-1/2} S)_l^- \{G_b^{-1/2} (B'(H_0 - \Psi_{c_J})/n - E[b^K(W_i)(h_0(X_i) - \Pi_J h_0(X_i))])\} \\ &\quad + \psi^J(x)' \{(\widehat{G}_b^{-1/2} \widehat{S})_l^- \widehat{G}_b^{-1/2} G_b^{1/2} - (G_b^{-1/2} S)_l^- \} G_b^{-1/2} B'(H_0 - \Psi_{c_J})/n \\ &=: T_1 + T_2 + T_3 \end{aligned}$$

where  $Q_J : L^2(X) \rightarrow \Psi_J$  is the sieve 2SLS projection operator given by

$$Q_J h(x) = \psi^J(x)' [S' G_b^{-1} S]^{-1} S' G_b^{-1} E[b^K(W_i)h(X_i)].$$

Note that  $Q_J h = h$  for all  $h \in \Psi_J$ .

Control of  $\|T_1\|_{\infty}$ :  $\|T_1\|_{\infty} = O(1) \times \|h_0 - \Pi_J h_0\|_{\infty}$  by Assumption 4(iii).

Control of  $\|T_2\|_{\infty}$ : Using equation (26), the Cauchy-Schwarz inequality, and Lemma F.9, we obtain:

$$\begin{aligned} \|T_2\|_{\infty} &\leq \sup_x \|\psi^J(x) (G_b^{-1/2} S)_l^- \|_{\ell^2} \|G_b^{-1/2} (B'(H_0 - \Psi_{c_J})/n - E[b^K(W_i)(h_0(X_i) - \Pi_J h_0(X_i))])\|_{\ell^2} \\ &\leq \zeta_{\psi, J} s_{JK}^{-1} \|G_b^{-1/2} (B'(H_0 - \Psi_{c_J})/n - E[b^K(W_i)(h_0(X_i) - \Pi_J h_0(X_i))])\|_{\ell^2} \\ &= \zeta_{\psi, J} s_{JK}^{-1} \times O_p(\sqrt{K/n}) \times \|h_0 - \Pi_J h_0\|_{\infty}. \end{aligned}$$

It then follows by the relations  $s_{JK}^{-1} \asymp \tau_J$  (Lemma A.1) and  $\zeta_{\psi,J} \geq \sqrt{J} \asymp \sqrt{K}$  and Assumption 3(ii) that:

$$\|T_2\|_\infty = O_p(\tau_J \zeta_{\psi,J} \sqrt{J/n}) \times \|h_0 - \Pi_J h_0\|_\infty = O_p(1) \times \|h_0 - \Pi_J h_0\|_\infty.$$

Control of  $\|T_3\|_\infty$ : Similar to  $T_2$  in the proof of Lemma 3.1, we may use Lemmas F.10(b) and A.1 to obtain:

$$\begin{aligned} \|T_3\|_\infty &\leq \zeta_{\psi,J} \|G_\psi^{1/2} \{(\widehat{G}_b^{-1/2} \widehat{S})_l^- \widehat{G}_b^{-1/2} G_b^{1/2} - (G_b^{-1/2} S)_l^-\} \|_{\ell^2} \|G_b^{-1/2} B'(H_0 - \Psi_{c_J})/n\|_{\ell^2} \\ &= \zeta_{\psi,J} \times O_p(\tau_J^2 \zeta \sqrt{(\log J)/n}) \times \|G_b^{-1/2} B'(H_0 - \Psi_{c_J})/n\|_{\ell^2}. \end{aligned} \quad (35)$$

Then by Lemma F.9 and the triangle inequality, we have:

$$\begin{aligned} \|G_b^{-1/2} B'(H_0 - \Psi_{c_J})/n\|_{\ell^2} &\leq O_p(\sqrt{K/n}) \times \|h_0 - \Pi_J h_0\|_\infty + \|\Pi_K T(h_0 - \Pi_J h_0)\|_{L^2(W)} \\ &\leq O_p(\sqrt{K/n}) \times \|h_0 - \Pi_J h_0\|_\infty + \|T(h_0 - \Pi_J h_0)\|_{L^2(W)}. \end{aligned} \quad (36)$$

Substituting (36) into (35) and using Assumptions 3(ii) and 4(ii):

$$\begin{aligned} \|T_3\|_\infty &\leq O_p(\tau_J \zeta^2 / \sqrt{n}) \times \left( O_p(\tau_J \sqrt{K(\log J)/n}) \times \|h_0 - \Pi_J h_0\|_\infty + \tau_J \|T(h_0 - \Pi_J h_0)\|_{L^2(W)} \right) \\ &= O_p(1) \times (o_p(1) \times \|h_0 - \Pi_J h_0\|_\infty + O_p(1) \times \|h_0 - \Pi_J h_0\|_{L^2(X)}) \\ &\leq O_p(1) \times \|h_0 - \Pi_J h_0\|_\infty \end{aligned}$$

where the final line is by the relation between the  $L^2(X)$  and sup norms.

**Result (2)** then follows because

$$\begin{aligned} \|\tilde{h} - h_0\|_\infty &\leq \|\tilde{h} - \Pi_J h_0\|_\infty + \|\Pi_J h_0 - h_0\|_\infty \\ &\leq (1 + O_p(1)) \|\Pi_J h_0 - h_0\|_\infty \\ &\leq (1 + O_p(1))(1 + \|\Pi_J\|_\infty) \|h_0 - h_{0,J}\|_\infty. \end{aligned}$$

where the second inequality is by Result (1) and the final line is by Lebesgue's lemma.  $\square$

### G.1.2 Proofs for Section 3.1

**Proof of Lemma 3.1.** Let  $u = (u_1, \dots, u_n)'$ . Let  $M_n$  be a sequence of positive constants diverging to  $+\infty$ , and decompose  $u_i = u_{1,i} + u_{2,i}$  where

$$\begin{aligned} u_{1,i} &= u_i \{|u_i| \leq M_n\} - E[u_i \{|u_i| \leq M_n\} | W_i] \\ u_{2,i} &= u_i \{|u_i| > M_n\} - E[u_i \{|u_i| > M_n\} | W_i] \\ u_1 &= (u_{1,1}, \dots, u_{1,n})' \\ u_2 &= (u_{2,1}, \dots, u_{2,n})'. \end{aligned}$$

For **Result (1)**, recall that  $\xi_{\psi,J} = \sup_x \|\psi^J(x)\|_{\ell^1}$ . By Hölder's inequality we have

$$\|\widehat{h} - \tilde{h}\|_\infty = \sup_x |\psi^J(x)'(\widehat{c} - \tilde{c})| \leq \xi_{\psi,J} \|(\widehat{c} - \tilde{c})\|_{\ell^\infty}.$$

To derive the sup-norm convergence rate of the standard deviation term  $\widehat{h} - \widetilde{h}$ , it suffices to bound the  $\ell^\infty$  norm of the  $J \times 1$  random vector  $(\widehat{c} - \widetilde{c})$ . Although this appears like a crude bound,  $\xi_{\psi,J}$  grows slowly in  $J$  for certain sieves whose basis functions have local support. For such bases the above bound, in conjunction with the following result

$$\|\widehat{c} - \widetilde{c}\|_{\ell^\infty} = O_p \left( s_{JK}^{-1} \sqrt{(\log J)/(ne_J)} \right) \quad (37)$$

leads to a tight bound on the convergence rate of  $\|\widehat{h} - \widetilde{h}\|_\infty$ .

To prove (37), we begin by writing

$$\begin{aligned} \widehat{c} - \widetilde{c} &= (\widehat{G}_b^{-1/2} \widehat{S})_l^- \widehat{G}_b^{-1/2} B' u/n \\ &= (G_b^{-1/2} S)_l^- G_b^{-1/2} B' u/n + \{(\widehat{G}_b^{-1/2} \widehat{S})_l^- \widehat{G}_b^{-1/2} - (G_b^{-1/2} S)_l^- G_b^{-1/2}\} B' u/n \\ &=: T_1 + T_2. \end{aligned}$$

We will show that  $\|T_1\|_{\ell^\infty} = O_p \left( s_{JK}^{-1} \sqrt{(\log J)/(ne_J)} \right)$  and  $\|T_2\|_{\ell^\infty} = O_p \left( s_{JK}^{-1} \sqrt{(\log J)/(ne_J)} \right)$ .

Control of  $\|T_1\|_{\ell^\infty}$ . Note that  $T_1 = (G_b^{-1/2} S)_l^- G_b^{-1/2} B' u_1/n + (G_b^{-1/2} S)_l^- G_b^{-1/2} B' u_2/n$ .

Let  $(a)_j$  denote the  $j$ th element of a vector  $a$ . By the definition of  $\|\cdot\|_{\ell^\infty}$  and the union bound,

$$\begin{aligned} \mathbb{P} \left( \|(G_b^{-1/2} S)_l^- G_b^{-1/2} B' u_1/n\|_{\ell^\infty} > t \right) &\leq \mathbb{P} \left( \bigcup_{j=1}^J |((G_b^{-1/2} S)_l^- G_b^{-1/2} B' u_1/n)_j| > t \right) \\ &\leq \sum_{j=1}^J \mathbb{P} \left( |((G_b^{-1/2} S)_l^- G_b^{-1/2} B' u_1/n)_j| > t \right) \\ &= \sum_{j=1}^J \mathbb{P} \left( \left| \sum_{i=1}^n q_{j,JK}(W_i) u_{1,i}/n \right| > t \right) \end{aligned} \quad (38)$$

where  $q_{j,JK}(W_i) = ((G_b^{-1/2} S)_l^- G_b^{-1/2} b^K(W_i))_j$ . The summands may be bounded by noting that

$$\begin{aligned} |q_{j,JK}(W_i)| &\leq \|(G_b^{-1/2} S)_l^- \|_{\ell^2} \|G_b^{-1/2} b^K(W_i)\|_{\ell^2} \\ &= \|G_\psi^{-1/2} [G_\psi^{-1/2} S' G_b^{-1} S G_\psi^{-1/2}]^{-1} G_\psi^{-1/2} S' G_b^{-1/2}\|_{\ell^2} \|G_b^{-1/2} b^K(W_i)\|_{\ell^2} \\ &\leq \|G_\psi^{-1/2}\|_{\ell^2} \| [G_\psi^{-1/2} S' G_b^{-1} S G_\psi^{-1/2}]^{-1} G_\psi^{-1/2} S' G_b^{-1/2}\|_{\ell^2} \|G_b^{-1/2} b^K(W_i)\|_{\ell^2} \\ &\leq \frac{\zeta_{b,K}}{s_{JK} \sqrt{e_J}} \end{aligned} \quad (39)$$

uniformly in  $i$  and  $j$ . Therefore,

$$|q_{j,JK}(W_i) u_{1,i}/n| \leq \frac{2M_n \zeta_{b,K}}{n s_{JK} \sqrt{e_J}} \quad (40)$$

uniformly in  $i$  and  $j$ .

Let  $(A)_{j\cdot}$  denote the  $j$ th row of the matrix  $A$  and let  $(A)_{jj}$  denote its  $j$ th diagonal element. The second moments

of the summands may be bounded by observing that

$$\begin{aligned}
E[q_{j,JK}(W_i)^2] &= E[((G_b^{-1/2}S)_l^-)_j | G_b^{-1/2} b^K(W_i)]^2 \\
&= E[((G_b^{-1/2}S)_l^-)_j | G_b^{-1/2} b^K(W_i) b^K(W_i)' G_b^{-1/2} ((G_b^{-1/2}S)_l^-)'_{jj}] \\
&= ((G_b^{-1/2}S)_l^- ((G_b^{-1/2}S)_l^-)')_{jj} \\
&= ((S' G_b^{-1} S)^{-1})_{jj} \\
&\leq \| (S' G_b^{-1} S)^{-1} \|_{\ell^2} \\
&= \| G_\psi^{-1/2} [G_\psi^{-1/2} S' G_b^{-1} S G_\psi^{-1/2}]^{-1} G_\psi^{-1/2} \|_{\ell^2} \\
&\leq \frac{1}{s_{JK}^2 e_J}
\end{aligned} \tag{41}$$

and so

$$E[(q_{j,JK}(W_i) u_{1,i}/n)^2] \leq \frac{\bar{\sigma}^2}{n^2 s_{JK}^2 e_J} \tag{42}$$

by Assumption 2(i) and the law of iterated expectations. Bernstein's inequality and expressions (38), (40) and (42) yield

$$\begin{aligned}
&\mathbb{P} \left( \| (G_b^{-1/2}S)_l^- G^{-1/2} B' u_1/n \|_{\ell^\infty} > C s_{JK}^{-1} \sqrt{(\log J)/(ne_J)} \right) \\
&\leq 2 \exp \left\{ \log J - \frac{C^2 (\log J)/(n s_{JK}^2 e_J)}{c_1/(n s_{JK}^2 e_J) + c_2 C M_n \zeta_{b,K} \sqrt{\log J}/(n^3 s_{JK}^2 e_J)} \right\} \\
&= 2 \exp \left\{ \log J - \frac{C^2 (\log J)/(n s_{JK}^2 e_J)}{1/(n s_{JK}^2 e_J) [c_1 + c_2 C M_n \zeta_{b,K} \sqrt{(\log J)/n}]} \right\}
\end{aligned} \tag{43}$$

for finite positive constants  $c_1$  and  $c_2$ . Then (43) is  $o(1)$  for all large  $C$  provided  $M_n \zeta_{b,K} \sqrt{(\log J)/n} = o(1)$ .

By the triangle and Markov inequalities and (39), we have

$$\begin{aligned}
\mathbb{P} \left( \| (G_b^{-1/2}S)_l^- G^{-1/2} B' u_2/n \|_{\ell^\infty} > t \right) &= \mathbb{P} \left( \max_{1 \leq j \leq J} \left| \sum_{i=1}^n q_{j,JK}(W_i) u_{2,i}/n \right| > t \right) \\
&\leq \mathbb{P} \left( \frac{\zeta_{b,K}}{s_{JK} \sqrt{e_J}} \sum_{i=1}^n |u_{2,i}/n| > t \right) \\
&\leq \frac{2\zeta_{b,K}}{t s_{JK} \sqrt{e_J}} E[|u_i| \{ |u_i| > M_n \}] \\
&\leq \frac{2\zeta_{b,K}}{t s_{JK} \sqrt{e_J} M_n^{1+\delta}} E[|u_i|^{2+\delta} \{ |u_i| > M_n \}]
\end{aligned}$$

which, by Assumption 2(ii), is  $o(1)$  when  $t = C s_{JK}^{-1} \sqrt{(\log J)/(ne_J)}$  provided  $\zeta_{b,K} \sqrt{n/(\log J)} = O(M_n^{1+\delta})$ .

Choosing  $M_n^{1+\delta} \asymp \zeta_{b,K} \sqrt{n/(\log J)}$  satisfies the condition  $\zeta_{b,K} \sqrt{n/(\log J)} = O(M_n^{1+\delta})$  trivially, and satisfies the condition  $M_n \zeta_{b,K} \sqrt{(\log J)/n} = o(1)$  provided  $\zeta_{b,K}^{(2+\delta)/\delta} \sqrt{(\log J)/n} = o(1)$ , which holds by Assumption 3(iii).

Control of  $\|T_2\|_{\ell^\infty}$ : Using the fact that  $\|\cdot\|_{\ell^\infty} \leq \|\cdot\|_{\ell^2}$  on  $\mathbb{R}^J$  and Lemmas F.10(a) and F.8, we have:

$$\begin{aligned}
\|T_2\|_{\ell^\infty} &= \|\{(\widehat{G}_b^{-1/2}\widehat{S})_l^- \widehat{G}_b^{-1/2} G_b^{1/2} - (G_b^{-1/2} S)_l^-\} G_b^{-1/2} B' u/n\|_{\ell^\infty} \\
&\leq \|(\widehat{G}_b^{-1/2}\widehat{S})_l^- \widehat{G}_b^{-1/2} G_b^{1/2} - (G_b^{-1/2} S)_l^-\|_{\ell^2} \|G_b^{-1/2} B' u/n\|_{\ell^2} \\
&= O_p\left(s_{JK}^{-2} \zeta \sqrt{(\log K)/(ne_J)}\right) \times O_p(\sqrt{K/n}) \\
&= O_p\left(s_{JK}^{-1} \sqrt{(\log J)/(ne_J)}\right) \times O_p(s_{JK}^{-1} \zeta \sqrt{K/n}) \\
&= O_p\left(s_{JK}^{-1} \sqrt{(\log J)/(ne_J)}\right)
\end{aligned}$$

where the last equality follows from Assumption 3(ii) and the facts that  $\zeta \geq \sqrt{K}$  and  $J \asymp K$ .

For **Result (2)**, we begin by writing

$$\begin{aligned}
\widehat{h}(x) - \widetilde{h}(x) &= \psi^J(x)' (\widehat{G}_b^{-1/2}\widehat{S})_l^- \widehat{G}_b^{-1/2} B' u/n \\
&= \psi^J(x)' (G_b^{-1/2} S)_l^- G_b^{-1/2} B' u/n + \psi^J(x)' \{(\widehat{G}_b^{-1/2}\widehat{S})_l^- \widehat{G}_b^{-1/2} - (G_b^{-1/2} S)_l^- G_b^{-1/2}\} B' u/n \\
&=: T_1 + T_2.
\end{aligned}$$

We will show that  $\|T_1\|_\infty = O_p\left(\tau_J \zeta_{\psi,J} \sqrt{(\log n)/n}\right)$  and  $\|T_2\|_\infty = O_p\left(\tau_J \zeta_{\psi,J} \sqrt{(\log n)/n}\right)$ .

Control of  $\|T_1\|_\infty$ . Note that  $T_1 = \psi^J(x)' (G_b^{-1/2} S)_l^- G_b^{-1/2} B' u_1/n + \psi^J(x)' (G_b^{-1/2} S)_l^- G_b^{-1/2} B' u_2/n$ .

Let  $\mathcal{X}_n \subset \mathcal{X}$  be a grid of finitely many points such that for each  $x \in \mathcal{X}$  there exists a  $\bar{x}_n(x) \in \mathcal{X}_n$  such that  $\|x - \bar{x}_n(x)\| \lesssim (\zeta_{\psi,J} J^{-(\omega+\frac{1}{2})})^{1/\omega'}$ , where  $\omega, \omega'$  are as in Assumption 3(i). By compactness and convexity of the support  $\mathcal{X}$  of  $X_i$ , we may choose  $\mathcal{X}_n$  to have cardinality  $\#(\mathcal{X}_n) \lesssim n^\beta$  for some  $0 < \beta < \infty$ . Therefore,

$$\begin{aligned}
&\sup_x \|\psi^J(x)' (G_b^{-1/2} S)_l^- G_b^{-1/2} B' u_1/n\|_\infty \\
&\leq \max_{x_n \in \mathcal{X}_n} |\psi^J(x_n)' (G_b^{-1/2} S)_l^- G_b^{-1/2} B' u_1/n| + \sup_x |\{\psi^J(x) - \psi^J(\bar{x}_n(x))\}' (G_b^{-1/2} S)_l^- G_b^{-1/2} B' u_1/n| \\
&\leq \max_{x_n \in \mathcal{X}_n} |\psi^J(x_n)' (G_b^{-1/2} S)_l^- G_b^{-1/2} B' u_1/n| + C_\omega J^\omega \zeta_{\psi,J} J^{-(\omega+\frac{1}{2})} s_{JK}^{-1} \|G_b^{-1/2} B' u_1/n\|_{\ell^2} \\
&= \max_{x_n \in \mathcal{X}_n} |\psi^J(x_n)' (G_b^{-1/2} S)_l^- G_b^{-1/2} B' u_1/n| + C_\omega J^\omega \zeta_{\psi,J} J^{-(\omega+\frac{1}{2})} s_{JK}^{-1} \times O_p(\sqrt{J/n}) \\
&= \max_{x_n \in \mathcal{X}_n} |\psi^J(x_n)' (G_b^{-1/2} S)_l^- G_b^{-1/2} B' u_1/n| + o_p(s_{JK}^{-1} \zeta_{\psi,J} \sqrt{(\log J)/n})
\end{aligned}$$

for some finite positive constant  $C_\omega$ , where the first inequality is by the triangle inequality, the second is by Hölder continuity of the basis for  $\Psi_J$  and similar reasoning to that used in equation (26), the first equality is by Lemma F.8 and the fact that  $J \asymp K$ , and the final equality is because  $(\log J)^{-1/2} = o(1)$ . For each  $x_n \in \mathcal{X}_n$  we may write

$$\begin{aligned}
\psi^J(x_n)' (G_b^{-1/2} S)_l^- G_b^{-1/2} B' u_1/n &= \frac{1}{n} \sum_{i=1}^n g_{n,i}(x_n) u_{1,i}, \text{ where} \\
g_{n,i}(x_n) &= \psi^J(x_n)' (G_b^{-1/2} S)_l^- G_b^{-1/2} b^K(W_i).
\end{aligned}$$

It follows from equation (26) and the Cauchy-Schwarz inequality that the bounds

$$\begin{aligned}
|g_{n,i}(x_n)| &\leq s_{JK}^{-1} \zeta_{\psi,J} \zeta_{b,K} \\
E[g_{n,i}(x_n)^2] &= \psi^J(x_n)' (G_b^{-1/2} S)_l^- ((G_b^{-1/2} S)_l^-)' \psi^J(x_n) \leq s_{JK}^{-2} \zeta_{\psi,J}^2
\end{aligned}$$

hold uniformly for  $x_n \in \mathcal{X}_n$ . Therefore, by Assumption 2(i) and iterated expectations, the bounds

$$\begin{aligned} |g_{n,i}(x_n)u_{1,i}| &\leq 2s_{JK}^{-1}\zeta_{\psi,J}\zeta_{b,K}M_n \\ E[g_{n,i}(x_n)^2u_{1,i}^2] &\leq \bar{\sigma}^2\psi^J(x_n)'(G_b^{-1/2}S)_l^-((G_b^{-1/2}S)_l^-)' \psi^J(x_n) \leq \bar{\sigma}^2s_{JK}^{-2}\zeta_{\psi,J}^2 \end{aligned}$$

hold uniformly for  $x_n \in \mathcal{X}_n$ . It follows by the union bound and Bernstein's inequality that

$$\begin{aligned} &\mathbb{P}\left(\max_{x_n \in \mathcal{X}_n} |\psi^J(x_n)'(G_b^{-1/2}S)_l^-G_b^{-1/2}B'u_1/n| > Cs_{JK}^{-1}\zeta_{\psi,J}\sqrt{(\log n)/n}\right) \\ &\leq \#(\mathcal{X}_n) \max_{x_n \in \mathcal{X}_n} \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n g_{n,i}(x_n)u_{1,i}\right| > Cs_{JK}^{-1}\zeta_{\psi,J}\sqrt{(\log n)/n}\right) \\ &\lesssim \exp\left\{\beta \log n - \frac{C^2\zeta_{\psi,J}^2(\log n)/(ns_{JK}^2)}{c_1\zeta_{\psi,J}^2/(ns_{JK}^2)[1 + (c_2/c_1)(CM_n\zeta_{b,K}\sqrt{(\log n)/n}]}\right\} \end{aligned} \quad (44)$$

for finite positive constants  $c_1$  and  $c_2$ . Then (44) is  $o(1)$  for all large  $C$  provided  $M_n\zeta_{b,K}\sqrt{(\log n)/n} = o(1)$ .

By the triangle and Markov inequalities and equation (26), we have

$$\begin{aligned} \mathbb{P}\left(\|\psi^J(x)(G_b^{-1/2}S)_l^-G_b^{-1/2}B'u_2/n\|_\infty > t\right) &\leq \mathbb{P}\left(s_{JK}^{-1}\zeta_{\psi,J}\|G_b^{-1/2}B'u_2/n\|_{\ell^2} > t\right) \\ &\leq \mathbb{P}\left(s_{JK}^{-1}\zeta_{\psi,J}\zeta_{b,K}\sum_{i=1}^n |u_{2,i}/n| > t\right) \\ &\leq \frac{2\zeta_{\psi,J}\zeta_{b,K}}{ts_{JK}}E[|u_i|\{ |u_i| > M_n \}] \\ &\leq \frac{2\zeta_{\psi,J}\zeta_{b,K}}{ts_{JK}M_n^{1+\delta}}E[|u_i|^{2+\delta}\{ |u_i| > M_n \}] \end{aligned}$$

which, by Assumption 2(ii), is  $o(1)$  when  $t = Cs_{JK}^{-1}\zeta_{\psi,J}\sqrt{(\log n)/n}$  provided  $\zeta_{b,K}\sqrt{n/(\log n)} = O(M_n^{1+\delta})$ .

Choosing  $M_n^{1+\delta} \asymp \zeta_{b,K}\sqrt{n/\log n}$  satisfies the condition  $\zeta_{b,K}\sqrt{n/(\log n)} = O(M_n^{1+\delta})$  trivially, and satisfies the condition  $M_n\zeta_{b,K}\sqrt{(\log n)/n} = o(1)$  provided  $\zeta_{b,K}^{(2+\delta)/\delta}\sqrt{(\log n)/n} = o(1)$ , which holds by Assumption 3(iii). We have therefore proved that  $\|T_1\|_\infty = O_p(s_{JK}^{-1}\zeta_{\psi,J}\sqrt{(\log n)/n})$ . It follows by the relation  $\tau_J \asymp s_{JK}^{-1}$  (Lemma A.1) that  $\|T_1\|_\infty = O_p(\tau_J\zeta_{\psi,J}\sqrt{(\log n)/n})$ .

Control of  $\|T_2\|_\infty$ : Using the fact that  $\|h\|_\infty \leq \zeta_{\psi,J}\|h\|_{L^2(X)}$  on  $\Psi_J$  and Lemmas F.10(b) and F.8 and the relation  $\tau_J \asymp s_{JK}^{-1}$ , we have:

$$\begin{aligned} \|T_2\|_\infty &\leq \zeta_{\psi,J}\|G_\psi^{1/2}\{(\widehat{G}_b^{-1/2}\widehat{S})_l^-\widehat{G}_b^{-1/2}G_b^{1/2} - (G_b^{-1/2}S)_l^-\}G_b^{-1/2}B'u/n\|_{\ell^2} \\ &\leq \zeta_{\psi,J}\|G_\psi^{1/2}\{(\widehat{G}_b^{-1/2}\widehat{S})_l^-\widehat{G}_b^{-1/2}G_b^{1/2} - (G_b^{-1/2}S)_l^-\}\|_{\ell^2}\|G_b^{-1/2}B'u/n\|_{\ell^2} \\ &= \zeta_{\psi,J}O_p\left(\tau_J^2\zeta\sqrt{(\log J)/n}\right) \times O_p(\sqrt{K/n}) \\ &= O_p\left(\tau_J\zeta_{\psi,J}\sqrt{(\log J)/n}\right) \times O_p(\tau_J\zeta\sqrt{K/n}) \\ &= O_p\left(\tau_J\zeta_{\psi,J}\sqrt{(\log J)/n}\right) \times O_p(1) \end{aligned}$$

where the last equality follows from Assumption 3(ii) and the fact that  $\zeta \geq \sqrt{J} \asymp \sqrt{K}$ .  $\square$

**Proof of Theorem 3.1.** We decompose  $\|\widehat{h} - h_0\|_\infty$  into three parts:

$$\|\widehat{h} - h_0\|_\infty \leq \|\widehat{h} - \widetilde{h}\|_\infty + \|\widetilde{h} - \Pi_J h_0\|_\infty + \|\Pi_J h_0 - h_0\|_\infty.$$

where  $\|\widehat{h} - \widetilde{h}\|_\infty = O_p(\tau_J \xi_{\psi, J} \sqrt{(\log J)/(ne_J)})$  by Lemma 3.1(1) and  $\|\widetilde{h} - \Pi_J h_0\|_\infty = O_p(1) \times \|\Pi_J h_0 - h_0\|_\infty$  by Lemma A.3.  $\square$

**Proof of Corollary 3.1.** For **Result (1)**, note that Assumption 3(ii) is satisfied with  $\zeta = O(J^{1/2})$  for  $\Psi_J$  and  $B_K$  being spline, or wavelet or cosine sieves. Next, by the lemmas in Appendix E,  $\xi_{\psi, J}/\sqrt{e_J} = O(J^{1/2})$  for  $\Psi_J$  being spline or wavelet sieves. Also,  $\|\Pi_J\|_\infty \lesssim 1$  for  $\Psi_J$  being a spline sieve (Huang (2003)) or a tensor product CDV wavelet sieve (Chen and Christensen (2015)). For  $h_0 \in B_\infty(p, L)$  and  $\Psi_J$  being spline or wavelet sieves, Lemma A.3 implies that

$$\|\widetilde{h} - h_0\|_\infty = O_p(J^{-p/d}).$$

Note that Bernstein inequalities (or inverse estimates) from approximation theory imply that

$$\|\partial^\alpha h\|_\infty = O(J^{|\alpha|/d}) \|h\|_\infty$$

for all  $h \in \Psi_J$  (see Schumaker (2007) for splines and Cohen (2003) for wavelets on domains). Therefore,

$$\begin{aligned} \|\partial^\alpha \widetilde{h} - \partial^\alpha h_0\|_\infty &\leq \|\partial^\alpha \widetilde{h} - \partial^\alpha (\Pi_J h_0)\|_\infty + \|\partial^\alpha (\Pi_J h_0) - \partial^\alpha h_0\|_\infty \\ &\leq O(J^{|\alpha|/d}) \|\widetilde{h} - \Pi_J h_0\|_\infty + \|\partial^\alpha (\Pi_J h_0) - \partial^\alpha h_0\|_\infty \\ &\leq O_p(J^{-(p-|\alpha|)/d}) + \|\partial^\alpha (\Pi_J h_0) - \partial^\alpha h_0\|_\infty \end{aligned}$$

Let  $h_J$  be any element of  $\Psi_J$ . Since  $\Pi_J h_J = h_J$ , we have:

$$\begin{aligned} \|\partial^\alpha (\Pi_J h_0) - \partial^\alpha h_0\|_\infty &= \|\partial^\alpha (\Pi_J (h_0 - h_J)) + \partial^\alpha h_J - \partial^\alpha h_0\|_\infty \\ &\leq O(J^{|\alpha|/d}) \|\Pi_J (h_0 - h_J)\|_\infty + \|\partial^\alpha h_J - \partial^\alpha h_0\|_\infty \\ &\leq O(J^{|\alpha|/d}) \times \text{const} \times \|h_0 - h_J\|_\infty + \|\partial^\alpha h_J - \partial^\alpha h_0\|_\infty. \end{aligned}$$

The above inequality holds uniformly in  $h_J \in \Psi_J$ . Choosing  $h_J$  such that  $\|h_0 - h_J\|_\infty = O(J^{-p/d})$  and  $\|\partial^\alpha h_J - \partial^\alpha h_0\|_\infty = O(J^{-(p-|\alpha|)/d})$  yields the desired result.

For **Result (2)**, Theorem 3.1 implies that

$$\|\widehat{h} - h_0\|_\infty = O_p(J^{-p/d} + \tau_J \sqrt{(J \log J)/n}).$$

By similar arguments to the above, we have:

$$\begin{aligned} \|\partial^\alpha \widehat{h} - \partial^\alpha h_0\|_\infty &\leq \|\partial^\alpha \widehat{h} - \partial^\alpha \widetilde{h}\|_\infty + \|\partial^\alpha \widetilde{h} - \partial^\alpha h_0\|_\infty \\ &\leq O(J^{|\alpha|/d}) \|\widehat{h} - \widetilde{h}\|_\infty + \|\partial^\alpha \widetilde{h} - \partial^\alpha h_0\|_\infty \\ &\leq O_p \left( J^{|\alpha|/d} \left( \tau_J \sqrt{(J \log J)/n} \right) \right) + \|\partial^\alpha \widetilde{h} - \partial^\alpha h_0\|_\infty \end{aligned}$$

and the result follows by Result (1).

For **Results (2.a) and (2.b)**, Assumption 3(ii)(iii) is satisfied if  $\tau_J \times J/\sqrt{n} = O(1)$  and  $J^{(2+\delta)/\delta} (\log n)/n = o(1)$ . This is satisfied given the stated conditions with the optimal choice of  $J$  for mildly ill-posed case and severely

ill-posed case respectively.  $\square$

## G.2 Proofs for Section 3.2

**Proof of Theorem 3.2.** Consider the Gaussian reduced-form NPIR model with known operator  $T$ :

$$\begin{aligned} Y_i &= Th_0(W_i) + u_i \\ u_i|W_i &\sim N(0, \sigma^2(W_i)) \end{aligned} \quad (45)$$

for  $1 \leq i \leq n$ , where  $W_i$  is continuously distributed over  $\mathcal{W}$  with density uniformly bounded away from 0 and  $\infty$ . As in [Chen and Reiss \(2011\)](#), Theorem 3.2 is proved by (i) noting that the risk (in sup-norm loss) for the NPIV model is at least as large as the risk (in sup-norm loss) for the NPIR model, and (ii) calculating a lower bound (in sup-norm loss) for the NPIR model. Theorem 3.2 therefore follows from a sup-norm analogue of Lemma 1 of [Chen and Reiss \(2011\)](#) and Theorem G.1, which establishes a lower bound on minimax risk over Hölder classes under sup-norm loss for the NPIR model.  $\square$

**Theorem G.1.** *Let Condition LB hold for the NPIR model (45) with a random sample  $\{(W_i, Y_i)\}_{i=1}^n$ . Then for any  $0 \leq |\alpha| < p$ :*

$$\liminf_{n \rightarrow \infty} \inf_{\hat{g}_n} \sup_{h \in B_\infty(p, L)} \mathbb{P}_h \left( \|\hat{g}_n - \partial^\alpha h\|_\infty \geq c(n/\log n)^{-(p-|\alpha|)/(2(p+\varsigma)+d)} \right) \geq c' > 0$$

in the mildly ill-posed case, and

$$\liminf_{n \rightarrow \infty} \inf_{\hat{g}_n} \sup_{h \in B_\infty(p, L)} \mathbb{P}_h \left( \|\hat{g}_n - \partial^\alpha h\|_\infty \geq c(\log n)^{-(p-|\alpha|/\varsigma)} \right) \geq c' > 0$$

in the severely ill-posed case, where  $\inf_{\hat{g}_n}$  denotes the infimum over all estimators of  $\partial^\alpha h$  based on the sample of size  $n$ ,  $\sup_{h \in B_\infty(p, L)} \mathbb{P}_h$  denotes the sup over  $h \in B_\infty(p, L)$  and distributions  $(W_i, u_i)$  which satisfy Condition LB with  $\nu$  fixed, and the finite positive constants  $c, c'$  depend only on  $p, L, d, \varsigma$  and  $\sigma_0$ .

**Proof of Theorem G.1.** We establish the lower bound by applying Theorem 2.5 of [Tsybakov \(2009\)](#) (see Theorem G.2 below). We first explain the scalar ( $d = 1$ ) case in detail. Let  $\{\phi_{j,k}, \psi_{j,k}\}_{j,k}$  be a wavelet basis of regularity  $\gamma > p$  for  $L^2([0, 1])$  as described in Appendix E. Recall that this basis is generated by a Daubechies pair  $(\varphi, \psi)$  where  $\varphi$  has support  $[-N + 1, N]$ . We will define a family of submodels in which we perturb  $h_0$  by elements of the wavelet space  $W_j$ , where we choose  $j$  deterministically with  $n$ . For given  $j$ , recall that the wavelet space  $W_j$  consists of  $2^j$  functions  $\{\psi_{j,k}\}_{0 \leq k \leq 2^j - 1}$ , such that  $\{\psi_{j,k}\}_{r \leq k \leq 2^j - N - 1}$  are interior wavelets for which  $\psi_{j,k}(\cdot) = 2^{j/2}\psi(2^j(\cdot) - k)$ .

By construction, the support of each interior wavelet is an interval of length  $2^{-j}(2r - 1)$ . Thus for all  $j$  sufficiently large (hence the  $\liminf$  in our statement of the Lemma) we may choose a set  $M \subset \{r, \dots, 2^j - N - 1\}$  of interior wavelets with  $\#(M) \gtrsim 2^j$  such that  $\text{support}(\psi_{j,m}) \cap \text{support}(\psi_{j,m'}) = \emptyset$  for all  $m, m' \in M$  with  $m \neq m'$ . Note also that by construction we have  $\#(M) \leq 2^j$  (since there are  $2^j - 2N$  interior wavelets).

Recall the norms  $\|\cdot\|_{b_{\infty, \infty}^p}$  defined in Appendix E. Let  $h_0 \in B_\infty(p, L)$  be such that  $\|h_0\|_{B_{\infty, \infty}^p} \leq L/2$ , and for each  $m \in M$  let

$$h_m = h_0 + c_0 2^{-j(p+1/2)} \psi_{j,m}$$

where  $c_0$  is a positive constant to be defined subsequently. Noting that

$$\begin{aligned} c_0 2^{-j(p+1/2)} \|\psi_{j,m}\|_{B_{\infty,\infty}^p} &\lesssim c_0 2^{-j(p+1/2)} \|\psi_{j,m}\|_{b_{\infty,\infty}^p} \\ &\leq c_0 \end{aligned}$$

it follows by the triangle inequality that  $\|h_m\|_{B_{\infty,\infty}^p} \leq L$  uniformly in  $m$  for all sufficiently small  $c_0$ . By Condition LB, let  $W_i$  be distributed such that  $X_i$  has uniform marginal distribution on  $[0, 1]$ . For  $m \in \{0\} \cup M$  let  $P_m$  be the joint distribution of  $\{(W_i, Y_i)\}_{i=1}^n$  with  $Y_i = Th_m(W_i) + u_i$  for the Gaussian NPIR model (45).

For any  $m \in M$

$$\begin{aligned} \|\partial^\alpha h_0 - \partial^\alpha h_m\|_\infty &= c_0 2^{-j(p+1/2)} \|\partial^\alpha \psi_{j,m}\|_\infty \\ &= c_0 2^{-j(p-|\alpha|)} \|\psi^{(|\alpha|)}\|_\infty \end{aligned}$$

where  $\psi^{(|\alpha|)}$  denotes the  $|\alpha|$ th derivative of  $\psi$ . Moreover, for any  $m, m' \in M$  with  $m \neq m'$

$$\begin{aligned} \|\partial^\alpha h_m - \partial^\alpha h_{m'}\|_\infty &= c_0 2^{-j(p+1/2)} \|\partial^\alpha \psi_{j,m} - \partial^\alpha \psi_{j,m'}\|_\infty \\ &= 2c_0 2^{-j(p-|\alpha|)} \|\psi^{(|\alpha|)}\|_\infty \end{aligned}$$

by virtue of the disjoint support of  $\{\psi_{j,m}\}_{m \in M}$ .

By Condition LB(iii),

$$\|T\psi_{j,m}(W_i)\|_{L^2(W)} \lesssim \nu(2^j)^2 \langle \psi_{j,m}, \psi_{j,m} \rangle_X^2 = \nu(2^j)^2$$

(because  $c_0 2^{-j(p+1/2)} \psi_{j,m} \in \mathcal{H}_2(p, L)$  for sufficiently small  $c_0$ ) where  $\nu(2^j) = 2^{-j\varsigma}$  in the mildly ill-posed case and  $\nu(2^j) = \exp(-2^{j\varsigma})$  in the severely ill-posed case. The KL distance  $K(P_m, P_0)$  is

$$\begin{aligned} K(P_m, P_0) &\leq \frac{1}{2} \sum_{i=1}^n (c_0 2^{-j(p+1/2)})^2 E \left[ \frac{(T\psi_{j,m}(W_i))^2}{\sigma^2(W_i)} \right] \\ &\leq \frac{1}{2} \sum_{i=1}^n (c_0 2^{-j(p+1/2)})^2 \frac{E[(T\psi_{j,m}(W_i))^2]}{\underline{\sigma}^2} \\ &\lesssim n (c_0 2^{-j(p+1/2)})^2 \nu(2^j)^2. \end{aligned}$$

In the mildly ill-posed case ( $\nu(2^j) = 2^{-j\varsigma}$ ) we choose  $2^j \asymp (n/(\log n))^{1/(2(p+\varsigma)+1)}$ . This yields:

$$\begin{aligned} K(P_m, P_0) &\lesssim c_0^2 \log n \quad \text{uniformly in } m \\ \log(\#(M)) &\gtrsim \log n + \log \log n. \end{aligned}$$

since  $\#(M) \asymp 2^j$ .

In the severely ill-posed case ( $\nu(2^j) = \exp(-\frac{1}{2}2^{j\varsigma})$ ) we choose  $2^j = (c_1 \log n)^{1/\varsigma}$  with  $c_1 > 1$ . This yields:

$$\begin{aligned} K(P_m, P_0) &\lesssim n^{-(c_1-1)} \quad \text{uniformly in } m \\ \log(\#(M)) &\gtrsim \log \log n. \end{aligned}$$

In both the mildly and severely ill-posed cases, we may choose  $c_0$  sufficiently small that both  $\|h_m\|_{B_{\infty,\infty}^p} \leq L$  and  $K(P_m, P_0) \leq \frac{1}{8} \log(\#(M))$  hold uniformly in  $m$  for all  $n$  sufficiently large. All conditions of Theorem 2.5 of [Tsybakov \(2009\)](#) are satisfied and hence we obtain the lower bound result.

In the multivariate case ( $d > 1$ ) we let  $\tilde{\psi}_{j,k,G}(x)$  denote an orthonormal tensor-product wavelet for  $L^2([0, 1]^d)$  at resolution level  $j$  (see Appendix E). We construct a family of submodels analogously to the univariate case, setting  $h_m = h_0 + c_0 2^{-j(p+d/2)} \tilde{\psi}_{j,m,G}$  where  $\tilde{\psi}_{j,m,G}$  is now the tensor product of  $d$  interior univariate wavelets at resolution level  $j$  with  $G = (w_\psi)^d$  and where  $\#(M) \asymp 2^{jd}$ . By condition LB we obtain

$$\|\partial^\alpha h_m - \partial^\alpha h_{m'}\|_\infty \gtrsim c_0 2^{-j(p-|\alpha|)}$$

for each  $m, m' \in \{0\} \cup M$  with  $m \neq m'$ , and

$$K(P_m, P_0) \lesssim n(c_0 2^{-j(p+d/2)})^2 \nu(2^j)^2$$

for each  $m \in M$ , where  $\nu(2^j) = 2^{-j\varsigma}$  in the mildly ill-posed case and  $\nu(2^j) \asymp \exp(-2^{j\varsigma})$  in the severely ill-posed case. We choose  $2^j \asymp (n/\log n)^{1/(2(p+\varsigma)+d)}$  in the mildly ill-posed case and  $2^j = (c_1 \log n)^{1/\varsigma}$  in the severely ill-posed case. The result follows as in the univariate case.  $\square$

The following theorem is a special case of Theorem 2.5 on p. 99 of [Tsybakov \(2009\)](#) which we use to prove the minimax lower bounds in sup- and  $L^2$ -norm loss for  $h_0$  and its derivatives. We state the result here for convenience.

**Theorem G.2** ([Tsybakov \(2009\)](#)). *Assume that  $\#(M) \geq 2$  and suppose that  $(\mathcal{H}, \|\cdot\|_{\mathcal{H}})$  contains elements  $\{h_m : m \in \{0\} \cup M\}$  such that:*

- (i)  $\|\partial^\alpha h_m - \partial^\alpha h_{m'}\|_{\mathcal{H}} \geq 2s > 0$  for each  $m, m' \in M \cup \{0\}$  with  $m \neq m'$ ;
- (ii)  $P_m \ll P_0$  for each  $m \in M$  and

$$\frac{1}{\#(M)} \sum_{m \in M} K(P_m, P_0) \leq \mathbf{a} \log(\#(M))$$

with  $0 < \mathbf{a} < \frac{1}{8}$  and where  $P_m$  denotes the distribution of the data when  $h = h_m$  for each  $m \in \{0\} \cup M$ . Then:

$$\inf_{\hat{g}} \sup_{h \in \mathcal{H}} \mathbb{P}_h(\|\hat{g} - \partial^\alpha h\|_{\mathcal{H}} \geq s) \geq \frac{\sqrt{\#(M)}}{1 + \sqrt{\#(M)}} \left( 1 - 2\mathbf{a} - \sqrt{\frac{2\mathbf{a}}{\log(\#(M))}} \right) > 0.$$

### G.3 Proofs for Section 3.3

**Proof of Lemma 3.2.** We first prove **Result (1)**. Let  $P_{J-1,z} = \text{clsp}\{\phi_{01,z}, \dots, \phi_{0J-1,z}\}$  and let  $P_{J-1,z}^\perp$  denote its orthogonal complement in  $L^2(X_1|Z=z)$ . Observe that by definition of the singular values, for each  $z$  we have:

$$\begin{aligned} \sup_{h_z \in P_{J-1,z}^\perp : \|h_z\|_{L^2(X_1|Z=z)}=1} \|T_z h_z\|_{L^2(W_1|Z=z)}^2 &= \sup_{h_z \in P_{J-1,z}^\perp : \|h_z\|_{L^2(X_1|Z=z)}=1} \langle (T_z^* T_z) h_z, h_z \rangle_{X_1|Z=z} \\ &= \mu_{J,z}^2. \end{aligned} \tag{46}$$

Then let  $P_{J-1}^\perp = \{h(x_1, z) \in L^2(X) : h(\cdot, z) \in P_{J-1, z}^\perp \text{ for each } z\}$ . Note that  $\phi_{0j} \in \{h \in P_{J-1}^\perp : \|h(\cdot, z)\|_{L^2(X_1|Z=z)} = 1 \forall z\}$  for each  $j \geq J$ . Then:

$$\begin{aligned}
\tau_J^{-2} &= \inf_{h \in \Psi_J : \|h\|_{L^2(X)}=1} \|Th\|_{L^2(W)}^2 \\
&\leq \inf_{h \in \Psi_J \cap P_{J-1}^\perp : \|h\|_{L^2(X_1|Z=z)}=1 \forall z} \|Th\|_{L^2(W)}^2 \\
&\leq \sup_{h \in \Psi_J \cap P_{J-1}^\perp : \|h\|_{L^2(X_1|Z=z)}=1 \forall z} \|Th\|_{L^2(W)}^2 \\
&\leq \sup_{h \in P_{J-1}^\perp : \|h\|_{L^2(X_1|Z=z)}=1 \forall z} \|Th\|_{L^2(W)}^2.
\end{aligned} \tag{47}$$

Let  $F_Z$  denote the distribution of  $Z$ . For any  $h \in P_{J-1}^\perp$  let  $h_z(x_1) = h(x_1, z)$  and observe that  $h_z \in P_{J-1, z}^\perp$ . By iterated expectations and (46), for any  $h \in P_{J-1}^\perp$  with  $\|h_z\|_{L^2(X_1|Z=z)} = 1$  for each  $z$ , we have:

$$\begin{aligned}
\|Th\|_{L^2(W)}^2 &= \int \|E[h(X_{1i}, z)|W_{1i}, Z_i = z]\|_{L^2(W_1|Z=z)}^2 dF_Z(z) \\
&= \int \|T_z h_z\|_{L^2(W_1|Z=z)}^2 dF_Z(z) \\
&\leq \int \mu_{J,z}^2 \|h_z\|_{L^2(X_1|Z=z)}^2 dF_Z(z) \\
&= \int \mu_{J,z}^2 dF_Z(z) = E[\mu_{J,Z_i}^2].
\end{aligned} \tag{48}$$

It follows by substituting (48) into (47) that  $\tau_J \geq E[\mu_{J,Z_i}^2]^{-1/2}$ .

To prove **Result (2)**, note that any  $h \in \Psi_J$  with  $h \neq 0$  can be written as  $\sum_{j=1}^J a_j \phi_{0j}$  for constants  $a_j = a_j(h)$  where

$$\|h\|_{L^2(X)}^2 = E \left[ E \left[ \left( \sum_{j=1}^J a_j \phi_{0j}(X_{1i}, Z_i) \right)^2 \middle| Z_i \right] \right] = \sum_{j=1}^J a_j^2$$

since  $E[\phi_{0j,z}(X_i)\phi_{0k,z}(X_i)|Z_i = z] = \delta_{jk}$  where  $\delta_{jk}$  denotes the Kronecker delta. Moreover:

$$\begin{aligned}
\|Th\|_{L^2(W)}^2 &= E \left[ \left( E \left[ \sum_{j=1}^J a_j \phi_{0j}(X_{1i}, Z_i) \middle| W_{1i}, Z_i \right] \right)^2 \right] \\
&= E \left[ \left( E \left[ \sum_{j=1}^J a_j \phi_{0j, Z_i}(X_{1i}) \middle| W_{1i}, Z_i \right] \right)^2 \right] \\
&= E \left[ \left( \sum_{j=1}^J a_j \mu_{j, Z_i} \phi_{1j, Z_i}(W_{1i}) \right)^2 \right] \\
&= E \left[ E \left[ \left( \sum_{j=1}^J a_j \mu_{j, Z_i} \phi_{1j, Z_i}(W_{1i}) \right)^2 \middle| Z_i \right] \right] = \sum_{j=1}^J a_j^2 E[\mu_{j, Z_i}^2] \geq \|h\|_{L^2(X)}^2 E[\mu_{J, Z_i}^2]
\end{aligned}$$

since  $E[\phi_{1j,z}(W_{1i})\phi_{1k,z}(W_{1i})|Z_i = z] = \delta_{jk}$ . Therefore,

$$\tau_J = \sup_{h \in \Psi_J} \frac{\|h\|_{L^2(X)}}{\|Th\|_{L^2(W)}} \leq \frac{1}{E[\mu_{J,Z_i}^2]^{1/2}}$$

as required.  $\square$

## G.4 Proofs for Appendix D and Section 4

Since the proofs for uniform inference theories (in Section 4) built upon that for the pointwise normality Theorem D.1 (in Appendix D), we shall present the proof of Theorem D.1 first.

### G.4.1 Proofs for Appendix D

*Proof of Theorem D.1.* We first prove **Result (1)**. By Assumption 5'(a) or 5'(b)(i)(ii) we have:

$$\sqrt{n} \frac{(f(\hat{h}) - f(h_0))}{\sigma_n(f)} = \sqrt{n} \frac{Df(h_0)[\hat{h} - \tilde{h}]}{\sigma_n(f)} + o_p(1).$$

Define

$$Z_n(W_i) = \frac{(Df(h_0)[\psi^J])'[S'G_b^{-1}S]^{-1}S'G_b^{-1}b^K(W_i)}{\sigma_n(f)} = \Pi_K T u_n(f)(W_i)$$

where  $u_n(f) = v_n(f)/\sigma_n(f)$  is the scaled sieve 2SLS Riesz representer. Note that  $E[(Z_n(W_i)u_i)^2] = 1$ . Then

$$\begin{aligned} \sqrt{n} \frac{Df(h_0)[\hat{h} - \tilde{h}]}{\sigma_n(f)} &= \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_n(W_i)u_i \\ &\quad + \frac{(Df(h_0)[\psi^J])'((\hat{S}'\hat{G}_b^- \hat{S}) - \hat{S}'\hat{G}_b^- - (S'G_b^{-1}S)^{-1}S'G_b^{-1})(B'u/\sqrt{n})}{\sigma_n(f)} \\ &=: T_1 + T_2. \end{aligned}$$

We first show  $T_1 \rightarrow_d N(0, 1)$  by the Lindeberg-Feller theorem. To verify the Lindeberg condition, note that

$$|Z_n(W_i)| \leq \frac{\left\| (Df(h_0)[\psi^J])'(S'G_b^{-1}S)^{-1}S'G_b^{-1/2} \right\| \left\| G_b^{-1/2}b^K(W_i) \right\|}{(\inf_w E[u_i^2|W_i = w])^{1/2} \left\| (Df(h_0)[\psi^J])'(S'G_b^{-1}S)^{-1}S'G_b^{-1/2} \right\|} \leq \underline{\sigma}^{-1}\zeta_b(K)$$

by the Cauchy-Schwarz inequality and Assumption 2(iii). Therefore,

$$E[u_i^2 Z_n(W_i)^2 \{ |Z_n(W_i)u_i| > \eta\sqrt{n} \}] \leq \sup_w E[u_i^2 \{ |u_i| \gtrsim \eta(\sqrt{n}/\zeta_b(K)) \} | W_i = w] = o(1)$$

by Assumption 2(iv') and the condition on  $J$ . Therefore,  $T_1 \rightarrow_d N(0, 1)$ .

For  $T_2$ , observe that

$$\begin{aligned}
|T_2| &= \left| \frac{(Df(h_0)[\psi^J])'((\widehat{G}_b^{-1/2}\widehat{S})_l^- \widehat{G}_b^{-1/2}G_b^{1/2} - (G_b^{-1/2}S)_l^-)(G_b^{-1/2}B'u/\sqrt{n})}{\sigma_n(f)} \right| \\
&= \left| \frac{[(Df(h_0)[\psi^J])'(G_b^{-1/2}S)_l^-]G_b^{-1/2}S\{(\widehat{G}_b^{-1/2}\widehat{S})_l^- \widehat{G}_b^{-1/2}G_b^{1/2} - (G_b^{-1/2}S)_l^-\}(G_b^{-1/2}B'u/\sqrt{n})}{\sigma_n(f)} \right| \\
&\leq \frac{\left\| (Df(h_0)[\psi^J])'(G_b^{-1/2}S)_l^- \right\| \left\| G_b^{-1/2}S\{(\widehat{G}_b^{-1/2}\widehat{S})_l^- \widehat{G}_b^{-1/2}G_b^{1/2} - (G_b^{-1/2}S)_l^-\} \right\| \left\| G_b^{-1/2}B'u/\sqrt{n} \right\|}{(\inf_w E[u_i^2|W_i = w])^{1/2} \left\| (Df(h_0)[\psi^J])'(G_b^{-1/2}S)_l^- \right\|} \\
&\leq \underline{\sigma}^{-1} \left\| G_b^{-1/2}S\{(\widehat{G}_b^{-1/2}\widehat{S})_l^- \widehat{G}_b^{-1/2}G_b^{1/2} - (G_b^{-1/2}S)_l^-\} \right\| \left\| G_b^{-1/2}B'u/\sqrt{n} \right\| \\
&= O_p(s_{JK}^{-1}\zeta\sqrt{(J\log J)/n})
\end{aligned}$$

where the first inequality is by the Cauchy-Schwarz inequality, the second is by Assumption 2(iii), and the final line is by Lemmas F.10(c) and F.8. The result follows by the equivalence  $\tau_J \asymp s_{JK}^{-1}$  (see Lemma A.1) and the condition  $\tau_J\zeta\sqrt{(J\log n)/n} = o(1)$ .

**Result (2)** follows directly from Result (1) and Lemma G.2.  $\square$

**Lemma G.2.** *Let Assumptions 1(iii), 2(i)-(iii), 3(iii) and 4(i) hold,  $\tau_J\zeta\sqrt{(\log n)/n} = o(1)$ , and Assumption 5'(b)(iii) hold (with  $\eta'_n = 0$  if  $f(\cdot)$  is linear). Let  $\|\widehat{h} - h_0\|_\infty = O_p(\delta_{h,n}) = o_p(1)$ , and  $\delta_{V,n} \equiv [\zeta_{b,K}^{(2+\delta)/\delta}\sqrt{(\log K)/n}]^{\delta/(1+\delta)} + \tau_J\zeta\sqrt{(\log J)/n} + \delta_{h,n}$ . Then:*

$$\left| \frac{\widehat{\sigma}_n(f)}{\sigma_n(f)} - 1 \right| = O_p(\delta_{V,n} + \eta'_n) = o_p(1).$$

**Proof of Lemma G.2.** First write

$$\begin{aligned}
\frac{\widehat{\sigma}_n(f)^2}{\sigma_n(f)^2} - 1 &= \left( \frac{\widehat{\gamma}'_n \Omega^\circ \widehat{\gamma}_n}{\sigma_n(f)^2} - 1 \right) + \frac{\widehat{\gamma}'_n (\widehat{\Omega}^\circ - \Omega^\circ) \widehat{\gamma}_n}{\sigma_n(f)^2} \\
&= \left( \frac{(\widehat{\gamma}_n - \gamma_n)' \Omega^\circ (\widehat{\gamma}_n + \gamma_n)}{\sigma_n(f)^2} \right) + \frac{\widehat{\gamma}'_n (\widehat{\Omega}^\circ - \Omega^\circ) \widehat{\gamma}_n}{\sigma_n(f)^2} =: T_1 + T_2
\end{aligned}$$

where

$$\begin{aligned}
\widehat{\Omega}^\circ &= G_b^{-1/2} \widehat{\Omega} G_b^{-1/2} & \widehat{\gamma}_n &= G_b^{1/2} \widehat{G}_b^{-1} \widehat{S} [\widehat{S}' \widehat{G}_b^{-1} \widehat{S}]^{-1} Df(\widehat{h})[\psi^J] \\
\Omega^\circ &= G_b^{-1/2} \Omega G_b^{-1/2} & \gamma_n &= G_b^{-1/2} S [S' G_b^{-1} S]^{-1} Df(h_0)[\psi^J]
\end{aligned}$$

and observe that  $\gamma'_n \Omega^\circ \gamma_n = \sigma_n(f)^2$  and  $\widehat{\gamma}'_n \widehat{\Omega}^\circ \widehat{\gamma}_n = \widehat{\sigma}_n(f)^2$ .

Control of  $T_1$ : We first show that

$$\frac{\|\widehat{\gamma}_n - \gamma_n\|_{\ell^2}}{\sigma_n(f)} = O_p(\tau_J\zeta\sqrt{(\log J)/n} + \eta'_n) = o_p(1). \quad (49)$$

To simplify notation, let

$$\partial = \frac{Df(h_0)[\psi^J]}{\sigma_n(f)} \quad \text{and} \quad \widehat{\partial} = \frac{Df(\widehat{h})[\psi^J]}{\sigma_n(f)}$$

and note that  $\|\partial'(G_b^{-1/2}S)_l^-\|_{\ell^2} = s_n(f)/\sigma_n(f) \asymp 1$  under Assumptions 2(i)(iii) and that  $\widehat{\partial} = \partial$  if  $f(\cdot)$  is linear.

Then we have:

$$\begin{aligned}
\frac{\|\widehat{\gamma}_n - \gamma_n\|_{\ell^2}}{\sigma_n(f)} &= \|\widehat{\partial}'(\widehat{G}_b^{-1/2}\widehat{S})_l^- \widehat{G}_b^{-1/2} G_b^{1/2} - \partial'(G_b^{-1/2}S)_l^-\|_{\ell^2} \\
&\leq \|\widehat{\partial}'(G_b^{-1/2}S)_l^-\|_{\ell^2} \|G_b^{-1/2}S\{(\widehat{G}_b^{-1/2}\widehat{S})_l^- \widehat{G}_b^{-1/2} G_b^{1/2} - (G_b^{-1/2}S)_l^-\}\|_{\ell^2} + \|(\widehat{\partial}' - \partial')(G_b^{-1/2}S)_l^-\|_{\ell^2} \\
&= O_p(1) \times O_p(s_{JK}^{-1}\zeta\sqrt{(\log J)/n}) + \frac{\|\Pi_K T(\widehat{v}_n(f) - v_n(f))\|_{L^2(W)}}{\sigma_n(f)} \\
&= O_p(1) \times O_p(s_{JK}^{-1}\zeta\sqrt{(\log J)/n}) + O_p(\eta'_n)
\end{aligned}$$

where the third line is Lemma F.10(c) and the final line is by Assumption 5'(b)(iii). Therefore, (49) holds by the equivalence  $s_{JK}^{-1} \asymp \tau_J$  (Lemma A.1) and the condition  $\tau_J \zeta \sqrt{(\log n)/n} = o(1)$ .

Finally, since all eigenvalues of  $\Omega^o$  are bounded between  $\underline{\sigma}^2$  and  $\overline{\sigma}^2$  under Assumption 2(i)(iii), it follows from (49) and Cauchy-Schwarz that  $|T_1| = o_p(1)$ .

Control of  $T_2$ : Equation (49) implies that  $\|\widehat{\gamma}_n\|/\sigma_n(f) = O_p(1)$ . Therefore,  $|T_2| \leq O_p(1) \times \|\widehat{\Omega}^o - \Omega^o\|_{\ell^2} = o_p(1)$  by Lemma G.3.  $\square$

**Lemma G.3.** *Let Assumptions 2(i)(ii) hold, let  $\zeta_{b,K}\sqrt{(\log K)/n} = o(1)$ , and let  $\|\widehat{h} - h_0\|_{\infty} = O_p(\delta_{h,n})$  with  $\delta_{h,n} = o(1)$ . Then:*

$$\|\widehat{\Omega}^o - \Omega^o\|_{\ell^2} = O_p\left(\left(\zeta_{b,K}^{(2+\delta)/\delta}\sqrt{(\log K)/n}\right)^{\delta/(1+\delta)} + \delta_{h,n}\right)$$

**Proof of Lemma G.3.** By the triangle inequality:

$$\begin{aligned}
\|\widehat{\Omega}^o - \Omega^o\|_{\ell^2} &\leq \left\| G_b^{-1/2} \left( \frac{1}{n} \sum_{i=1}^n u_i^2 b^K(W_i) b^K(W_i)' \right) G_b^{-1/2} \right\|_{\ell^2} \\
&\quad + \left\| G_b^{-1/2} \left( \frac{1}{n} \sum_{i=1}^n 2u_i(\widehat{u}_i - u_i) b^K(W_i) b^K(W_i)' \right) G_b^{-1/2} \right\|_{\ell^2} \\
&\quad + \left\| G_b^{-1/2} \left( \frac{1}{n} \sum_{i=1}^n (\widehat{u}_i - u_i)^2 b^K(W_i) b^K(W_i)' \right) G_b^{-1/2} \right\|_{\ell^2} \\
&\leq O_p\left(\left(\zeta_{b,K}^{(2+\delta)/\delta}\sqrt{(\log K)/n}\right)^{\delta/(1+\delta)}\right) + \|\widehat{h} - h_0\|_{\infty} \times O_p(1) + \|\widehat{h} - h_0\|_{\infty}^2 \times O_p(1)
\end{aligned}$$

where the first term may easily be deduced from the proof of Lemma 3.1 of Chen and Christensen (2015), the second then follows because  $2u_i(\widehat{u}_i - u_i) \leq (1 + u_i^2)\|\widehat{h} - h_0\|_{\infty}$ , and the third follows similarly because  $\|\widehat{G}_b^o\|_{\ell^2} = O_p(1)$  by Lemma F.7.  $\square$

## G.4.2 Proofs for Section 4

**Proof of Lemma 4.1.** Recall that

$$\begin{aligned}
\widehat{\mathbb{Z}}_n(t) &= \frac{(Df_t(h_0)[\psi^J])'[S'G_b^{-1}S]^{-1}S'G_b^{-1/2}}{\sigma_n(f_t)} \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n G_b^{-1/2} b^K(W_i) u_i \right), \\
\mathbb{Z}_n(t) &= \frac{(Df_t(h_0)[\psi^J])'[S'G_b^{-1}S]^{-1}S'G_b^{-1/2}}{\sigma_n(f_t)} \mathbb{Z}_n \quad \text{where } \mathbb{Z}_n \sim N(0, \Omega^o) \text{ with } \Omega^o = G_b^{-1/2} \Omega G_b^{-1/2}.
\end{aligned}$$

**Step 1: Uniform Bahadur representation.** By Assumption 5(a) or (b)(i)(ii), we have

$$\begin{aligned}
\sup_{t \in \mathcal{T}} \left| \sqrt{n} \frac{f_t(\hat{h}) - f_t(h_0)}{\hat{\sigma}_n(f_t)} - \hat{\mathbb{Z}}_n(t) \right| &\leq \sup_{t \in \mathcal{T}} \left| \sqrt{n} \frac{Df_t(h_0)[\hat{h} - \tilde{h}]}{\sigma_n(f_t)} - \hat{\mathbb{Z}}_n(t) \right| + O_p(\eta_n) \times \sup_{t \in \mathcal{T}} \left| \frac{\sigma_n(f_t)}{\hat{\sigma}_n(f_t)} \right| \\
&\quad + \sup_{t \in \mathcal{T}} \left| \frac{\sigma_n(f_t)}{\hat{\sigma}_n(f_t)} - 1 \right| \times \sup_{t \in \mathcal{T}} \left| \sqrt{n} \frac{Df_t(h_0)[\hat{h} - \tilde{h}]}{\sigma_n(f_t)} \right| \\
&=: T_1 + T_2 + T_3.
\end{aligned}$$

Control of  $T_1$ : As in the proof of Theorem D.1,

$$\begin{aligned}
T_1 &= \sup_{t \in \mathcal{T}} \left| \frac{(Df_t(h_0)[\psi^J])'((\hat{G}_b^{-1/2}\hat{S})_l^- \hat{G}_b^{-1/2} G_b^{1/2} - (G_b^{-1/2} S)_l^-)(G_b^{-1/2} B'u/\sqrt{n})}{\sigma_n(f_t)} \right| \\
&= \sup_{t \in \mathcal{T}} \left| \frac{[(Df_t(h_0)[\psi^J])'(G_b^{-1/2} S)_l^-] G_b^{-1/2} S \{(\hat{G}_b^{-1/2}\hat{S})_l^- \hat{G}_b^{-1/2} G_b^{1/2} - (G_b^{-1/2} S)_l^-\} (G_b^{-1/2} B'u/\sqrt{n})}{\sigma_n(f_t)} \right| \\
&\leq \sup_{t \in \mathcal{T}} \frac{\left\| (Df_t(h_0)[\psi^J])'(G_b^{-1/2} S)_l^- \right\| \left\| G_b^{-1/2} S \{(\hat{G}_b^{-1/2}\hat{S})_l^- \hat{G}_b^{-1/2} G_b^{1/2} - (G_b^{-1/2} S)_l^-\} \right\| \left\| G_b^{-1/2} B'u/\sqrt{n} \right\|}{(\inf_w E[u_i^2 | W_i = w])^{1/2} \left\| (Df_t(h_0)[\psi^J])'(G_b^{-1/2} S)_l^- \right\|} \\
&\leq \underline{\sigma}^{-1} \left\| G_b^{-1/2} S \{(\hat{G}_b^{-1/2}\hat{S})_l^- \hat{G}_b^{-1/2} G_b^{1/2} - (G_b^{-1/2} S)_l^-\} \right\| \left\| G_b^{-1/2} B'u/\sqrt{n} \right\| \\
&= O_p(\tau_J \zeta \sqrt{(J \log J)/n}) = o_p(r_n)
\end{aligned}$$

where the first inequality is by the Cauchy-Schwarz inequality, the second is by Assumption 2(iii), and the final line is by Lemmas F.10(c) and F.8 and the equivalence  $\tau_J \asymp s_{JK}^{-1}$  (see Lemma A.1), and the last  $o_p(r_n)$  is by Assumption 6(ii.2).

Control of  $T_2$ : Lemma G.4 below shows that

$$\sup_{t \in \mathcal{T}} \left| \frac{\hat{\sigma}_n(f_t)}{\sigma_n(f_t)} - 1 \right| = O_p(\delta_{V,n} + \eta'_n) = o_p(1)$$

from which it follows that  $T_2 = O_p(\eta_n) \times O_p(1) = O_p(\eta_n)$ .

Control of  $T_3$ : By Lemma G.4 below and the bound for  $T_1$ , we have:

$$\begin{aligned}
T_3 &= O_p(\delta_{V,n} + \eta'_n) \times \sup_{t \in \mathcal{T}} \left| \sqrt{n} \frac{Df_t(h_0)[\hat{h} - \tilde{h}]}{\sigma_n(f_t)} \right| \\
&= O_p(\delta_{V,n} + \eta'_n) \times \left[ \sup_{t \in \mathcal{T}} |\hat{\mathbb{Z}}_n(t)| + o_p(r_n) \right] \\
&= O_p(\delta_{V,n} + \eta'_n) \times \left[ o_p(r_n) + \sup_{t \in \mathcal{T}} |\mathbb{Z}_n(t)| + o_p(r_n) \right] \\
&= O_p(\delta_{V,n} + \eta'_n) \times \left[ o_p(r_n) + O_p(c_n) \right]
\end{aligned}$$

where the second-last line is by display (54) step 2 below and the final line is by Lemma G.5 below. Therefore we have proved:

$$\begin{aligned}
\sup_{t \in \mathcal{T}} \left| \sqrt{n} \frac{f_t(\hat{h}) - f_t(h_0)}{\hat{\sigma}_n(f_t)} - \hat{\mathbb{Z}}_n(t) \right| &= O_p(\tau_J \zeta \sqrt{(J \log J)/n}) + O_p(\eta_n) + O_p(\delta_{V,n} + \eta'_n) \times [o_p(r_n) + O_p(c_n)] \\
&= o_p(r_n)
\end{aligned} \tag{50}$$

where the final line is by Assumption 6(ii.2).

**Step 2: Approximating  $\widehat{\mathbb{Z}}_n(t)$  by a Gaussian process  $\mathbb{Z}_n(t)$ .** We use Yurinskii's coupling (Pollard, 2002, Theorem 10, p. 244) to show that there exists a sequence of  $N(0, \Omega^o)$  random vectors  $\mathcal{Z}_n$  such that

$$\left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n G_b^{-1/2} b^K(W_i) u_i - \mathcal{Z}_n \right\|_{\ell^2} = o_p(r_n). \quad (51)$$

By Assumption 2(iv) we have

$$\sum_{i=1}^n E[\|n^{-1/2} G_b^{-1/2} b^K(W_i) u_i\|_{\ell^2}^3] \lesssim n^{-1/2} \zeta_{b,K} E[\|G_b^{-1/2} b^K(W_i)\|_{\ell^2}^2] = \frac{\zeta_{b,K} K}{\sqrt{n}} = O\left(\frac{\zeta_{b,K} J}{\sqrt{n}}\right).$$

Existence of  $\mathcal{Z}_n$  follows under the condition (Assumption 6(ii.1))

$$\frac{\zeta_{b,K} J^2}{r_n^3 \sqrt{n}} = o(1).$$

The process  $\mathbb{Z}_n(t)$  is a centered Gaussian process with the covariance function

$$E[\mathbb{Z}_n(t_1) \mathbb{Z}_n(t_2)] = \frac{(Df_{t_1}(h_0)[\psi^J])' [S' G_b^{-1} S]^{-1} S' G_b^{-1} \Omega G_b^{-1} S [S' G_b^{-1} S]^{-1} Df_{t_2}(h_0)[\psi^J]}{\sigma_n(f_{t_1}) \sigma_n(f_{t_2})}. \quad (52)$$

Now observe that

$$\sup_{t \in \mathcal{T}} \left\| \frac{(Df_t(h_0)[\psi^J])' [S' G_b^{-1} S]^{-1} S' G_b^{-1/2}}{\sigma_n(f_t)} \right\|_{\ell^2} = \sup_{t \in \mathcal{T}} \frac{s_n(f_t)}{\sigma_n(f_t)} \asymp 1 \quad (53)$$

by Assumption 2(i)(iii). Therefore,

$$\sup_{t \in \mathcal{T}} \left| \widehat{\mathbb{Z}}_n(t) - \mathbb{Z}_n(t) \right| = o_p(r_n) \quad (54)$$

by equations (51) and (53) and Cauchy-Schwarz.  $\square$

**Lemma G.4.** *Let Assumptions 1(iii), 2(i)–(iii), 3(ii)(iii) and 4(i) hold,  $\tau_J \zeta \sqrt{(\log n)/n} = o(1)$ , and Assumption 5(b)(iii) hold (with  $\eta'_n = 0$  if  $f_t(\cdot)$  is linear). Let  $\|\widehat{h} - h_0\|_\infty = O_p(\delta_{h,n})$  with  $\delta_{h,n} = o(1)$ . Then:*

$$\sup_{t \in \mathcal{T}} \left| \frac{\sigma_n(f_t)}{\widehat{\sigma}_n(f_t)} - 1 \right| = O_p(\delta_{V,n} + \eta'_n) = o_p(1).$$

**Proof of Lemma G.4.** The proof follows by identical arguments to the proof of Lemma G.2.  $\square$

**Proof of Theorem 4.1.** Recall that

$$\mathbb{Z}_n^*(t) = \frac{(Df_t(\widehat{h})[\psi^J])' [\widehat{S}' \widehat{G}_b^{-1} \widehat{S}]^{-1} \widehat{S}' \widehat{G}_b^{-1}}{\widehat{\sigma}(f_t)} \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n b^K(W_i) \widehat{u}_i \varpi_i \right) \quad \text{for each } t \in \mathcal{T}.$$

**Step 1: Approximating  $\mathbb{Z}_n^*(t)$  by a Gaussian process  $\widetilde{\mathbb{Z}}_n^*$ .** Each of the terms  $n^{-1/2} G_b^{-1/2} b^K(W_i) \widehat{u}_i \varpi_i$  is centered under  $\mathbb{P}^*$  because  $E[\varpi_i | Z^n] = 0$  for  $i = 1, \dots, n$ . Moreover,

$$\sum_{i=1}^n E[(n^{-1/2} G_b^{-1/2} b^K(W_i) \widehat{u}_i \varpi_i)(n^{-1/2} G_b^{-1/2} b^K(W_i) \widehat{u}_i \varpi_i)' | Z^n] = \widehat{\Omega}^o$$

where  $G_b^{-1/2}\widehat{\Omega}G_b^{-1/2} = \widehat{\Omega}^o$ , and

$$\sum_{i=1}^n E[\|n^{-1/2}G_b^{-1/2}b^K(W_i)\widehat{u}_i\varpi_i\|_{\ell^2}^3 | Z^n] \lesssim n^{-3/2} \sum_{i=1}^n E[\|G_b^{-1/2}b^K(W_i)\|_{\ell^2}^2 |\widehat{u}_i|^3]$$

because  $E[|\varpi_i|^3 | Z^n] < \infty$  uniformly in  $i$ , and where

$$n^{-3/2} \sum_{i=1}^n E[\|G_b^{-1/2}b^K(W_i)\|_{\ell^2}^2 |\widehat{u}_i|^3] \lesssim \frac{\zeta_{b,K}K}{\sqrt{n}}$$

holds wpa1 (by Markov's inequality using  $|\widehat{u}_i|^3 \lesssim |u_i|^3 + \|\widehat{h} - h_0\|_\infty^3$  and Assumption 2(iv)). A second application of Yurinskii's coupling conditional on the data  $Z^n$  then yields existence of a sequence of  $N(0, \widehat{\Omega}^o)$  random vectors  $\mathcal{Z}_n^*$  such that

$$\left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n G_b^{-1/2}b^K(W_i)\widehat{u}_i\varpi_i - \mathcal{Z}_n^* \right\|_{\ell^2} = o_{p^*}(r_n)$$

wpa1. Therefore:

$$\sup_{t \in \mathcal{T}} \left| \mathbb{Z}_n^*(t) - \frac{(Df_t(\widehat{h})[\psi^J])'[\widehat{S}'\widehat{G}_b^{-1}\widehat{S}]^{-1}\widehat{S}'\widehat{G}_b^{-1/2}}{\widehat{\sigma}_n(f_t)} \mathcal{Z}_n^* \right| = o_{p^*}(r_n) \quad (55)$$

wpa1. Now observe that we can define a centered Gaussian process  $\widetilde{\mathbb{Z}}_n^*$  under  $\mathbb{P}^*$  by

$$\widetilde{\mathbb{Z}}_n^*(t) \equiv \frac{(Df_t(h_0)[\psi^J])'[\widehat{S}'G_b^{-1}S]^{-1}\widehat{S}'G_b^{-1/2}}{\sigma_n(f_t)} (\Omega^o)^{1/2}(\widehat{\Omega}^o)^{-1/2} \mathcal{Z}_n^*$$

which has the same covariance function as  $\mathbb{Z}_n$  (see equation (52)) whenever  $\widehat{\Omega}^o$  is invertible (which it is wpa1). Therefore, by Lemma G.6 below we have:

$$\sup_{t \in \mathcal{T}} \left| \frac{(Df_t(\widehat{h})[\psi^J])'[\widehat{S}'\widehat{G}_b^{-1}\widehat{S}]^{-1}\widehat{S}'\widehat{G}_b^{-1/2}}{\widehat{\sigma}_n(f_t)} \mathcal{Z}_n^* - \widetilde{\mathbb{Z}}_n^*(t) \right| = o_{p^*}(r_n) \quad (56)$$

wpa1. It follows from equations (55) and (56) and Assumption 6(ii) that

$$\sup_{t \in \mathcal{T}} \left| \mathbb{Z}_n^*(t) - \widetilde{\mathbb{Z}}_n^*(t) \right| = o_{p^*}(r_n) + o_{p^*}(r_n) = o_{p^*}(r_n) \quad (57)$$

wpa1.

**Step 2: Consistency.** By Lemma 4.1 and display (54), we have:

$$\sup_{t \in \mathcal{T}} \left| \frac{\sqrt{n}(f_t(\widehat{h}) - f_t(h_0))}{\widehat{\sigma}_n(f_t)} - \mathbb{Z}_n(t) \right| = o_p(r_n) + o_p(r_n) = o_p(r_n).$$

Therefore, we may choose a sequence of positive constants  $\epsilon_n$  with  $\epsilon_n = o(1)$  such that

$$\sup_{t \in \mathcal{T}} \left| \frac{\sqrt{n}(f_t(\widehat{h}) - f_t(h_0))}{\widehat{\sigma}_n(f_t)} - \mathbb{Z}_n(t) \right| \leq \epsilon_n r_n \quad (58)$$

holds wpa1. By an anti-concentration inequality (Chernozhukov, Chetverikov, and Kato, 2014, Theorem 2.1)

and Lemma G.5 below, we have:

$$\sup_{s \in \mathbb{R}} \mathbb{P} \left( \sup_{t \in \mathcal{T}} |\mathbb{Z}_n(t) - s| \leq \epsilon_n r_n \right) \lesssim \epsilon_n r_n E[\sup_{t \in \mathcal{T}} |\mathbb{Z}_n(t)|] \lesssim \epsilon_n r_n c_n = o(1)$$

due to  $r_n c_n \lesssim 1$  (Assumption 6(ii.1)). This, together with (58), yields:

$$\sup_{s \in \mathbb{R}} \left| \mathbb{P} \left( \sup_{t \in \mathcal{T}} \left| \frac{\sqrt{n}(f_t(\hat{h}) - f_t(h_0))}{\hat{\sigma}_n(f_t)} \right| \leq s \right) - \mathbb{P} \left( \sup_{t \in \mathcal{T}} |\mathbb{Z}_n(t)| \leq s \right) \right| = o(1). \quad (59)$$

Moreover, by (57) we may choose a sequence of positive constants  $\epsilon'_n$  with  $\epsilon'_n = o(1)$  such that

$$\sup_{t \in \mathcal{T}} \left| \mathbb{Z}_n^*(t) - \tilde{\mathbb{Z}}_n^*(t) \right| \leq \epsilon'_n r_n$$

holds wpa1. Similar arguments then yield:

$$\sup_{s \in \mathbb{R}} \mathbb{P}^* \left( \sup_{t \in \mathcal{T}} |\tilde{\mathbb{Z}}_n^*(t) - s| \leq \epsilon'_n r_n \right) \lesssim \epsilon'_n = o(1)$$

wpa1. This, together with equation (57), yields:

$$\sup_{s \in \mathbb{R}} \left| \mathbb{P}^* \left( \sup_{t \in \mathcal{T}} |\mathbb{Z}_n^*(t)| \leq s \right) - \mathbb{P}^* \left( \sup_{t \in \mathcal{T}} |\tilde{\mathbb{Z}}_n^*(t)| \leq s \right) \right| = o_p(1). \quad (60)$$

The result is immediate from equations (59) and (60) and the fact that

$$\mathbb{P} \left( \sup_{t \in \mathcal{T}} |\mathbb{Z}_n(t)| \leq s \right) = \mathbb{P}^* \left( \sup_{t \in \mathcal{T}} |\tilde{\mathbb{Z}}_n^*(t)| \leq s \right) \quad \text{wpa1 in } \mathbb{P}$$

holds uniformly in  $s$ . □

**Lemma G.5.** *Let Assumption 6(i) hold. Then:  $E[\sup_{t \in \mathcal{T}} |\mathbb{Z}_n(t)|] \lesssim c_n$  and  $\sup_{t \in \mathcal{T}} |\mathbb{Z}_n(t)| = O_p(c_n)$ .*

**Proof of Lemma G.5.** Observe that  $d_n(t_1, t_2) := E[(\mathbb{Z}_n(t_1) - \mathbb{Z}_n(t_2))^2]^{1/2}$ . By Corollary 2.2.8 of van der Vaart and Wellner (1996) and Assumption 6(i), there exists a universal constant  $C$  such that

$$E[\sup_t |\mathbb{Z}_n(t)|] \leq E[|\mathbb{Z}_n(\bar{t})|] + C \int_0^\infty \sqrt{\log N(\mathcal{T}, d_n, \epsilon)} d\epsilon$$

for any  $\bar{t} \in \mathcal{T}$ , where  $E[|\mathbb{Z}_n(\bar{t})|] = \sqrt{2/\pi}$  because  $\mathbb{Z}_n(\bar{t}) \sim N(0, 1)$ . Therefore,  $E[\sup_t |\mathbb{Z}_n(t)|] \lesssim c_n$ . The second result follows by Markov's inequality. □

**Lemma G.6.** *Let Assumptions 1(iii), 2, 3(ii)(iii), 4(i) and 6 hold,  $\tau_J \zeta \sqrt{(\log n)/n} = o(1)$  and  $\|\hat{h} - h_0\|_\infty = O_p(\delta_{h,n})$  with  $\delta_{h,n} = o(1)$ . Let Assumption 5(b)(iii) hold with  $\eta'_n \sqrt{J} = o(r_n)$  for nonlinear  $f_t(\cdot)$ . Let  $\mathcal{Z}_n^*$  and  $\tilde{\mathbb{Z}}_n^*(t)$  be as in the proof of Theorem 4.1. Then:*

$$\sup_{t \in \mathcal{T}} \left| \frac{(Df_t(\hat{h})[\psi^J])' [\hat{S}' \hat{G}_b^{-1} \hat{S}]^{-1} \hat{S}' \hat{G}_b^{-1/2}}{\hat{\sigma}_n(f_t)} \mathcal{Z}_n^* - \tilde{\mathbb{Z}}_n^*(t) \right| = o_{p^*}(r_n) \quad \text{wpa1 in } \mathbb{P}.$$

**Proof of Lemma G.6.** First note that because  $\mathcal{Z}_n^* \sim N(0, \hat{\Omega}^o)$  and the minimum and maximum eigenvalues

of  $\widehat{\Omega}^\circ$  are uniformly bounded away from 0 and  $\infty$  wpa1 (by Lemma G.3 and Assumptions 2(i)(iii)), we have  $\|\mathcal{Z}_n^*\| = O_{p^*}(\sqrt{K})$  wpa1 by Chebyshev's inequality.

Now, writing out term by term we have:

$$\begin{aligned}
& \sup_{t \in \mathcal{T}} \left| \left( \frac{(Df_t(\widehat{h}))[\psi^J]}' [\widehat{S}' \widehat{G}_b^{-1} \widehat{S}]^{-1} \widehat{S}' \widehat{G}_b^{-1/2}}{\widehat{\sigma}_n(f_t)} - \frac{(Df_t(h_0))[\psi^J]}' [S' G_b^{-1} S]^{-1} S' G_b^{-1/2} (\Omega^\circ)^{1/2} (\widehat{\Omega}^\circ)^{-1/2}}{\sigma_n(f_t)} \right) \mathcal{Z}_n^* \right| \\
& \leq \sup_{t \in \mathcal{T}} \left| \frac{\left( (Df_t(\widehat{h}))[\psi^J] - Df_t(h_0)[\psi^J] \right)' (G_b^{-1/2} S)_l^{-1} G_b^{-1/2} S [\widehat{S}' \widehat{G}_b^{-1} \widehat{S}]^{-1} \widehat{S}' \widehat{G}_b^{-1/2}}{\widehat{\sigma}_n(f_t)} \mathcal{Z}_n^* \right| \\
& + \sup_{t \in \mathcal{T}} \left| \frac{(Df_t(h_0))[\psi^J]}' \left( [\widehat{S}' \widehat{G}_b^{-1} \widehat{S}]^{-1} \widehat{S}' \widehat{G}_b^{-1/2} - [S' G_b^{-1} S]^{-1} S' G_b^{-1/2} (\Omega^\circ)^{1/2} (\widehat{\Omega}^\circ)^{-1/2} \right)}{\sigma_n(f_t)} \mathcal{Z}_n^* \right| \times \sup_{t \in \mathcal{T}} \frac{\sigma_n(f_t)}{\widehat{\sigma}_n(f_t)} \\
& + \sup_{t \in \mathcal{T}} \left| \frac{\sigma_n(f_t)}{\widehat{\sigma}_n(f_t)} - 1 \right| \times \sup_{t \in \mathcal{T}} \left| \frac{(Df_t(h_0))[\psi^J]}' [S' G_b^{-1} S]^{-1} S' G_b^{-1/2} (\Omega^\circ)^{1/2} (\widehat{\Omega}^\circ)^{-1/2}}{\sigma_n(f_t)} \mathcal{Z}_n^* \right| =: T_1 + T_2 + T_3.
\end{aligned}$$

Control of  $T_1$ : By Cauchy-Schwarz, we have:

$$\begin{aligned}
T_1 & \leq \sup_{t \in \mathcal{T}} \frac{\|\Pi_K T(\widehat{v}_n(f_t) - v_n(f_t))\|_{L^2(W)}}{\sigma_n(f_t)} \times \sup_{t \in \mathcal{T}} \frac{\sigma_n(f_t)}{\widehat{\sigma}_n(f_t)} \times \left\| G_b^{-1/2} S [\widehat{S}' \widehat{G}_b^{-1} \widehat{S}]^{-1} \widehat{S}' \widehat{G}_b^{-1/2} \right\|_{\ell^2} \times \|(\Omega^\circ)^{1/2} (\widehat{\Omega}^\circ)^{-1/2}\|_{\ell^2} \times \|\mathcal{Z}_n^*\|_{\ell^2} \\
& = o_p(\eta'_n) \times O_p(1) \times O_p(1) \times O_p(1) \times O_p(\sqrt{K}) = o_{p^*}(r_n)
\end{aligned}$$

where the first term is by Assumption 5(b)(iii) (or zero if the  $f_t$  are linear functionals), the second term is by Lemma G.4, the third is by Lemma F.10(c) (using the fact that  $s_{JK} \asymp \tau_J$ , see Lemma A.1) and the fact that  $\|G_b^{-1/2} S [S' G_b^{-1} S]^{-1} S' G_b^{-1/2}\|_{\ell^2} = 1$ , and the fourth term is by Lemma G.3. Therefore,  $T_1 = O_{p^*}(\eta'_n \sqrt{J})$  wpa1 (since  $K \asymp J$ ), and is therefore  $= o_{p^*}(r_n)$  wpa1 by the condition stated in this Lemma.

Control of  $T_2$ : Let  $\Delta \mathbb{Z}_n(t)$  denote the Gaussian process (under  $\mathbb{P}^*$ ) defined by

$$\Delta \mathbb{Z}_n(t) = \frac{(Df_t(h_0))[\psi^J]}' \left( [\widehat{S}' \widehat{G}_b^{-1} \widehat{S}]^{-1} \widehat{S}' \widehat{G}_b^{-1/2} - [S' G_b^{-1} S]^{-1} S' G_b^{-1/2} (\Omega^\circ)^{1/2} (\widehat{\Omega}^\circ)^{-1/2} \right)}{\widehat{\sigma}_n(f_t)} \mathcal{Z}_n^*$$

for each  $t \in \mathcal{T}$ . The intrinsic semi-metric  $\Delta d_n(t_1, t_2)$  of  $\Delta \mathbb{Z}_n(t)$  is  $\Delta d_n(t_1, t_2)^2 = E^*[(\Delta \mathbb{Z}_n(t_1) - \Delta \mathbb{Z}_n(t_2))^2]$  for each  $t_1, t_2 \in \mathcal{T}$ , where  $E^*$  denotes expectation under the measure  $\mathbb{P}^*$ . Observe that:

$$\begin{aligned}
\Delta d_n(t_1, t_2) & = \left\| \left( \frac{(Df_{t_1}(h_0))[\psi^J]}'}{\sigma_n(f_{t_1})} - \frac{(Df_{t_2}(h_0))[\psi^J]}'}{\sigma_n(f_{t_2})} \right)' [S' G_b^{-1} S]^{-1} S' G_b^{-1/2} (\Omega^\circ)^{1/2} \right. \\
& \quad \times \left. (\Omega^\circ)^{-1/2} G_b^{-1/2} S \left\{ [\widehat{S}' \widehat{G}_b^{-1} \widehat{S}]^{-1} \widehat{S}' \widehat{G}_b^{-1/2} - [S' G_b^{-1} S]^{-1} S' G_b^{-1/2} (\Omega^\circ)^{1/2} (\widehat{\Omega}^\circ)^{-1/2} \right\} (\widehat{\Omega}^\circ)^{1/2} \right\|_{\ell^2} \\
& \lesssim d_n(t_1, t_2) \times \left\| G_b^{-1/2} S' \left\{ [\widehat{S}' \widehat{G}_b^{-1} \widehat{S}]^{-1} \widehat{S}' \widehat{G}_b^{-1/2} - [S' G_b^{-1} S]^{-1} S' G_b^{-1/2} (\Omega^\circ)^{1/2} (\widehat{\Omega}^\circ)^{-1/2} \right\} \right\|_{\ell^2}
\end{aligned}$$

wpa1, where the first line uses the fact that  $\widehat{\Omega}^\circ$  is invertible wpa1 and the second line uses the fact that  $\Omega^\circ$  and  $\widehat{\Omega}^\circ$  have eigenvalue uniformly bounded away from 0 and  $\infty$  wpa1. It follows by Lemma F.10(c) and Lemmas

G.3 and F.3 that

$$\begin{aligned}
& \left\| G_b^{-1/2} S' \left\{ [\widehat{S}' \widehat{G}_b^{-1} \widehat{S}]^{-1} \widehat{S}' \widehat{G}_b^{-1/2} - [S' G_b^{-1} S]^{-1} S' G_b^{-1/2} (\Omega^o)^{1/2} (\widehat{\Omega}^o)^{-1/2} \right\} \right\|_{\ell^2} \\
& \leq \left\| G_b^{-1/2} S' \left\{ [\widehat{S}' \widehat{G}_b^{-1} \widehat{S}]^{-1} \widehat{S}' \widehat{G}_b^{-1/2} - [S' G_b^{-1} S]^{-1} S' G_b^{-1/2} \right\} \right\|_{\ell^2} + \left\| I - (\Omega^o)^{1/2} (\widehat{\Omega}^o)^{-1/2} \right\|_{\ell^2} \\
& = O_p(\tau_J \zeta \sqrt{(\log J)/n}) + O_p\left( (\zeta_{b,K}^{(2+\delta)/\delta} \sqrt{(\log K)/n})^{\delta/(1+\delta)} + \delta_{h,n} \right) \\
& = O_p(\delta_{V,n}).
\end{aligned}$$

Therefore,

$$\Delta d_n(t_1, t_2) \leq O_p(\delta_{V,n}) \times d_n(t_1, t_2)$$

wpa1. Moreover, by similar arguments we have

$$\sup_{t \in \mathcal{T}} E^* [(\Delta Z_n(t))^2]^{1/2} \lesssim O_p(\delta_{V,n})$$

wpa1. Therefore, we can scale  $\Delta Z_n(t)$  by dividing through by a sequence of positive constants of order  $\delta_{V,n}$  to obtain

$$E^* [\sup_{t \in \mathcal{T}} |\Delta Z_n(t)|] \lesssim O_p(\delta_{V,n}) \times c_n$$

wpa1 by identical arguments to the proof of Lemma G.5. Therefore,

$$T_2 \leq O_p(\delta_{V,n} \times c_n) \times \sup_{t \in \mathcal{T}} \frac{\sigma_n(f_t)}{\widehat{\sigma}_n(f_t)} = O_p(\delta_{V,n} \times c_n) \times O_p(1)$$

wpa1 by Lemma G.4 and so  $T_2 = o_{p^*}(r_n)$  under Assumption 6(ii.2).

Control of  $T_3$ : The second term in  $T_3$  is the supremum of a Gaussian process with the same distribution (under  $\mathbb{P}^*$ ) as  $Z_n(t)$  (under  $\mathbb{P}$ ). Therefore, by Lemmas G.4 and G.5 we have:

$$T_3 = O_p(\delta_{V,n} + \eta'_n) \times O_{p^*}(c_n)$$

and so  $T_3 = o_{p^*}(r_n)$  wpa1 under Assumption 6(ii.2).  $\square$

**Proof of Remark 4.2.** For any  $t_1, t_2 \in \mathcal{T}$  we have:

$$\begin{aligned}
d_n(t_1, t_2) & \leq \frac{2}{\sigma_n(f_{t_1}) \vee \sigma_n(f_{t_2})} \left\| \Omega^{1/2} G_b^{-1} S [S' G_b^{-1} S]^{-1} (Df_{t_1}(h_0)[\psi^J] - Df_{t_2}(h_0)[\psi^J]) \right\|_{\ell^2} \\
& \leq \frac{2\bar{\sigma} s_{JK}^{-1}}{\underline{\sigma}_n} \left\| G_\psi^{-1/2} (Df_{t_1}(h_0)[\psi^J] - Df_{t_2}(h_0)[\psi^J]) \right\|_{\ell^2} \\
& \lesssim \frac{\tau_J \Gamma_n}{\underline{\sigma}_n} \|t_1 - t_2\|_{\ell^2}^{\gamma_n}
\end{aligned} \tag{61}$$

where the first inequality is because  $\|x/\|x\| - y/\|y\|\| \leq 2\|x - y\|/(\|x\| \vee \|y\|)$  whenever  $\|x\|, \|y\| \neq 0$  and the third is by the equivalence  $s_{JK}^{-1} \asymp \tau_J$  (see Lemma A.1). By (61) and compactness of  $\mathcal{T}$ , we have  $N(\mathcal{T}, d_n, \epsilon) \leq C(\tau_J \Gamma_n / (\epsilon \underline{\sigma}_n))^{d_{\mathcal{T}}/\gamma_n} \vee 1$  for some finite constant  $C$ .  $\square$

**Proof of Corollary 4.1.** We verify the conditions of Lemma 4.1 (or Theorem 4.1). By assumption we may take  $\underline{\sigma}_n \asymp \tau_J J^a$  with  $a = \frac{1}{2} + \frac{|\alpha|}{d}$ . Assumption 5 is therefore satisfied with  $\eta_n = \sqrt{n} \tau_J^{-1} J^{-(p/d+1/2)}$  by Remark 4.1(a') and Lemma A.3.

The continuity condition in Remark 4.2 holds with  $\Gamma_n = O(J^{a'})$  for some  $a' > 0$  and  $\gamma_n = 1$  since  $\Psi_J$  is spanned by a B-spline basis of order  $\gamma > (p \vee 2 + |\alpha|)$  (DeVore and Lorentz, 1993, Section 5.3). Assumption 6(i) therefore holds with  $c_n = O(\sqrt{\log J})$  by Remark 4.2 because  $(\tau_J \Gamma_n / (\epsilon \underline{\sigma}_n)) \lesssim (J^{a'-a} \epsilon^{-1})$ . We can therefore take  $r_n = (\log J)^{-\kappa}$  for  $\kappa \in [1/2, 1]$  in Assumption 6(ii). The first condition in Assumption 6(ii) then holds provided  $J^5 (\log J)^{6\kappa} / n = o(1)$ . Since  $\eta'_n = 0$ , the second condition in Assumption 6(ii) holds provided

$$\tau_J J \sqrt{(\log J)/n} + \sqrt{n} \tau_J^{-1} J^{-(p/d+1/2)} + \left( [J^{\frac{2+\delta}{2\delta}} \sqrt{(\log J)/n}]^{\frac{\delta}{1+\delta}} + J^{-p/d} + \tau_J \sqrt{J(\log J)/n} \right) \sqrt{\log J} = o((\log J)^{-\kappa})$$

(using Corollary 3.1 for  $\delta_{h,n}$ ). In applying Corollary 3.1 we require that the conditions  $\tau_J J / \sqrt{n} = O(1)$  and  $J^{(2+\delta)/2\delta} \sqrt{(\log n)/n} = o(1)$  hold. Finally to apply Theorem 4.1 we also need  $\tau_J J \sqrt{(\log J)/n} = o(1)$ . Sufficient conditions for all these restrictions on  $J$  are provided in the statement of this corollary. In particular, we note that  $[J^{\frac{2+\delta}{2\delta}} \sqrt{(\log n)/n}] (\log J)^{\frac{1+\delta}{\delta}}$  decreases as  $\delta > 0$  increases. Hence the condition  $J^5 (\log n)^{6\kappa} / n = o(1)$  (for  $\kappa \in [1/2, 1]$ ) implies that  $[J^{\frac{2+\delta}{2\delta}} \sqrt{(\log n)/n}]^{\frac{\delta}{1+\delta}} (\log J)^{\kappa+0.5} = o(1)$  holds for all  $\delta \geq 1$ .  $\square$

## G.5 Proofs for Section 5

**Proof of Theorem 5.1.** The result will follow from Theorem D.1. Assumption 2(i)–(iii)(iv') is satisfied under Assumption CS(iii). Assumption 3(i)(ii)(iii) is satisfied by Assumption CS (iv) and the second part of Assumption CS(v), noting that  $\zeta_{\psi,J} = O(\sqrt{J})$  and  $\zeta_{b,K} = O(\sqrt{K}) = O(\sqrt{J})$ . Since the basis spanning  $\Psi_J$  is a Riesz basis for  $T$  Assumption 4 is satisfied with  $\tau_J \asymp \mu_J^{-1}$ .

It remains to verify Assumption 5'(b). By the Riesz basis property and Assumption 2(i)–(iii) we have  $[\sigma_n(f_{CS})]^2 \asymp \sum_{j=1}^J (a_j / \mu_j)^2 \lesssim J \mu_J^{-2}$  (see Section 6 of Chen and Pouzo (2015)). For  $f_{CS}$  we have

$$Df_{CS}(h_0)[h - h_0] = \int_0^1 \left( \{h(\mathbf{p}(t), y - S_y(t)) - h_0(t, y - S_y(t))\} e^{-\int_0^t \partial_2 h_0(\mathbf{p}(v), y - S_y(v)) \mathbf{p}'(v) dv} \mathbf{p}'(t) \right) dt$$

(Hausman and Newey, 1995, p. 1471) which is clearly a linear functional (Assumption 5'(b)(i)). Note that  $\sigma_n(f_{CS}) \lesssim \sqrt{J} \mu_J^{-1}$  and Assumption CS(v) together imply  $\mu_J^{-1} J^{3/2} \sqrt{(\log J)/n} = o(1)$ . This,  $p > 2$  and Corollary 3.1 together imply that  $\|\hat{h} - h_0\|_{B_{\infty,\infty}^2} = o_p(1)$  and  $\|\tilde{h} - h_0\|_{B_{\infty,\infty}^2} = o_p(1)$ , and

$$\|\hat{h} - h_0\|_{\infty} = O_p \left( J^{-p/2} + \mu_J^{-1} \sqrt{(J \log J)/n} \right), \quad \|\hat{h} - h_0\|_{B_{\infty,\infty}^1} = O_p \left( \sqrt{J} \left( J^{-p/2} + \mu_J^{-1} \sqrt{(J \log J)/n} \right) \right) \quad (62)$$

$$\|\tilde{h} - h_0\|_{\infty} = O_p \left( J^{-p/2} \right), \quad \|\tilde{h} - h_0\|_{B_{\infty,\infty}^1} = O_p \left( \sqrt{J} \left( J^{-p/2} \right) \right). \quad (63)$$

Applying Lemma A1 of Hausman and Newey (1995), we obtain

$$\begin{aligned} \left| f_{CS}(\hat{h}) - f_{CS}(h_0) - Df_{CS}(h_0)[\hat{h} - h_0] \right| &= O_p \left( \sqrt{J} \left( J^{-p} + \mu_J^{-2} \frac{J \log J}{n} \right) \right) \\ \left| Df_{CS}(h_0)[\tilde{h} - h_0] \right| &= O_p(J^{-p/2}). \end{aligned}$$

Since  $p > 2$ , Assumption CS(v) guarantees that Assumption 5'(b)(ii) holds with

$$\eta_n = \frac{\sqrt{n}}{\sigma_n(f_{CS})} \times \left( J^{-p/2} + \mu_J^{-2} \frac{J^{3/2} \log J}{n} \right) = o(1).$$

Finally, for Assumption 5'(b)(iii), we have

$$\frac{\|\Pi_K T(\hat{v}_n(f_{CS}) - v_n(f_{CS}))\|_{L^2(W)}}{\sigma_n(f_{CS})} \lesssim \frac{\tau_J \sqrt{\sum_{j=1}^J \left( Df_{CS}(\hat{h})[(G_\psi^{-1/2} \psi^J)_j] - Df_{CS}(h_0)[(G_\psi^{-1/2} \psi^J)_j] \right)^2}}{\sigma_n(f_{CS})}$$

where  $\tau_J \asymp \mu_J^{-1}$ . Moreover,

$$\left| Df_{CS}(\hat{h})[(G_\psi^{-1/2} \psi^J)_j] - Df_{CS}(h_0)[(G_\psi^{-1/2} \psi^J)_j] \right| \lesssim \sqrt{J} \times O_p \left( \sqrt{J} \left( J^{-p/2} + \mu_J^{-1} \sqrt{(J \log J)/n} \right) \right)$$

(uniformly in  $j = 1, \dots, J$ ) by Lemma A1 of Hausman and Newey (1995). Therefore

$$\frac{\|\Pi_K T(\hat{v}_n(f_{CS}) - v_n(f_{CS}))\|_{L^2(W)}}{\sigma_n(f_{CS})} \lesssim \frac{J^{3/2} \mu_J^{-1} O_p \left( J^{-p/2} + \mu_J^{-1} \sqrt{(J \log J)/n} \right)}{\left( \sum_{j=1}^J (a_j/\mu_j)^2 \right)^{1/2}} = O_p(\eta'_n)$$

which is  $o_p(1)$  by Assumption CS(v). Finally we note that the condition  $\mu_J^{-1} J \sqrt{(\log J)/n} = o(1)$  of Theorem D.1 is trivially implied by  $\mu_J^{-1} J^{3/2} \sqrt{(\log J)/n} = o(1)$  (which is in turn implied by Assumption CS(v)). This proves the result.  $\square$

**Proof of Corollary 5.1.** For **Result (1)**, since  $\sigma_n(f_{CS}) \asymp J^{(a+\varsigma+1)/2}$ , the first part of Assumption CS(v) is satisfied provided

$$\frac{\sqrt{n}}{J^{(a+\varsigma+1)/2}} \left( J^{-p/2} + J^{\varsigma+2} \sqrt{(\log J)/n} \right) = o(1)$$

for which a sufficient condition is  $nJ^{-(p+a+\varsigma+1)} = o(1)$  and  $J^{3+\varsigma-a}(\log n)/n = o(1)$ . Moreover, the condition  $\mu_J^{-1} J^{3/2} \sqrt{(\log J)/n} = o(1)$  is implied by  $J^{3+\varsigma-(a \wedge 0)}(\log n)/n = o(1)$ . The condition  $J^{3+\varsigma-(a \wedge 0)}(\log n)/n = o(1)$  also implies that  $J^{(2+\delta)/(2\delta)} \sqrt{(\log n)/n} = o(1)$  holds whenever  $\delta \geq 2/(2 + \varsigma - (a \wedge 0))$ .

For **Result (2)**, we have  $\sigma_n(f_{CS})^2 \gtrsim a_J^2/\mu_J^2 \asymp \exp(J^{\varsigma/2} + a \log J)$ . Take  $J = (\log n/(\log n)^\varrho)^{2/\varsigma}$ . Then

$$\begin{aligned} \sigma_n(f_{CS})^2 &\gtrsim \exp \left( \log(n/(\log n)^\varrho) + \log[(\log(n/(\log n)^\varrho))^{2a/\varsigma}] \right) \\ &= \exp \left( \log[n/(\log n)^\varrho \times (\log(n/(\log n)^\varrho))^{2a/\varsigma}] \right) \\ &= n/(\log n)^\varrho \times (\log(n/(\log n)^\varrho))^{2a/\varsigma} \end{aligned}$$

and so

$$\sigma_n(f_{CS}) \gtrsim \frac{\sqrt{n}}{(\log n)^{\varrho/2}} \times (\log(n/(\log n)^\varrho))^{a/\varsigma}.$$

The first part of Assumption CS(v) is then satisfied provided

$$\frac{(\log n)^{\varrho/2}}{(\log(n/(\log n)^\varrho))^{a/\varsigma}} \left( (\log(n/(\log n)^\varrho))^{-p/\varsigma} + (\log n)^{-\varrho} \times (\log(n/(\log n)^\varrho))^{4/\varsigma} \log \log n \right) = o(1)$$

which holds provided  $2p > \varrho\varsigma - 2a$  and  $\varrho\varsigma > 8 - 2a$ . The condition  $J^{(2+\delta)/(2\delta)} \sqrt{(\log n)/n} = o(1)$  holds for any  $\delta > 0$ . The remaining condition  $\mu_J^{-1} J^{3/2} \sqrt{(\log J)/n} = o(1)$  is implied by

$$\frac{\sqrt{n}}{(\log n)^{\varrho/2}} (\log(n/(\log n)^\varrho))^{3/\varsigma} \sqrt{(\log \log n)/n} = o(1)$$

for which a sufficient condition is  $\varrho\varsigma > 6$ . Now, we may always choose  $\varrho > 0$  so that  $\varrho\varsigma > 6 \vee (8 - 2a)$ . The

remaining condition then holds provided  $2p > 6 \vee (8 - 2a) - 2a$ .  $\square$

**Proof of Theorem 5.2.** The proof follows by identical arguments to those of Theorem 5.1, noting that

$$f_{DL}(h) - f_{CS}(h) = (\mathbf{p}^1 - \mathbf{p}^0)h(\mathbf{p}^1, y)$$

and so

$$\begin{aligned} f_{DL}(\widehat{h}) - f_{DL}(h_0) &= f_{CS}(\widehat{h}) - f_{CS}(h_0) + (\mathbf{p}^1 - \mathbf{p}^0) \left( \widehat{h}(\mathbf{p}^1, y) - h_0(\mathbf{p}^1, y) \right) \\ Df_{DL}(\widehat{h})[h - h_0] &= Df_{CS}(h_0)[h - h_0] + (\mathbf{p}^1 - \mathbf{p}^0) \left( \widehat{h}(\mathbf{p}^1, y) - h_0(\mathbf{p}^1, y) \right) \\ Df_{DL}(\widehat{h})[v] - Df_{DL}(h_0)[v] &= Df_{CS}(\widehat{h})[v] - Df_{CS}(h_0)[v] \end{aligned}$$

where clearly  $|(\mathbf{p}^1 - \mathbf{p}^0)(\widehat{h}(\mathbf{p}^1, y) - h_0(\mathbf{p}^1, y))| \leq \text{const} \times \|\widehat{h} - h_0\|_\infty$ . Since  $\sigma_n(f_{DL}) \asymp \mu_J^{-1} \sqrt{J}$ , the stated conditions on  $J$  in this theorem imply that Assumption CS(v) holds.  $\square$

**Proof of Theorem 5.3.** The result will follow from Theorem D.1, and is very similar to that of Theorem 5.1. Assumptions 2(i)–(iii)(iv'), 3(i)(ii)(iii), and 4 are verified as in the proof of Theorem 5.1. It remains to verify Assumption 5'(b). As in the proof of Theorem 5.1 we have  $\tau_J \asymp \mu_J^{-1}$  and  $[\sigma_n(f_A)]^2 \asymp \sum_{j=1}^J (a_j/\mu_j)^2$  (see Section 6 of Chen and Pouzo (2015)). Simple expansion of  $f_A$  yields

$$Df_A(h_0)[h - h_0] = \int w(\mathbf{p}) e^{h_0(\log \mathbf{p}, \log y)} (h(\log \mathbf{p}, \log y) - h_0(\log \mathbf{p}, \log y)) d\mathbf{p}$$

which is clearly a linear functional (Assumption 5'(b)(i)), and

$$\begin{aligned} & \left| f_A(\widehat{h}) - f_A(h_0) - Df_A(h_0)[\widehat{h} - h_0] \right| \\ &= \int w(\mathbf{p}) \left( e^{\widehat{h}(\log \mathbf{p}, \log y) - h_0(\log \mathbf{p}, \log y)} - 1 - (\widehat{h}(\log \mathbf{p}, \log y) - h_0(\log \mathbf{p}, \log y)) \right) e^{h_0(\log \mathbf{p}, \log y)} d\mathbf{p}. \end{aligned}$$

Therefore, by Corollary 3.1 we have

$$\begin{aligned} \left| f_A(\widehat{h}) - f_A(h_0) - Df_A(h_0)[\widehat{h} - h_0] \right| &= O_p \left( J^{-p} + \mu_J^{-2} \frac{J \log J}{n} \right) \\ \left| Df_A(h_0)[\widehat{h} - h_0] \right| &= O_p(J^{-p/2}). \end{aligned}$$

Since  $p > 0$ , the stated conditions on  $J$  in this theorem guarantees that Assumption 5'(b)(ii) holds with

$$\eta_m = \frac{\sqrt{n}}{\sigma_n(f_A)} \times \left( J^{-p/2} + \mu_J^{-2} \frac{J \log J}{n} \right) = o(1).$$

Finally, for Assumption 5'(b)(iii), we have

$$\frac{\|\Pi_K T(\widehat{v}_n(f_A) - v_n(f_A))\|_{L^2(W)}}{\sigma_n(f_A)} \lesssim \frac{\tau_J}{\sigma_n(f_A)} \sqrt{\left( \sum_{j=1}^J \left( Df_A(\widehat{h})[(G_\psi^{-1/2} \psi^J)_j] - Df_A(h_0)[(G_\psi^{-1/2} \psi^J)_j] \right)^2 \right)}$$

where  $\tau_J \asymp \mu_J^{-1}$  and where a first-order Taylor expansion of  $Df_A$  yields

$$\begin{aligned} \left| Df_A(\hat{h})[(G_\psi^{-1/2}\psi^J)_j] - Df_A(h_0)[(G_\psi^{-1/2}\psi^J)_j] \right| &\lesssim \|(G_\psi^{-1/2}\psi^J)_j\|_\infty \times \|\hat{h} - h_0\|_\infty \\ &= O_p\left(\sqrt{J}\left(J^{-p/2} + \mu_J^{-1}\sqrt{(J\log J)/n}\right)\right). \end{aligned}$$

It follows that

$$\frac{\|\Pi_K T(\hat{v}_n(f_A) - v_n(f_A))\|_{L^2(W)}}{[\sigma_n(f_A)]} \lesssim \frac{\mu_J^{-1}J \times O_p\left(J^{-p/2} + \mu_J^{-1}\sqrt{(J\log J)/n}\right)}{\left(\sum_{j=1}^J (a_j/\mu_j)^2\right)^{1/2}} = O_p(\eta'_n)$$

which is  $o_p(1)$  by the displayed condition on  $J$  in this theorem. Finally we note that the condition  $\mu_J^{-1}J\sqrt{(\log J)/n} = o(1)$  of Theorem D.1 is implied by the displayed condition on  $J$  in this theorem and the fact that  $\sigma_n(f_A) \lesssim \sqrt{J}\mu_J^{-1}$ . This proves the result.  $\square$

**Proof of Theorem 5.4.** We verify the conditions of Lemma 4.1 and Theorem 4.1. Assumptions 1 and 2 are satisfied by Assumption CS(i)(ii)(iv) and U-CS(i). Assumption 3(iii) is satisfied by Assumption U-CS(iii). Assumption 4(i) is satisfied by the Riesz basis condition. For Assumption 5(b), we check the conditions of Remark 4.1(b'). It is clear that  $Df_{CS,t}[h - h_0]$  (see display (14)) is a linear functional of  $h - h_0$  for each  $t \in \mathcal{T}$ . As in the proof of Theorem 5.1 we have  $[\sigma_n(f_{CS,t})]^2 \asymp \sum_{j=1}^J (a_{j,t}/\mu_j)^2$  uniformly in  $t$ . Thus,  $\underline{\sigma}_n \lesssim \sqrt{J}\mu_J^{-1}$ . This and Assumption U-CS(iv.1) together imply  $\mu_J^{-1}J^{3/2}\sqrt{(\log J)/n} = o(1)$ . Also we note that the first part of Assumption U-CS(iii) and  $\delta \geq 1$  imply that  $J^{(2+\delta)/(2\delta)}\sqrt{(\log n)/n} = o(1)$  holds. These results and  $p > 2$  and Corollary 3.1 together imply that  $\|\hat{h} - h_0\|_{B_{\infty,\infty}^2} = o_p(1)$  and  $\|\tilde{h} - h_0\|_{B_{\infty,\infty}^2} = o_p(1)$ , and equations (62) and (63) hold. Therefore,  $\hat{h}$  and  $\tilde{h}$  are within an  $\epsilon$  neighborhood (in Hölder norm of smoothness 2) of  $h_0$  wpa1. As  $\mathcal{T} = [\underline{p}^0, \bar{p}^0] \times [\underline{p}^1, \bar{p}^1] \times [\underline{y}, \bar{y}]$  where the intervals  $[\underline{p}^0, \bar{p}^0]$  and  $[\underline{p}^1, \bar{p}^1]$  are in the interior of the support of  $P_i$  and  $[\underline{y}, \bar{y}]$  is in the interior of the support of  $Y_i$  and  $h_0 \in B_\infty(p, L)$  with  $p > 2$  and  $0 < L < \infty$ , it is straightforward to extend Lemma A1 of Hausman and Newey (1995) to show

$$\begin{aligned} \sup_{t \in \mathcal{T}} \left| f_{CS,t}(\hat{h}) - f_{CS,t}(h_0) - Df_{CS,t}(h_0)[\hat{h} - h_0] \right| &= O_p\left(\sqrt{J}\left(J^{-p} + \mu_J^{-2}\frac{J\log J}{n}\right)\right), \\ \sup_{t \in \mathcal{T}} \left| Df_{CS,t}(h_0)[\tilde{h} - h_0] \right| &= O_p(J^{-p/2}) \end{aligned}$$

by Corollary 3.1. Since  $\underline{\sigma}_n \lesssim \sqrt{J}\mu_J^{-1}$ , Assumption U-CS(iii) guarantees that Assumption 5(b)(ii) holds with

$$\eta_n = \frac{\sqrt{n}}{\underline{\sigma}_n} \times \left( J^{-p/2} + \mu_J^{-2}\frac{J^{3/2}\log J}{n} \right).$$

For Assumption 5(b)(iii), we have

$$\sup_{t \in \mathcal{T}} \frac{\|\Pi_K T(\hat{v}_n(f_{CS,t}) - v_n(f_{CS,t}))\|_{L^2(W)}}{[\sigma_n(f_{CS,t})]} \lesssim \sup_{t \in \mathcal{T}} \tau_J \frac{\sqrt{\sum_{j=1}^J \left( Df_{CS,t}(\hat{h})[(G_\psi^{-1/2}\psi^J)_j] - Df_{CS,t}(h_0)[(G_\psi^{-1/2}\psi^J)_j] \right)^2}}{[\sigma_n(f_{CS,t})]}$$

where  $\tau_J \asymp \mu_J^{-1}$ . By straightforward extension of Lemma A1 of Hausman and Newey (1995) and (62) and (63):

$$\sup_{t \in \mathcal{T}} \left| Df_{CS,t}(\hat{h})[(G_\psi^{-1/2}\psi^J)_j] - Df_{CS,t}(h_0)[(G_\psi^{-1/2}\psi^J)_j] \right| \lesssim \sqrt{J} \times O_p\left(\sqrt{J}\left(J^{-p/2} + \mu_J^{-1}\sqrt{(J\log J)/n}\right)\right)$$

whence Assumption 5(b)(iii) holds with

$$\eta'_n = \frac{J^{3/2}\mu_J^{-1}}{\underline{\sigma}_n} \times \left( J^{-p/2} + \mu_J^{-1} \sqrt{J(\log J)/n} \right).$$

which is  $o(1)$  by Assumption U-CS(iv). This verifies Assumption 5(b).

Finally, Assumption 6(i) holds with  $c_n = O(\sqrt{\log J})$  by Assumption U-CS(ii) and Remark 4.2. For Assumption 6(ii) we take  $r_n = \lceil \log J \rceil^{-1/2}$ . Assumption 6(ii.1) then holds provided  $J^5(\log J)^3/n = o(1)$ . Assumption 6(ii.2) holds provided

$$\tau_J J \sqrt{(\log J)/n} + \eta_n + \left( [J^{\frac{2+\delta}{2\delta}} \sqrt{(\log J)/n}]^{\frac{\delta}{1+\delta}} + J^{-p/d} + \tau_J \sqrt{J(\log J)/n} + \eta'_n \right) \sqrt{\log J} = o((\log J)^{-1/2})$$

(using Corollary 3.1 for  $\delta_{h,n}$ ), which is satisfied provided

$$\tau_J J \sqrt{(\log J)/n} + \eta_n + \eta'_n \sqrt{\log J} = o((\log J)^{-1/2})$$

which is in turn implied by Assumption U-CS(iii) and U-CS(iv.1) and the property  $\underline{\sigma}_n \lesssim \sqrt{J}\mu_J^{-1}$ . Thus Lemma 4.1 applies to  $f_t = f_{CS,t}$  with a rate  $r_n = \lceil \log J \rceil^{-1/2}$ .

Next we note that the condition  $\eta'_n \sqrt{J} = o((\log J)^{-1/2})$  needed for Theorem 4.1 is directly implied by Assumption U-CS(iv.2).  $\square$

**Proof of Theorem 5.5.** Follows by similar arguments to the proofs of Theorems 5.2 and 5.4, noting that

$$|Df_{DL,t_1}(h_0)[h] - Df_{DL,t_2}(h_0)[h]| \leq |Df_{CS,t_1}(h_0)[h] - Df_{CS,t_2}(h_0)[h]| + |(\mathbf{p}_1^1 - \mathbf{p}_1^0)h(\mathbf{p}_1^1, y_1) - (\mathbf{p}_2^1 - \mathbf{p}_2^0)h(\mathbf{p}_2^1, y_2)|$$

and so  $c_n = O(\sqrt{\log J})$  by Assumption U-CS(ii) and Remark 4.2 (see the proof of Corollary 4.1). We can then take  $r_n = \lceil \log J \rceil^{-1/2}$ .  $\square$

## G.6 Proofs for Appendix B

**Proof of Theorem B.1.** As with the proof of Theorem 3.1, we first decompose the error into three parts:

$$\begin{aligned} \|\widehat{h} - h_0\|_{L^2(X)} &\leq \|\widehat{h} - \widetilde{h}\|_{L^2(X)} + \|\widetilde{h} - \Pi_J h_0\|_{L^2(X)} + \|\Pi_J h_0 - h_0\|_{L^2(X)} \\ &=: T_1 + T_2 + \|h_0 - \Pi_J h_0\|_{L^2(X)}. \end{aligned}$$

To prove **Result (1)** it is enough to show that  $T_2 \leq O_p(1) \times \|h_0 - \Pi_J h_0\|_{L^2(X)}$ . To do this, bound

$$\begin{aligned} T_2 &\leq \|G_\psi^{1/2} (S' G_b^{-1/2})_l^- G_b^{-1/2} B'(H_0 - \Psi_{C_J})/n\|_{\ell^2} \\ &\quad + \|G_\psi^{1/2} \{(\widehat{G}_b^{-1/2} \widehat{S})_l^- \widehat{G}_b^{-1/2} G_b^{1/2} - (G_b^{-1/2} S)_l^-\} G_b^{-1/2} B'(H_0 - \Psi_{C_J})/n\|_{\ell^2} =: T_{21} + T_{22}. \end{aligned}$$

For  $T_{21}$ ,

$$\begin{aligned}
T_{21} &\leq s_{JK}^{-1} \|G_b^{-1/2} B'(H_0 - \Psi_{c_J})/n\|_{\ell^2} \\
&\leq O_p(\tau_J \zeta_{b,K}/\sqrt{n}) \times \|h_0 - \Pi_J h_0\|_{L^2(X)} + \tau_J \|\Pi_K T(h_0 - \Pi_J h_0)\|_{L^2(W)} \\
&\leq O_p(\tau_J \zeta_{b,K}/\sqrt{n}) \times \|h_0 - \Pi_J h_0\|_{L^2(X)} + \tau_J \|T(h_0 - \Pi_J h_0)\|_{L^2(W)} \\
&= O_p(1) \times \|h_0 - \Pi_J h_0\|_{L^2(X)}
\end{aligned}$$

where the second line is by Lemma F.9 and the relations  $J \asymp K$  and  $\tau_J \asymp s_{JK}^{-1}$ , and the final line is by Assumption 4(ii) and the condition  $\tau_J \zeta \sqrt{(\log J)/n} = o(1)$ . Similarly,

$$\begin{aligned}
T_{22} &\leq \|G_\psi^{1/2} \{(\widehat{G}_b^{-1/2} \widehat{S})_l^- \widehat{G}_b^{-1/2} G_b^{1/2} - (G_b^{-1/2} S)_l^-\}\|_{\ell^2} \|G_b^{-1/2} B'(H_0 - \Psi_{c_J})/n\|_{\ell^2} \\
&\leq O_p(s_{JK}^{-2} \zeta \sqrt{(\log J)/n}) \times (O_p(\zeta_{b,K}/\sqrt{n}) \times \|h_0 - \Pi_J h_0\|_{L^2(X)} + \|T(h_0 - \Pi_J h_0)\|_{L^2(W)}) \\
&= o_p(\tau_J \zeta \sqrt{(\log J)/n})^2 \times \|h_0 - \Pi_J h_0\|_{L^2(X)} + O_p(1) \times \tau_J \|T(h_0 - \Pi_J h_0)\|_{L^2(W)} \\
&= O_p(1) \times \|h_0 - \Pi_J h_0\|_{L^2(X)}
\end{aligned}$$

where the second line is by Lemmas F.9 and F.10(b) and the relations  $J \asymp K$  and  $\tau_J \asymp s_{JK}^{-1}$ , and the final line is by the condition  $\tau_J \zeta \sqrt{(\log J)/n} = o(1)$  and Assumption 4(ii). This proves Result (1).

To prove **Result (2)** it remains to control  $T_1$ . To do this, bound

$$\begin{aligned}
T_1 &\leq \|G_\psi^{1/2} (S' G_b^{-1/2})_l^- G_b^{-1/2} B' u/n\|_{\ell^2} + \|G_\psi^{1/2} \{(\widehat{G}_b^{-1/2} \widehat{S})_l^- \widehat{G}_b^{-1/2} G_b^{1/2} - (G_b^{-1/2} S)_l^-\} G_b^{-1/2} B' u/n\|_{\ell^2} \\
&=: T_{11} + T_{12}.
\end{aligned}$$

For  $T_{11}$ , by definition of  $s_{JK}$  and Lemma F.8 we have:

$$T_{11} \leq s_{JK}^{-1} \|G_b^{-1/2} B' u/n\|_{\ell^2} = O_p(s_{JK}^{-1} \sqrt{K/n}) = O_p(\tau_J \sqrt{J/n})$$

where the final line is because  $J \asymp K$  and  $\tau_J \asymp s_{JK}^{-1}$  (Lemma A.1). Similarly,

$$\begin{aligned}
T_{12} &\leq \|G_\psi^{1/2} \{(\widehat{G}_b^{-1/2} \widehat{S})_l^- \widehat{G}_b^{-1/2} G_b^{1/2} - (G_b^{-1/2} S)_l^-\}\|_{\ell^2} \|G_b^{-1/2} B' u/n\|_{\ell^2} \\
&= O_p(\tau_J^2 \zeta \sqrt{(\log J)/n}) \times O_p(\sqrt{J/n}) \\
&= o_p(\tau_J \sqrt{J/n})
\end{aligned}$$

where the second line is by Lemmas F.8 and F.10(b) and the relations  $J \asymp K$  and  $\tau_J \asymp s_{JK}^{-1}$ , and the final line is by the condition  $\tau_J \zeta \sqrt{(\log J)/n} = o(1)$ .  $\square$

**Proof of Corollary B.1.** Analogous to the proof of Corollary 3.1.  $\square$

**Proof of Theorem B.2.** As in the proof of Theorem 3.2, it suffices to prove a lower bound for the Gaussian reduced-form NPIR model (45). Theorem G.3 below does just this.  $\square$

**Theorem G.3.** *Let Condition LB hold with  $B_2(p, L)$  in place of  $B_\infty(p, L)$  hold for the NPIR model (45) with a random sample  $\{(W_i, Y_i)\}_{i=1}^n$ . Then for any  $0 \leq |\alpha| < p$ :*

$$\liminf_{n \rightarrow \infty} \inf_{\widehat{g}_n} \sup_{h \in B_2(p, L)} \mathbb{P}_h \left( \|\widehat{g}_n - \partial^\alpha h\|_{L^2(X)} \geq cn^{-(p-|\alpha|)/(2(p+\varsigma)+d)} \right) \geq c' > 0$$

in the mildly ill-posed case, and

$$\liminf_{n \rightarrow \infty} \inf_{\hat{g}_n} \sup_{h \in B_2(p, L)} \mathbb{P}_h \left( \|\hat{g}_n - \partial^\alpha h\|_{L^2(X)} \geq c(\log n)^{-(p-|\alpha|)/\varsigma} \right) \geq c' > 0$$

in the severely ill-posed case, in the severely ill-posed case, where  $\inf_{\hat{g}_n}$  denotes the infimum over all estimators of  $\partial^\alpha h$  based on the sample of size  $n$ ,  $\sup_{h \in B_2(p, L)} \mathbb{P}_h$  denotes the sup over  $h \in B_2(p, L)$  and distributions  $(W_i, u_i)$  which satisfy Condition LB with  $\nu$  fixed, and the finite positive constants  $c, c'$  depend only on  $p, L, d, \varsigma$  and  $\sigma_0$ .

**Proof of Theorem G.3.** We use similar arguments to the proof of Theorem G.1, using Theorem 2.5 of [Tsybakov \(2009\)](#) (see Theorem G.2). Again, we first explain the scalar ( $d = 1$ ) case in detail. Let  $\{\phi_{j,k}, \psi_{j,k}\}_{j,k}$  be a wavelet basis of regularity  $\gamma > p$  for  $L^2([0, 1])$  as described in Appendix E.

By construction, the support of each interior wavelet is an interval of length  $2^{-j}(2r-1)$ . Thus for all  $j$  sufficiently large (hence the lim inf in our statement of the Lemma) we may choose a set  $M \subset \{r, \dots, 2^j - r - 1\}$  of interior wavelets with cardinality  $\#(M) \asymp 2^j$  such that  $\text{support}(\psi_{j,m}) \cap \text{support}(\psi_{j,m'}) = \emptyset$  for all  $m, m' \in M$  with  $m \neq m'$ .

Take  $g_0 \in B(p, L/2)$  and for each  $m \in M$  define  $\theta = \{\theta_m\}_{m \in M}$  where each  $\theta_m \in \{0, 1\}$  and define

$$h_\theta = g_0 + c_0 2^{-j(p+1/2)} \sum_{m \in M} \theta_m \psi_{j,m}$$

for each  $\theta$ , where  $c_0$  is a positive constant to be defined subsequently. Note that this gives  $2^{\#(M)}$  such choices of  $h_\theta$ . By the equivalence  $\|\cdot\|_{B_{2,2}^p} \asymp \|\cdot\|_{b_{2,2}^p}$ , for each  $\theta$  we have:

$$\begin{aligned} \|h_\theta\|_{B_{2,2}^p} &\leq L/2 + \left\| c_0 2^{-j(p+1/2)} \sum_{m \in M} \theta_m \psi_{j,m} \right\|_{B_{2,2}^p} \\ &\leq L/2 + \text{const} \times \left\| c_0 2^{-j(p+1/2)} \sum_{m \in M} \theta_m \psi_{j,m} \right\|_{b_{2,2}^p} \\ &= L/2 + \text{const} \times c_0 2^{-j(p+1/2)} \left( \sum_{m \in M} \theta_m^2 2^{2jp} \right)^{1/2} \\ &\leq L/2 + \text{const} \times c_0. \end{aligned}$$

Therefore, we can choose  $c_0$  sufficiently small that  $h_\theta \in B_2(p, L)$  for each  $\theta$ .

Since  $\psi_{j,m} \in C^\gamma$  with  $\gamma > |\alpha|$  is compactly supported and  $X_i$  has density bounded away from 0 and  $\infty$ , we have  $\|2^{j/2} \psi^{(|\alpha|)}(2^j x - m)\|_{L^2(X)} \asymp 1$  (uniformly in  $m$ ). By this and the disjoint support of the  $\psi_{j,m}$ , for each  $\theta, \theta'$  we have:

$$\begin{aligned} \|\partial^\alpha h_\theta - \partial^\alpha h_{\theta'}\|_{L^2(X)} &= c_0 2^{-j(p-|\alpha|+1/2)} \left( \sum_{m \in M} (\theta_m - \theta'_m)^2 \|2^{j/2} \psi^{(|\alpha|)}(2^j \cdot - m)\|_{L^2(X)}^2 \right)^{1/2} \\ &\gtrsim c_0 2^{-j(p-|\alpha|+1/2)} \sqrt{\rho(\theta, \theta')} \end{aligned}$$

where  $\rho(\theta, \theta')$  is the Hamming distance between  $\theta$  and  $\theta'$ . Take  $j$  large enough that  $\#(M) \geq 8$ . By the Varshamov-Gilbert bound ([Tsybakov, 2009](#), Lemma 2.9) we may choose a subset  $\theta^{(0)}, \theta^{(1)}, \dots, \theta^{(M^*)}$  such that  $\theta^0 = (0, \dots, 0)$ ,  $\rho(\theta^{(a)}, \theta^{(b)}) \geq \#(M)/8 \gtrsim 2^j$  for all  $0 \leq a < b \leq M^*$  and  $M^* \geq 2^{\#(M)/8}$  (recall that there were

$2^{\#(M)}$  distinct vectors  $\theta$  and  $\#(M) \asymp 2^j$ . For each  $m \in \{0, 1, \dots, m^*\}$  let  $h_m := h_{\theta^{(m)}}$ . Then we have

$$\|\partial^\alpha h_m - \partial^\alpha h_{m'}\|_{L^2(X)} \gtrsim c_0 2^{-j(p-|\alpha|)}$$

for each  $0 \leq m < m' \leq M^*$ .

For each  $0 \leq m \leq M^*$ , let  $P_m$  denote the joint distribution of  $\{(W_i, Y_i)\}_{i=1}^n$  with  $Y_i = Th_m(W_i) + u_i$  for the Gaussian NPIR model (45). It follows from Condition LB(ii)(iii) that for each  $1 \leq m \leq M^*$  the KL distance  $K(P_m, P_0)$  is

$$\begin{aligned} K(P_m, P_0) &\leq \frac{1}{2} \sum_{i=1}^n (c_0 2^{-j(p+\frac{1}{2})})^2 E \left[ \frac{(T \sum_{k \in M} \theta_k^{(m)} (\psi_{j,k}(W_i)))^2}{\sigma^2(W_i)} \right] \\ &\leq \frac{n}{2\sigma^2} (c_0 2^{-j(p+\frac{1}{2})})^2 \nu (2^j)^2 \sum_{k \in M} (\theta_k^{(m)})^2 \|\psi_{j,k}\|_{L^2(X)}^2 \\ &\lesssim n c_0^2 2^{-2jp} \nu (2^j)^2 \end{aligned}$$

where the final line is because  $X_i$  has density bounded away from 0 and  $\infty$  and  $\sum_{k \in M} (\theta_k^{(m)})^2 \leq \#(M) \asymp 2^j$  for each  $1 \leq m \leq M^*$ .

In the mildly ill-posed case ( $\nu(2^j) = 2^{-j\varsigma}$ ) we choose  $2^j \asymp n^{1/(2(p+\varsigma)+1)}$ . This yields:

$$\begin{aligned} K(P_m, P_0) &\lesssim c_0^2 n^{\frac{1}{2(p+\varsigma)+1}} \quad \text{uniformly in } m \\ \log(M^*) &\gtrsim 2^j \asymp n^{1/(2(p+\varsigma)+1)}. \end{aligned}$$

since  $M^* \geq 2^{\#(M)/8}$  and  $\#(M) \asymp 2^j$ .

In the severely ill-posed case ( $\nu(2^j) = \exp(-\frac{1}{2}2^{j\varsigma})$ ) we choose  $2^j = (c_1 \log n)^{1/\varsigma}$  with  $c_1 > 1$ . This yields:

$$\begin{aligned} K(P_m, P_0) &\lesssim c_0^2 n^{-(c_1-1)} \quad \text{uniformly in } m \\ \log(M^*) &\gtrsim 2^j \asymp (\log n)^{1/\varsigma}. \end{aligned}$$

In both the mildly and severely ill-posed cases, the result follows by choosing  $c_0$  sufficiently small that both  $\|h_m\|_{B_{2,2}^p} \leq L$  and  $K(P_m, P_0) < \frac{1}{8} \log(M^*)$  hold uniformly in  $m$  for all  $n$  sufficiently large. All conditions of Theorem 2.5 of [Tsybakov \(2009\)](#) are satisfied and hence we obtain the lower bound result.

In the multivariate case ( $d > 1$ ) we let  $\tilde{\psi}_{j,k,G}(x)$  denote an orthonormal tensor-product wavelet for  $L^2([0, 1]^d)$  at resolution level  $j$ . We construct a family of submodels analogously to the univariate case, setting  $h_\theta = g_0 + c_0 2^{-j(p+d/2)} \sum_{m \in M} \theta_m \tilde{\psi}_{j,m,G}$  where  $\psi_{j,m}$  is now the product of  $d$  interior univariate wavelets at resolution level  $j$  with  $G = (w_\psi)^d$  (see Appendix E) and where  $\#(M) \asymp 2^{jd}$ . We then use the Varshamov-Gilbert bound to reduce this to a family of models  $h_m$  with  $0 \leq m \leq M^*$  and  $M^* \asymp 2^{jd}$ . We then have:

$$\|\partial^\alpha h_m - \partial^\alpha h_{m'}\|_\infty \gtrsim c_0 2^{-j(p-|\alpha|)}$$

for each  $0 \leq m < m' \leq M^*$ , and

$$K(P_m, P_0) \lesssim n (c_0 2^{-j(p+d/2)})^2 \nu (2^j)^2$$

for each  $1 \leq m \leq M^*$ , where  $\nu(2^j) = 2^{-j\varsigma}$  in the mildly ill-posed case and  $\nu(2^j) \asymp \exp(-2^{j\varsigma})$  in the severely ill-posed case. We choose  $2^j \asymp n^{1/(2(p+\varsigma)+d)}$  in the mildly ill-posed case and  $2^j = (c_1 \log n)^{1/\varsigma}$  in the severely

ill-posed case. The result follows as in the univariate case.  $\square$

## G.7 Proofs for Appendix C

**Proof of Theorem C.1.** As in the proof of Theorem 3.2, this follows from the lower bound for NPIR in Theorem G.4.  $\square$

The following is a slightly stronger “in probability” version of Lemma 1 in Yu (1997), which is used to prove Theorem G.4. Let  $\mathcal{P}$  be a family of probability measures, let  $\theta(P)$  be a parameter with values in a pseudo-metric space  $(\mathcal{D}, d)$  for some distribution  $P \in \mathcal{P}$ , and let  $\hat{\theta}(P)$  be an estimator of  $\theta(P)$  taking values in  $(\mathcal{D}, d)$ . If  $\theta \in \mathcal{D}$  and  $D \subset \mathcal{D}$ , we let  $d(\theta, D) = \inf_{\theta' \in D} d(\theta, \theta')$ . Let  $co(\mathcal{P})$  denote the convex hull of a set of measures  $\mathcal{P}$ . Finally, if  $\mathbb{P}, \mathbb{Q} \in \mathcal{P}$  we let  $\|\mathbb{P} - \mathbb{Q}\|_{TV}$  denote the total variation distance and  $\text{aff}(\mathbb{P}, \mathbb{Q}) = 1 - \|\mathbb{P} - \mathbb{Q}\|_{TV}$  denote the affinity between  $\mathbb{P}$  and  $\mathbb{Q}$ .

**Lemma G.7.** *Suppose there are subsets  $D_1, D_2 \subset \mathcal{D}$  that are  $2\delta$  separated for some  $\delta > 0$  (i.e.  $d(s_1, s_2) \geq 2\delta$  for all  $s_1 \in D_1$  and  $s_2 \in D_2$ ) and subsets  $\mathcal{P}_1, \mathcal{P}_2 \subset \mathcal{P}$  for which  $\theta(\mathbb{P}) \in D_1$  for all  $\mathbb{P} \in \mathcal{P}_1$  and  $\theta(\mathbb{P}) \in D_2$  for all  $\mathbb{P} \in \mathcal{P}_2$ . Then:*

$$2 \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}(d(\hat{\theta}, \theta(\mathbb{P})) \geq \delta) \geq \sup_{\mathbb{P}_1 \in co(\mathcal{P}_1), \mathbb{P}_2 \in co(\mathcal{P}_2)} \text{aff}(\mathbb{P}_1, \mathbb{P}_2).$$

**Proof of Lemma G.7.** We proceed as in the proof of Lemma 1 in Yu (1997). Let  $\mathcal{P}_1 \in \mathcal{P}_1$  and  $\mathcal{P}_2 \in \mathcal{P}_2$ . Then:

$$\begin{aligned} 2 \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}(d(\hat{\theta}, \theta(\mathbb{P})) \geq \delta) &\geq \mathbb{P}_1(d(\hat{\theta}, \theta(\mathcal{P}_1)) \geq \delta) + \mathbb{P}_2(d(\hat{\theta}, \theta(\mathcal{P}_2)) \geq \delta) \\ &\geq \mathbb{P}_1(d(\hat{\theta}, D_1) \geq \delta) + \mathbb{P}_2(d(\hat{\theta}, D_2) \geq \delta). \end{aligned}$$

Since the inequality  $2 \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}(d(\hat{\theta}, \theta(\mathbb{P})) \geq \delta) \geq \mathbb{P}_1(d(\hat{\theta}, D_1) \geq \delta) + \mathbb{P}_2(d(\hat{\theta}, D_2) \geq \delta)$  holds for any fixed  $\mathbb{P}_1 \in \mathcal{P}_1$  and  $\mathbb{P}_2 \in \mathcal{P}_2$ , it must also hold for any  $\mathbb{P}_1 \in co(\mathcal{P}_1)$  and  $\mathbb{P}_2 \in co(\mathcal{P}_2)$ . Also note that

$$\begin{aligned} \mathbb{1}\{d(\hat{\theta}, D_1) \geq \delta\} + \mathbb{1}\{d(\hat{\theta}, D_2) \geq \delta\} &\geq \mathbb{1}\{d(\hat{\theta}, D_1) + d(\hat{\theta}, D_2) \geq 2\delta\} \\ &\geq \mathbb{1}\{d(D_1, D_2) \geq 2\delta\} = 1 \end{aligned}$$

because  $d(D_1, D_2) \geq 2\delta$ . Now by definition of  $\alpha(\cdot, \cdot)$ , for any  $\mathbb{P}_1 \in co(\mathcal{P}_1)$  and  $\mathbb{P}_2 \in co(\mathcal{P}_2)$  we have:

$$\begin{aligned} 2 \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P}(d(\hat{\theta}, \theta(\mathbb{P})) \geq \delta) &\geq \mathbb{P}_1(d(\hat{\theta}, D_1) \geq \delta) + \mathbb{P}_2(d(\hat{\theta}, D_2) \geq \delta) \\ &\geq \inf\{\mathbb{P}_1 f + \mathbb{P}_2 g : f, g \text{ non negative and measurable with } f + g \geq 1\} \\ &= \text{aff}(\mathbb{P}_1, \mathbb{P}_2). \end{aligned}$$

The result follows by taking the supremum of the right-hand side over  $\mathbb{P}_1$  and  $\mathbb{P}_2$ .  $\square$

**Theorem G.4.** *Let Condition LB hold with  $B_2(p, L)$  in place of  $B_\infty(p, L)$  for the NPIR model (45) with a random sample  $\{(W_i, Y_i)\}_{i=1}^n$ . Then for any  $0 \leq |\alpha| < p$ :*

$$\liminf_{n \rightarrow \infty} \inf_{\hat{g}_n} \sup_{h \in B_2(p, L)} \mathbb{P}_h(|\hat{g}_n - f(h)| > cr_n) \geq c' > 0$$

where

$$r_n = \begin{cases} n^{-1/2} & \text{in the mildly ill-posed case when } p \geq \varsigma + 2|\alpha| + d/4 \\ n^{-4(p-|\alpha|)/(4(p+\varsigma)+d)} & \text{in the mildly ill-posed case when } \varsigma < p < \varsigma + 2|\alpha| + d/4 \\ (\log n)^{-2(p-|\alpha|)/\varsigma} & \text{in the severely ill-posed case,} \end{cases}$$

$\inf_{\hat{g}_n}$  denotes the infimum over all estimators of  $f(h_0)$  based on the sample of size  $n$ ,  $\sup_{h \in B_2(p,L)} \mathbb{P}_h$  denotes the sup over  $h \in B_2(p,L)$  and distributions  $(W_i, u_i)$  which satisfy Condition LB with  $\nu$  fixed, and the finite positive constants  $c, c'$  do not depend on  $n$ .

**Proof of Theorem G.4.** We first prove the result for the scalar ( $d = 1$ ) case, then describe the modifications required in the multivariate case.

Let  $\{\phi_{j,k}, \psi_{j,k}\}_{j,k}$  be a CDV wavelet basis of regularity  $\gamma > p$  for  $L^2([0,1])$ , as described in Appendix E. As in the proof of Theorem G.3, we choose a set  $M \subset \{r, \dots, 2^j - r - 1\}$  of interior wavelets with cardinality  $\mathfrak{m} := \#(M) \asymp 2^j$  such that  $\text{support}(\psi_{j,m}) \cap \text{support}(\psi_{j,m'}) = \emptyset$  for all  $m, m' \in M$  with  $m \neq m'$ . Let  $\theta = \{\theta_m\}_{m \in M}$  where each  $\theta_m \in \{-1, 1\}$  and for each  $\theta \in \{-1, 1\}^{\mathfrak{m}}$  define:

$$h_\theta = \sum_{m \in M} \frac{\theta_m c_0 2^{-jp}}{\sqrt{\mathfrak{m}}} \psi_{j,m}.$$

and let  $h_0 = 0$ . By the equivalence  $\|\cdot\|_{b_{2,2}^p} \asymp \|\cdot\|_{B_{2,2}^p}$ , we have:

$$\|h_\theta\|_{B_{2,2}^p} \lesssim \|h_\theta\|_{b_{2,2}^p} = \left( 2^{2jp} \sum_{m \in M} \frac{\theta_m^2 c_0^2 2^{-2jp}}{\mathfrak{m}} \right)^{1/2} = c_0.$$

Therefore, we may choose  $c_0$  sufficiently small that  $h_\theta \in B_2(p,L)$  for all  $\theta \in \{-1, 1\}^{\mathfrak{m}}$ .

Let  $\psi^{(|\alpha|)}$  denote the  $|\alpha|$ th derivative of  $\psi$ . By disjoint support of the  $\psi_{j,m}(x) = 2^{j/2} \psi(2^j x - m)$ ,  $\mu(x) \geq \underline{\mu} > 0$ , and a change of variables, we have:

$$\begin{aligned} |f(h_\theta) - f(h_0)| &= \mathfrak{m}^{-1} \sum_{m \in M} \int (c_0 \theta_m 2^{-jp} \psi_{j,m}^{(|\alpha|)}(x))^2 \mu(x) dx \\ &\gtrsim \mathfrak{m}^{-1} \sum_{m \in M} \int (c_0 \theta_m 2^{-jp} \psi_{j,m}^{(|\alpha|)}(x))^2 dx \\ &= c_0^2 2^{-2jp} \mathfrak{m}^{-1} \sum_{m \in M} \int 2^{(2|\alpha|+1)j} \psi^{(|\alpha|)}(2^j x - m)^2 dx \\ &= c_0^2 2^{-2j(p-|\alpha|)} \int \psi^{(|\alpha|)}(u)^2 du \gtrsim c_0^2 2^{-2j(p-|\alpha|)}. \end{aligned}$$

Therefore, there exists a constant  $c_* > 0$  such that

$$|f(h_\theta) - f(h_0)| > 2c_* 2^{-2j(p-|\alpha|)} \tag{64}$$

holds for all for each  $\theta \in \{-1, 1\}^{\mathfrak{m}}$  whenever  $j$  is sufficiently large.

Let  $P_0$  (respectively  $P_\theta$ ) denote the joint distribution of  $\{(W_i, Y_i)\}_{i=1}^n$  with  $Y_i = Th_0(W_i) + u_i$  (respectively  $Y_i = Th_\theta(W_i) + u_i$ ) for the Gaussian NPIR model (45) where, under Condition LB, we may assume that  $X_i$

and  $W_i$  have uniform marginals and that the joint density  $f_{XW}(x, w)$  of  $(X_i, W_i)$  has wavelet expansion

$$f_{XW}(x, w) = \sum_{k=0}^{2^{r_0}-1} \lambda_{r_0} \varphi_{r_0,k}(x) \varphi_{r_0,k}(w) + \sum_{j=r_0}^{\infty} \sum_{k=0}^{2^j-1} \lambda_j \psi_{j,k}(x) \psi_{j,k}(w).$$

Observe that

$$T\psi_{j,k}(w) = \int \psi_{j,k}(x) f_{XW}(x, w) dx = \lambda_j \psi_{j,k}(w)$$

for each  $0 \leq k \leq 2^j - 1$  and each  $j \geq r_0$  ( $r_0$  is fixed) and that  $|\lambda_j| \asymp \nu(2^j)$  by Condition LB(iii). Let  $P^*$  denote the mixture distribution obtained by assigning weight  $2^{-m}$  to  $P_\theta$  for each of the  $2^m$  realizations of  $\theta$ . Lemma G.8 yields

$$\|P^* - P_0\|_{TV}^2 \lesssim \frac{n^2 2^{-4jp} \nu(2^j)^4}{\mathfrak{m}}. \quad (65)$$

In the mildly ill-posed case ( $\nu(2^j) = 2^{j^\varsigma}$ ) we have

$$\|P^* - P_0\|_{TV}^2 \lesssim n^2 2^{-j(4(p+\varsigma)+1)}$$

because  $\mathfrak{m} \asymp 2^j$ . Choose  $2^j \asymp cn^{2/(4(p+\varsigma)+1)}$  with  $c$  sufficiently small so  $\|P^* - P_0\|_{TV} \leq 1 - \epsilon$  for some  $1 > \epsilon > 0$  and all  $n$  large enough, whence:

$$\text{aff}(P^*, P_0) = 1 - \|P^* - P_0\|_{TV} \geq \epsilon \quad (66)$$

for all  $n$  sufficiently large. It now follows by Lemma G.7 and equations (64) and (66) that for all  $n$  sufficiently large, any estimator  $\widehat{g}_n$  of  $f(h)$  obeys the bound

$$\sup_{h \in B_2(p, L)} \mathbb{P}_h \left( |\widehat{g}_n - f(h)| > c_* 2^{-2j(p-|\alpha|)} \right) \geq \epsilon/2 \quad (67)$$

where  $2^{-2j(p-|\alpha|)} \asymp n^{-4(p-|\alpha|)/(4(p+\varsigma)+1)}$ . This is slower than  $n^{-1/2}$  whenever  $p \leq \varsigma + 2|\alpha| + 1/4$ .

In the severely ill-posed case ( $\nu(2^j) = \exp(-\frac{1}{2}2^{\varsigma j})$ ) we choose  $2^j = (c \log n)^{1/\varsigma}$  for some  $c \in (0, 1)$ . This yields  $\|P^* - P_0\|_{TV} = o(1)$  by (65) and hence there exists  $\epsilon > 0$  such that  $\text{aff}(P^*, P_0) \geq \epsilon$  for all  $n$  sufficiently large. Then by Lemma G.7 and equation (64), for all  $n$  sufficiently large, any estimator  $\widetilde{f}_n$  of  $f(h)$  obeys the same bound (67) with  $2^{-2j(p-|\alpha|)} \asymp (\log n)^{-2(p-|\alpha|)/\varsigma}$ .

In the multivariate case ( $d > 1$ ) we let  $\widetilde{\psi}_{j,k,G}(x)$  denote an orthonormal tensor-product wavelet for  $L^2([0, 1]^d)$  at resolution level  $j$ , as described in Appendix E. We may choose a subset  $M$  of  $\{0, \dots, 2^j - 1\}^d$  with  $\mathfrak{m} := \#(M) \asymp 2^{dj}$  for which each  $m \in M$  indexes a tensor-product of interior wavelets of the form  $2^{j/2} \psi(2^j x_l - m_l)$ , which we denote by  $\widetilde{\psi}_{j,m}(x)$ , such that  $\widetilde{\psi}_{j,m}$  and  $\widetilde{\psi}_{j,m'}$  have disjoint support for each  $m, m' \in M$  with  $m \neq m'$ . For each  $\theta \in \{-1, 1\}^{\mathfrak{m}}$  we define

$$h_\theta = \sum_{m \in M} \frac{\theta_m c_0 2^{-jp}}{\sqrt{\mathfrak{m}}} \widetilde{\psi}_{j,m}(x)$$

with  $c_0$  sufficiently small such that  $h_\theta \in B_2(p, L)$  for each  $\theta$ . Let  $h_0(x) = 0$  for all  $x \in [0, 1]^d$ . By disjoint support

of the  $\tilde{\psi}_{j,m}$  and a change of variables, we have:

$$\begin{aligned} |f(h_\theta) - f(h_0)| &= c_0^2 2^{-2jp} \mathfrak{m}^{-1} \sum_{m \in M} \int \left( \prod_{i=1}^d 2^{(2\alpha_i+1)j} \psi^{(\alpha_i)}(2^j x_i - m_i)^2 \right) \mu(x) dx \\ &\gtrsim c_0^2 2^{-2jp} \mathfrak{m}^{-1} \sum_{m \in M} \int \left( \prod_{i=1}^d 2^{(2\alpha_i+1)j} \psi^{(\alpha_i)}(2^j x_i - m_i)^2 \right) dx \gtrsim c_0^2 2^{-2j(p-|\alpha|)}. \end{aligned}$$

Letting  $P_0$ ,  $P_\theta$ , and  $P^*$  be defined analogously to in the univariate case, we let  $X_i$  and  $W_i$  have uniform marginals on  $[0, 1]^d$  and their joint density  $f_{XW}(x, w)$  has wavelet expansion

$$f_{XW}(x, w) = \sum_{j=r_0}^{\infty} \sum_{G \in G_{j,r_0}} \sum_k \lambda_j \tilde{\psi}_{j,k,G}(x) \tilde{\psi}_{j,k,G}(w) \quad (68)$$

with  $|\lambda_j| \asymp \nu(2^j)$ . Lemma G.8 again yields

$$\|P^* - P_0\|_{TV}^2 \lesssim \frac{n^2 2^{-4jp} \nu(2^j)^4}{\mathfrak{m}}.$$

The result follows by choosing  $2^j \asymp cn^{2/(4(p+\varsigma)+d)}$  with sufficiently small  $c$  in the mildly ill-posed case and  $2^j = (c \log n)^{1/\varsigma}$  for some  $c \in (0, 1)$  in the severely ill-posed case.  $\square$

**Lemma G.8.** *Let the Condition LB hold with  $B_2(p, L)$  in place of  $B_\infty(p, L)$  for the NPIR model (45), let  $P^*$  and  $P_0$  be as described in the proof of Theorem G.4, and let  $2^{-jp} \nu(2^j) = o(1)$ . Then:*

$$\|P^* - P_0\|_{TV}^2 \lesssim \frac{n^2 2^{-4jp} \nu(2^j)^4}{\mathfrak{m}}.$$

**Proof of Lemma G.8.** We prove the result for the multivariate case. For each  $\theta \in \{-1, 1\}^m$ , the density of  $P_\theta$  with respect to  $P_0$  is

$$\begin{aligned} \frac{dP_\theta}{dP_0} &= \prod_{i=1}^n \exp \left\{ \frac{-1}{2\sigma_0^2} ([T(h_\theta - h_0)(W_i)]^2 - 2u_i[T(h_\theta - h_0)(W_i)]) \right\} \\ &= \prod_{i=1}^n \exp \left\{ \frac{-1}{2\sigma_0^2} \left( \left[ \sum_{m \in M} \frac{\theta_m c_0 2^{-jp} \lambda_j}{\sqrt{\mathfrak{m}}} \tilde{\psi}_{j,m}(W_i) \right]^2 - 2u_i \left[ \sum_{m \in M} \frac{\theta_m c_0 2^{-jp} \lambda_j}{\sqrt{\mathfrak{m}}} \tilde{\psi}_{j,m}(W_i) \right] \right) \right\}. \end{aligned}$$

Since the  $\tilde{\psi}_{j,m}$  have disjoint support, we have

$$\begin{aligned} \frac{dP_\theta}{dP_0} &= \prod_{i=1}^n \exp \left\{ -\frac{1}{2} \sum_{m \in M} \frac{c_0^2 2^{-2jp} \lambda_j^2}{\sigma_0^2 \mathfrak{m}} \tilde{\psi}_{j,m}(W_i)^2 + \frac{u_i}{\sigma_0} \sum_{m \in M} \theta_m \frac{c_0 2^{-jp} \lambda_j}{\sigma_0 \sqrt{\mathfrak{m}}} \tilde{\psi}_{j,m}(W_i) \right\} \\ &= \prod_{i=1}^n \exp \left\{ \sum_{m \in M} \left( -\frac{1}{2} \Delta_{i,j,m}^2 + \theta_m \frac{u_i}{\sigma_0} \Delta_{i,j,m} \right) \right\} \end{aligned}$$

where  $u_i \sim N(0, \sigma_0^2)$  under  $P_0$  and

$$\Delta_{i,j,m} = \frac{c_0 2^{-jp} \lambda_j}{\sigma_0 \sqrt{\mathfrak{m}}} \tilde{\psi}_{j,m}(W_i).$$

Therefore:

$$\frac{dP_\theta}{dP_0} = \prod_{i=1}^n (1 + A_{i,j}(\theta))$$

where

$$\begin{aligned} A_{i,j}(\theta) &= \exp \left\{ \sum_{m \in M} \left( -\frac{1}{2} \Delta_{i,j,m}^2 + \theta_m \frac{u_i}{\sigma_0} \Delta_{i,j,m} \right) \right\} - 1 \\ &= \left\{ \sum_{m \in M} \exp \left( -\frac{1}{2} \Delta_{i,j,m}^2 + \theta_m \frac{u_i}{\sigma_0} \Delta_{i,j,m} \right) \right\} - 1 \end{aligned}$$

and the second line is again by disjoint support of the  $\tilde{\psi}_{j,m}$  (which implies  $\Delta_{i,j,m}$  is nonzero for at most one  $m$  for each  $i$ ).

Let  $E_0$  be expectation under the measure  $P_0$  and observe that  $E_0[A_{i,j}(\theta)] = 0$  for each  $\theta \in \{-1, 1\}^m$ . For each  $\theta, \theta'$  we define the vector  $\kappa_{\theta, \theta'} \in \mathbb{R}^n$  whose  $i$ th element is:

$$\begin{aligned} \kappa_{\theta, \theta'}(i) &= E_0[A_{i,j}(\theta)A_{i,j}(\theta')] \\ &= \sum_{m \in M} \sum_{m' \in M} E_0 \left[ \exp \left\{ -\frac{1}{2} \Delta_{i,j,m}^2 - \frac{1}{2} \Delta_{i,j,m'}^2 + \frac{u_i}{\sigma_0} (\theta_m \Delta_{i,j,m} + \theta'_{m'} \Delta_{i,j,m'}) \right\} - 1 \right] \\ &= \sum_{m \in M} \sum_{m' \in M} E_0 \left[ e^{\theta_m \theta'_{m'} \Delta_{i,j,m} \Delta_{i,j,m'}} - 1 \right] \\ &= \sum_{m \in M} E_0 \left[ e^{\theta_m \theta'_m \Delta_{i,j,m}^2} - 1 \right]. \end{aligned}$$

where the final line is again by disjoint support of the  $\tilde{\psi}_{j,m}$ . Using  $|\tilde{\psi}_{j,k}| \lesssim 2^{dj/2}$ ,  $E_0[\tilde{\psi}_{j,k}(X_i)^2] = 1$ , and  $m \asymp 2^{dj}$ , it is straightforward to derive the bounds:

$$|\Delta_{i,j,m}| \lesssim \frac{c_0 2^{-jp} \lambda_j}{\sigma_0} \tag{69}$$

$$E_0[\Delta_{i,j,m}^2] = \frac{c_0^2 2^{-2jp} \lambda_j^2}{\sigma_0^2 m}. \tag{70}$$

By Taylor's theorem:

$$\begin{aligned} \kappa_{\theta, \theta'}(i) &= \sum_{m \in M} \theta_m \theta'_m E_0[\Delta_{i,j,m}^2] + \sum_{m \in M} \frac{1}{2} E_0[\Delta_{i,j,m}^4] + r_3(\theta, \theta') \\ &= \sum_{m \in M} \theta_m \theta'_m E_0[\Delta_{i,j,m}^2] + r_2(\theta, \theta') \end{aligned}$$

where  $r_2$  and  $r_3$  are remainder terms. Using the Lagrange remainder formula and (69) and (70), we may deduce that

$$r_2(\theta, \theta') \leq \frac{1}{2} \sum_m E_0 \left[ e^{\Delta_{i,j,m}^2} \Delta_{i,j,m}^4 \right] \lesssim \exp \left\{ \frac{C c_0^2 2^{-2jp} \lambda_j^2}{\sigma_0^2} \right\} \frac{c_0^4 2^{-4jp} \lambda_j^4}{\sigma_0^4 m}.$$

and

$$r_3(\theta, \theta') \leq \frac{1}{6} \sum_m E_0 \left[ e^{\Delta_{i,j,m}^2} \Delta_{i,j,m}^6 \right] \lesssim \exp \left\{ \frac{C c_0^2 2^{-2jp} \lambda_j^2}{\sigma_0^2} \right\} \frac{c_0^6 2^{-6jp} \lambda_j^6}{\sigma_0^6 m}.$$

where  $C$  is a finite positive constant and we again used the fact that  $\Delta_{i,j,m}$  is nonzero for at most one  $m$  for each  $i$ .

Lemma 22 of [Pollard \(2000\)](#) provides the bound:

$$\|P^* - P_0\|_{TV}^2 \leq 2^{-2m} \sum_{\theta \in \{-1,1\}^m} \sum_{\theta' \in \{-1,1\}^m} \Upsilon(\kappa_{\theta,\theta'})$$

where for any vector  $c = (c_1, \dots, c_n)' \in \mathbb{R}^n$  the function  $\Upsilon(c)$  is defined as

$$\Upsilon(c) = -1 + \prod_{i=1}^n (1 + c_i) = \sum_{i=1}^n c_i + \sum_{i_1=1}^n \sum_{i_2=i_1+1}^n c_{i_1} c_{i_2} + \text{higher-order terms}$$

where the higher-order terms are sums over triples, quadruples, etc, with all distinct indices, up to  $c_1 c_2 \dots c_n$ . Therefore:

$$\begin{aligned} \|P^* - P_0\|_{TV}^2 &\leq 2^{-2m} \sum_{\theta, \theta'} \left[ \sum_{i=1}^n \left( \sum_{m \in M} \theta_m \theta'_m E_0[\Delta_{i,j,m}^2] + \sum_{m \in M} \frac{1}{2} E_0[\Delta_{i,j,m}^4] + r_3(\theta, \theta') \right) \right. \\ &\quad \left. + \sum_{i_1=1}^n \sum_{i_2=i_1+1}^n \left\{ \left( \sum_{m \in M} \theta_m \theta'_m E_0[\Delta_{i_1,j,m}^2] + r_2(\theta, \theta') \right) \left( \sum_{m' \in M} \theta_{m'} \theta'_{m'} E_0[\Delta_{i_2,j,m'}^2] + r_2(\theta, \theta') \right) \right\} \right. \\ &\quad \left. + \text{higher order terms} \right]. \end{aligned} \quad (71)$$

Since  $\sum_{\theta, \theta'} \theta_m \theta'_{m'} = 0$  for all  $m, m' \in M$  and  $\sum_{\theta, \theta'} 1 = 2^{2m}$ , the first-order sum in (71) is:

$$\begin{aligned} &2^{-2m} \sum_{i=1}^n \sum_{\theta, \theta'} \left( \sum_m \theta_m \theta'_m E_0[\Delta_{i,j,m}^2] + \frac{1}{2} \sum_m E_0[\Delta_{i,j,m}^4] + r_3(\theta, \theta') \right) \\ &= \frac{n}{2} \sum_m E_0[\Delta_{i,j,m}^4] + n 2^{-2m} \sum_{\theta, \theta'} r_3(\theta, \theta') \\ &\lesssim \frac{n c_0^4 2^{-4jp} \lambda_j^4}{\sigma_0^4 m} \left( 1 + \exp \left\{ \frac{c_0^2 2^{-2jp} \lambda_j^2}{\sigma_0^2} \right\} \frac{c_0^2 2^{-2jp} \lambda_j^2}{\sigma_0^2 m} \right). \end{aligned} \quad (72)$$

Also observe that

$$\sum_{\theta} \sum_{\theta'} (\theta_m \theta'_m) (\theta_{m'} \theta'_{m'}) = \begin{cases} 0 & \text{if } m \neq m' \\ 2^{2m} & \text{if } m = m'. \end{cases}$$

The second-order sum in (71) is therefore:

$$\begin{aligned} &\frac{n(n-1)}{2} \left( m E_0[\Delta_{i,j,m}^2]^2 + O(m \max_{m \in M} E[\Delta_{i,j,m}^2] \times \max_{\theta, \theta'} r_2(\theta, \theta')) + O(\max_{\theta, \theta'} r_2(\theta, \theta')^2) \right) \\ &\asymp \frac{n^2 c_0^4 2^{-4jp} \lambda_j^4}{\sigma_0^4 m} \left( 1 + O \left( \exp \left\{ \frac{2C c_0^2 2^{-2jp} \lambda_j^2}{\sigma_0^2} \right\} \left( 2^{-2jp} \lambda_j^2 + \frac{2^{-4jp} \lambda_j^4}{m} \right) \right) \right). \end{aligned} \quad (73)$$

The higher-order terms in (71) will be of asymptotically smaller order because  $2^{-jp} \lambda_j \asymp 2^{-jp} \nu(2^j) = o(1)$ . Substituting (72) and (73) into (71) yields:

$$\|P^* - P_0\|_{TV}^2 \lesssim \frac{n^2 2^{-4jp} \lambda_j^4}{m} (1 + o(1)) \asymp \frac{n^2 2^{-4jp} \nu(2^j)^4}{m}$$

as required.  $\square$

## G.8 Proofs for Appendix E

**Proof of Lemma E.1.** Part (a) is equation (3.4) on p. 141 of DeVore and Lorentz (1993). For part (b), let  $v \in \mathbb{R}^J$ , let  $f_X(x)$  denote the density of  $X_i$  and let  $\underline{f}_X = \inf_x f_X(x)$  and  $\bar{f}_X = \sup_x f_X(x)$ . Then for any  $v \in \mathbb{R}^J$ :

$$\begin{aligned} v' E[\psi^J(X_i)\psi^J(X_i)']v &\geq \underline{f}_X \int_0^1 (\psi^J(x)'v)^2 dx \\ &\geq \underline{f}_X c_1^2 \frac{\min_{-r+1 \leq j \leq m} (t_{j+r} - t_r)}{r} \|v\|_{\ell^2}^2 \\ &\geq \underline{f}_X c_1^2 \frac{c_2 J^{-1}}{r} \|v\|_{\ell^2}^2 \end{aligned}$$

for some finite positive constant  $c_1$ , where the first inequality is by Assumption 1(i), the second is by Theorem 4.2 (p. 145) of DeVore and Lorentz (1993) with  $p = 2$ , and the third is by uniform boundedness of the mesh ratio. By the variational characterization of eigenvalues of selfadjoint matrices, we have:

$$\lambda_{\min}(G_\psi) = \min_{v \in \mathbb{R}^J, v \neq 0} \frac{v' E[\psi^J(X_i)\psi^J(X_i)']v}{\|v\|_{\ell^2}^2} \geq \underline{f}_X c_1^2 \frac{c_2 J^{-1}}{r}$$

This establishes the upper bound on  $\lambda_{\min}(G_\psi)^{-1}$ . The proof of the lower bound for  $\lambda_{\max}(G_\psi)^{-1}$  follows analogously by Theorem 4.2 (p. 145) of DeVore and Lorentz (1993) with  $p = 2$ . Part (c) then follows directly from part (b).  $\square$

**Proof of Lemma E.2.** The  $\ell^1$  norm of the tensor product of vectors equals the product of the  $\ell^1$  norms of the factors, whence part (a) follows from Lemma E.1. As  $\psi^J(x)$  is formed as the tensor-product of univariate B-splines, each element of  $\psi^J(x)$  is of the form  $\prod_{l=1}^d \psi_{J_{i_l}}(x_l)$  where  $\psi_{J_{i_l}}(x_l)$  denotes the  $i_l$ th element of the vector of univariate B-splines. Let  $v \in \mathbb{R}^J$ . We may index the elements of  $v$  by the multi-indices  $i_1, \dots, i_d \in \{1, \dots, m+r\}^d$ . By boundedness of  $f_X$  away from zero and Fubini's theorem, we have:

$$\begin{aligned} v' E[\psi^J(X_i)\psi^J(X_i)']v &\geq \underline{f}_X \int_0^1 \cdots \int_0^1 \left( \sum_{i_1, \dots, i_d} v_{i_1, \dots, i_d} \prod_{l=1}^d \psi_{J_{i_l}}(x_l) \right)^2 dx_1 \cdots dx_d \\ &= \underline{f}_X \int_0^1 \cdots \int_0^1 \sum_{i_2, \dots, i_d} \sum_{j_2, \dots, j_d} \left( \prod_{l=2}^d \psi_{J_{i_l}}(x_l) \right) \left( \prod_{l=2}^d \psi_{J_{j_l}}(x_l) \right) \\ &\quad \left\{ \int_0^1 \sum_{i_1} \sum_{j_1} v_{i_1, \dots, i_d} v_{j_1, \dots, j_d} \psi_{J_{i_1}}(x_1) \psi_{J_{j_1}}(x_1) dx_1 \right\} dx_2 \cdots dx_d. \end{aligned}$$

Applying Theorem 4.2 (p. 145) of DeVore and Lorentz (1993) to the term in braces, and repeating for  $x_2, \dots, x_d$ , we have:

$$\begin{aligned} v' E[\psi^J(X_i)\psi^J(X_i)']v &\geq \underline{f}_X \left( c_1^2 \frac{\min_{-r+1 \leq j \leq m} (t_{j+r} - t_r)}{r} \right)^d \|v\|_{\ell^2}^2 \\ &\geq \underline{f}_X c_1^{2d} \frac{c_2^d J^{-1}}{r^d} \|v\|_{\ell^2}^2 \end{aligned}$$

where the second inequality is by uniform boundedness of the mesh ratio. The rest of the proof follows by identical arguments to Lemma E.1.  $\square$

**Proof of Lemma E.3.** Each of the interior  $\varphi_{j,k}$  and  $\psi_{j,k}$  have support  $[2^{-j}(-N+1+k), 2^{-j}(N+k)]$ , therefore  $\varphi_{j,k}(x) \neq 0$  (respectively  $\psi_{j,k}(x) \neq 0$ ) for less than or equal to  $2N$  interior  $\varphi_{j,k}$  (resp.  $\psi_{j,k}$ ) and for any  $x \in [0, 1]$ . Further, there are only  $N$  left and right  $\varphi_{j,k}$  and  $\psi_{j,k}$ . Therefore,  $\varphi_{j,k}(x) \neq 0$  (respectively  $\psi_{j,k}(x) \neq 0$ ) for less than or equal to  $3N$  of the  $\varphi_{j,k}$  (resp.  $\psi_{j,k}$ ) at resolution level  $j$  for each  $x \in [0, 1]$ . By construction of the basis, each of  $\varphi_{j,k}$ ,  $\varphi_{j,k}^l$ ,  $\psi_{j,k}^l$  for  $k = 0, \dots, N-1$  and  $\varphi_{j,-k}^r$ ,  $\psi_{j,-k}^r$  for  $k = 1, \dots, N$  are continuous and therefore attain a finite maximum on  $[0, 1]$ . Therefore, each of the  $\varphi_{j,k}$  and  $\psi_{j,k}$  are uniformly bounded by some multiple of  $2^{j/2}$  and so:

$$\xi_{\psi,J} \lesssim 3N \times \left( \underbrace{2^{L_0/2}}_{\text{for the } \varphi_{L_0,k}} + \underbrace{2^{L_0/2}}_{\text{for the } \psi_{L_0,k}} + \dots + \underbrace{2^{L/2}}_{\text{for the } \psi_{L,k}} \right) \lesssim 2^{L/2}$$

The result then follows because  $J = 2^{L+1}$ . For part (b), because  $f_X$  is uniformly bounded away from 0 and  $\infty$  and the wavelet basis is orthonormal for  $L^2[0, 1]$ , we have

$$v' E[\psi^J(X_i)\psi^J(X_i)'] v \asymp v' \left( \int_0^1 \psi^J(x)\psi^J(x)' dx \right) v = \|v\|_{\ell^2}^2$$

and so all eigenvalues of  $G_\psi$  are uniformly (in  $J$ ) bounded away from 0 and  $\infty$ . Part (c) follows directly.  $\square$

**Proof of Lemma E.4.** Lemma E.3 implies that each of the factor vectors in the tensor product at level  $j$  has  $\ell^1$  norm of order  $O(2^{dj/2})$  uniformly for  $x = (x_1, \dots, x_d)' \in [0, 1]^d$  and in  $j$ . There are at most  $2^d$  such tensor products at each resolution level. Therefore,  $\xi_{\psi,J} = O(2^{dL/2}) = O(\sqrt{J})$  since  $J = O(2^{dL})$ . Parts (b) and (c) follow by the same arguments of the proof of Lemma E.3 since the tensor-product basis is orthonormal for  $L^2([0, 1]^d)$ .  $\square$

## G.9 Proofs for Appendix F

**Proof of Lemma F.2.**  $\|A^{-1} - I_r\|_{\ell^2} = \|A^{-1}(A - I_r)\|_{\ell^2} \leq \|A^{-1}\|_{\ell^2} \|A - I_r\|_{\ell^2}$ .  $\square$

**Proof of Lemma F.4.** The first assertion is immediate by Theorem 3.3 of Stewart (1977) and definition of  $A_l^-$  and  $B_l^-$ . For the second part, Weyl's inequality implies that  $s_{\min}(B) \geq \frac{1}{2}s_{\min}(A)$  whenever  $\|A - B\|_{\ell^2} \leq \frac{1}{2}s_{\min}(A)$ .  $\square$

**Proof of Lemma F.5.**  $\|A_l^-\|_{\ell^2}^2 = \lambda_{\max}(A_l^-(A_l^-)') = \lambda_{\max}((A' A)^{-1}) = 1/\lambda_{\min}(A' A) = s_{\min}(A)^{-2}$ .  $\square$

**Proof of Lemma F.6.** The result follows from Li, Li, and Cui (2013) (see also Stewart (1977)).  $\square$

**Proof of Lemma F.7.** We prove the results for  $\widehat{S}^o$ ; convergence of  $\widehat{G}_\psi^o$  and  $\widehat{G}_b^o$  is proved in Lemma 2.1 of Chen and Christensen (2015). Note that

$$\widehat{S}^o - S^o = \sum_{i=1}^n n^{-1} G_b^{-1/2} \{b^K(W_i)\psi^J(X_i)' - E[b^K(W_i)\psi^J(X_i)']\} G_\psi^{-1/2} =: \sum_{i=1}^n \Xi_i^o$$

where  $\|\Xi_i\|_{\ell^2} \leq 2n^{-1}\zeta_{b,K}\zeta_{\psi,J}$ . Also,

$$\begin{aligned} \left\| \sum_{i=1}^n E[\Xi_i^o \Xi_i^{o'}] \right\|_{\ell^2} &\leq n^{-1} \|E[G_b^{-1/2} b^K(W_i) \psi^J(X_i)' G_\psi^{-1} \psi^J(X_i) b^K(W_i)' G_b^{-1/2}]\|_{\ell^2} \\ &\leq n^{-1} \zeta_{\psi,J}^2 \|E[G_b^{-1/2} b^K(W_i) b^K(W_i)' G_b^{-1/2}]\|_{\ell^2} \\ &= n^{-1} \zeta_{\psi,J}^2 \|I\|_{\ell^2} \\ &= n^{-1} \zeta_{\psi,J}^2 \end{aligned}$$

by the fact that  $\|I\|_{\ell^2} = 1$ . An identical argument yields the bound  $\|\sum_{i=1}^n E[\Xi_i^{o'} \Xi_i^o]\|_{\ell^2} \leq n^{-1} \zeta_{b,K}^2$ . Applying a Bernstein inequality for random matrices (Tropp, 2012, Theorem 1.6) yields

$$\mathbb{P}\left(\|\widehat{S}^o - S^o\|_{\ell^2} > t\right) \leq 2 \exp\left\{\log K - \frac{-t^2/2}{(\zeta_{b,K}^2 \vee \zeta_{\psi,J}^2)/n + 2\zeta_{b,K}\zeta_{\psi,J}t/(3n)}\right\}.$$

The convergence rate  $\|\widehat{S}^o - S^o\|_{\ell^2}$  from this inequality under appropriate choice of  $t$ .  $\square$

**Proof of Lemma F.9.** Let  $\widetilde{b}^K(x) = G_b^{-1/2} b^K(x)$  and denote  $\widetilde{b}^K(x)' = (\widetilde{b}_{K1}(x), \dots, \widetilde{b}_{KK}(x))$ . As the summands have expectation zero, we have

$$\begin{aligned} E\left[\|G_b^{-1/2}\{B'(H_0 - H_J)/n - E[b^K(W_i)(h_0(X_i) - h_J(X_i))]\}\|_{\ell^2}^2\right] &\leq \frac{1}{n} E\left[\sum_{k=1}^K (\widetilde{b}_{Kk}(W_i))^2 (h_0(X_i) - h_J(X_i))^2\right] \\ &\leq \frac{K}{n} \|h_0 - h_J\|_{\infty}^2 \wedge \frac{\zeta_{b,K}^2}{n} \|h_0 - h_J\|_{L^2(X)}^2. \end{aligned}$$

The result follows by Chebyshev's inequality.  $\square$

**Proof of Lemma F.10.** We begin by rewriting the target in terms of the orthonormalized matrices

$$\begin{aligned} (\widehat{G}_b^{-1/2} \widehat{S})_l^- \widehat{G}_b^{-1/2} G_b^{1/2} - (G_b^{-1/2} S)_l^- &= G_\psi^{-1/2} \{(\widehat{S}^{o'} \widehat{G}_b^{o-1} \widehat{S}^o)^{-1} \widehat{S}^{o'} \widehat{G}_b^{o-1} - (S^{o'} S^o)^{-1} S^{o'}\} \\ &= G_\psi^{-1/2} \{((\widehat{G}_b^o)^{-1/2} \widehat{S}^o)_l^- (\widehat{G}_b^o)^{-1/2} - (S^o)_l^-\}. \end{aligned} \quad (74)$$

We first bound the term in braces. By the triangle inequality,

$$\begin{aligned} &\|((\widehat{G}_b^o)^{-1/2} \widehat{S}^o)_l^- (\widehat{G}_b^o)^{-1/2} - (S^o)_l^-\|_{\ell^2} \\ &\leq \|((\widehat{G}_b^o)^{-1/2} \widehat{S}^o)_l^- - (S^o)_l^-\|_{\ell^2} \|(\widehat{G}_b^o)^{-1/2}\|_{\ell^2} + \|(\widehat{G}_b^o)^{-1/2} - I\|_{\ell^2} \|(S^o)_l^-\|_{\ell^2}. \end{aligned} \quad (75)$$

Lemma F.7 provides that

$$\|\widehat{G}_b^o - I_K\|_{\ell^2} = O_p(\zeta_{b,K} \sqrt{(\log K)/n}) \quad (76)$$

$$\|\widehat{S}^o - S^o\|_{\ell^2} = O_p((\zeta_{b,K} \vee \zeta_{\psi,J}) \sqrt{(\log K)/n}). \quad (77)$$

Let  $\mathcal{A}_n$  denote the event upon which  $\|\widehat{G}_b^o - I_K\|_{\ell^2} \leq \frac{1}{2}$  and note that  $\mathbb{P}(\mathcal{A}_n^c) = o(1)$  because  $\|\widehat{G}_b^o - I_K\|_{\ell^2} = o_p(1)$ . Then by Lemmas F.2 and F.3 we have

$$\begin{aligned} \|(\widehat{G}_b^o)^{-1/2} - I_K\|_{\ell^2} &\leq \sqrt{2} \|(\widehat{G}_b^o)^{1/2} - I_K\|_{\ell^2} \\ &\leq \frac{2}{1 + \sqrt{2}} \|\widehat{G}_b^o - I_K\|_{\ell^2} \end{aligned}$$

on  $\mathcal{A}_n$ . It follows by expression (76) and the fact that  $\mathbb{P}(\mathcal{A}_n^c) = o(1)$  that

$$\|(\widehat{G}_b^o)^{-1/2} - I\|_{\ell^2} = O_p(\zeta_{b,K} \sqrt{(\log K)/n}) \quad (78)$$

which in turn implies that  $\|(\widehat{G}_b^o)^{-1/2}\| = 1 + o_p(1)$ .

To bound  $\|((\widehat{G}_b^o)^{-1/2} \widehat{S}^o)_l^- - (S^o)_l^-\|_{\ell^2}$ , it follows by equations (77) and (78) and the fact that  $\|S^o\|_{\ell^2} \leq 1$  that:

$$\begin{aligned} \|(\widehat{G}_b^o)^{-1/2} \widehat{S}^o - S^o\|_{\ell^2} &\leq \|(\widehat{G}_b^o)^{-1/2} - I_K\|_{\ell^2} \|\widehat{S}^o\|_{\ell^2} + \|\widehat{S}^o - S^o\|_{\ell^2} \\ &= O_p((\zeta_{b,K} \vee \zeta_{\psi,J}) \sqrt{(\log K)/n}). \end{aligned} \quad (79)$$

Let  $\mathcal{A}_{n,1} \subseteq \mathcal{A}_n$  denote the event on which  $\|(\widehat{G}_b^o)^{-1/2} \widehat{S}^o - S^o\|_{\ell^2} \leq \frac{1}{2} s_{JK}$  and note that  $\mathbb{P}(\mathcal{A}_{n,1}^c) = o(1)$  by virtue of the condition  $s_{JK}^{-1}(\zeta_{b,K} \vee \zeta_{\psi,J}) \sqrt{(\log K)/n} = o(1)$ . Lemma F.4 provides that

$$\|((\widehat{G}_b^o)^{-1/2} \widehat{S}^o)_l^- - (S^o)_l^-\|_{\ell^2} \leq 2(1 + \sqrt{5}) s_{JK}^{-2} \|(\widehat{G}_b^o)^{-1/2} \widehat{S}^o - S^o\|_{\ell^2} \quad (80)$$

on  $\mathcal{A}_{n,1}$ , and so

$$\|((\widehat{G}_b^o)^{-1/2} \widehat{S}^o)_l^- - (S^o)_l^-\|_{\ell^2} = O_p(s_{JK}^{-2} (\zeta_{b,K} \vee \zeta_{\psi,J}) \sqrt{(\log K)/n}) \quad (81)$$

by (79) and (80). It follows from equations (81) and (74) that:

$$\|(\widehat{G}_b^{-1/2} \widehat{S})_l^- \widehat{G}_b^{-1/2} G_b^{1/2} - (G_b^{-1/2} S)_l^-\|_{\ell^2} = O_p\left(s_{JK}^{-2} (\zeta_{b,K} \vee \zeta_{\psi,J}) \sqrt{(\log K)/(ne_J)}\right)$$

which, together with the condition  $J \leq K = O(J)$ , proves part (a). Part (b) follows similarly.

For part (c), we pre and post multiply terms in the product by  $G_b^{-1/2}$  and  $G_\psi^{-1/2}$  to obtain:

$$\begin{aligned} &\|G_b^{-1/2} S \{(\widehat{G}_b^{-1/2} \widehat{S})_l^- \widehat{G}_b^{-1/2} G_b^{1/2} - (G_b^{-1/2} S)_l^-\}\|_{\ell^2} \\ &= \|S^o [\widehat{S}^o (\widehat{G}_b^o)^- \widehat{S}^o]^- \widehat{S}^o (\widehat{G}_b^o)^- - S^o [S^o S^o]^{-1} S^o\|_{\ell^2} \\ &\leq \|S^o [\widehat{S}^o (\widehat{G}_b^o)^- \widehat{S}^o]^- \widehat{S}^o (\widehat{G}_b^o)^{-1/2} ((\widehat{G}_b^o)^{-1/2} - I_K)\|_{\ell^2} \\ &\quad + \|(S^o - (\widehat{G}_b^o)^{-1/2} \widehat{S}^o) [\widehat{S}^o (\widehat{G}_b^o)^- \widehat{S}^o]^- \widehat{S}^o (\widehat{G}_b^o)^{-1/2}\|_{\ell^2} \\ &\quad + \|(\widehat{G}_b^o)^{-1/2} \widehat{S}^o [\widehat{S}^o (\widehat{G}_b^o)^- \widehat{S}^o]^- \widehat{S}^o (\widehat{G}_b^o)^{-1/2} - S^o [S^o S^o]^{-1} S^o\|_{\ell^2}. \end{aligned} \quad (82)$$

Note that  $\|((\widehat{G}_b^o)^{-1/2} \widehat{S}^o)_l^-\|_{\ell^2} \leq 2s_{JK}^{-1}$  on  $\mathcal{A}_{n,1}$  by Lemma F.5, so

$$\|((\widehat{G}_b^o)^{-1/2} \widehat{S}^o)_l^-\|_{\ell^2} = O_p(s_{JK}^{-1}). \quad (83)$$

It follows by substituting (78), (79), and (83) into (82) that

$$\begin{aligned} &\|G_b^{-1/2} S \{(\widehat{G}_b^{-1/2} \widehat{S})_l^- \widehat{G}_b^{-1/2} G_b^{1/2} - (G_b^{-1/2} S)_l^-\}\|_{\ell^2} \\ &\leq O_p(s_{JK}^{-1} (\zeta_{b,K} \vee \zeta_{\psi,J}) \sqrt{(\log K)/n}) \\ &\quad + \|(\widehat{G}_b^o)^{-1/2} \widehat{S}^o [\widehat{S}^o (\widehat{G}_b^o)^- \widehat{S}^o]^- \widehat{S}^o (\widehat{G}_b^o)^{-1/2} - S^o [S^o S^o]^{-1} S^o\|_{\ell^2}. \end{aligned} \quad (84)$$

The remaining term on the right-hand side of (84) is the  $\ell^2$  norm of the difference between the orthogonal projection matrices associated with  $S^o$  and  $(\widehat{G}_b^o)^{-1/2} \widehat{S}^o$ . Applying Lemma F.6, we obtain:

$$\|(\widehat{G}_b^o)^{-1/2} \widehat{S}^o [\widehat{S}^o (\widehat{G}_b^o)^- \widehat{S}^o]^- \widehat{S}^o (\widehat{G}_b^o)^{-1/2} - S^o [S^o S^o]^{-1} S^o\|_{\ell^2} \leq 2s_{JK}^{-1} \|(\widehat{G}_b^o)^{-1/2} \widehat{S}^o - S^o\|_{\ell^2}$$

on  $\mathcal{A}_{n,1}$ . Result (c) then follows by (79) and (84). □

## References

- Chen, X. and T. M. Christensen (2015). Optimal uniform convergence rates and asymptotic normality for series estimators under weak dependence and weak conditions. *Journal of Econometrics* 188(2), 447 – 465.
- Chen, X. and D. Pouzo (2015). Sieve Wald and QLR inferences on semi/nonparametric conditional moment models. *Econometrica* 83(3), 1013–1079.
- Chen, X. and M. Reiss (2011). On rate optimality for ill-posed inverse problems in econometrics. *Econometric Theory* 27(3), 497–521.
- Chernozhukov, V., D. Chetverikov, and K. Kato (2014). Anti-concentration and honest, adaptive confidence bands. *The Annals of Statistics* 42(5), 1787–1818.
- Cohen, A. (2003). *Numerical Analysis of Wavelet Methods*. Elsevier.
- DeVore, R. A. and G. G. Lorentz (1993). *Constructive Approximation*. Grundlehren der mathematischen Wissenschaften 303. Springer-Verlag, Berlin.
- Hausman, J. A. and W. K. Newey (1995). Nonparametric estimation of exact consumers surplus and deadweight loss. *Econometrica* 63(6), pp. 1445–1476.
- Huang, J. Z. (2003). Local asymptotics for polynomial spline regression. *The Annals of Statistics* 31(5), 1600–1635.
- Li, B., W. Li, and L. Cui (2013). New bounds for perturbation of the orthogonal projection. *Calcolo* 50(1), 69–78.
- Pollard, D. (2000). *Asymptopia*, Chapter 3: Distances and affinities between measures. Unpublished Manuscript.
- Pollard, D. (2002). *A User’s Guide to Measure Theoretic Probability*. Cambridge University Press, Cambridge.
- Schumaker, L. L. (2007). *Spline Functions: Basic Theory*. Cambridge University Press, Cambridge.
- Stewart, G. W. (1977). On the perturbation of pseudo-inverses, projections and linear least squares problems. *SIAM Review* 19(4), 634–662.
- Tropp, J. A. (2012). User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics* 12, 389–434.
- Tsybakov, A. B. (2009). *Introduction to Nonparametric Estimation*. Springer, New York.
- van der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes*. Springer, New York.
- Yu, B. (1997). Assouad, Fano, and Le Cam. In D. Pollard, E. Torgersen, and G. L. Yang (Eds.), *Festschrift for Lucien Le Cam*. Springer-Verlag New York.