Submitted to the Annals of Statistics

PROOFS FOR "POSTERIOR CONSISTENCY OF NONPARAMETRIC CONDITIONAL MOMENT RESTRICTED MODELS"

BY YUAN LIAO AND WENXIN JIANG

Princeton University and Northwestern University

This document contains the technical proofs of all the results developed in the main paper Liao and Jiang (2011).

Throughout the proof, we denote by C as a generic constant.

APPENDIX A: PROOFS FOR SECTION 2

A.1. Lemma 2.1. For any $\delta_n > 0$,

$$E\{P(G(g_b) - \inf_{g\in\mathcal{H}} G(g) > 5\delta_n|X^n)\} \leq P(\sup_{g\in\mathcal{F}_n} |\bar{G}(g) - G(g)| \geq \delta_n)$$
$$+\frac{e^{-2n\delta_n}}{\pi(G(g_b) - \inf_{g\in\mathcal{H}} G(g) < \delta_n \cap g_b \in \mathcal{F}_n)} + EP(g_b \in \mathcal{F}_n^c|X^n). \tag{A.1}$$

In addition,

$$EP(g_b \in \mathcal{F}_n^c|X^n) \leq P(\sup_{g\in\mathcal{F}_n} |\bar{G}(g) - G(g)| \geq \delta_n)$$
$$+\frac{\pi(\mathcal{F}_n^c)e^{2n\delta_n}}{\pi(G(g_b) - \inf_{g\in\mathcal{H}} G(g) < \delta_n \cap g_b \in \mathcal{F}_n)}. \tag{A.2}$$

PROOF. With probability one, we have

$$P(G(g_b) - \inf_{g\in\mathcal{H}} G(g) > 5\delta_n|X^n) \leq P(G(g_b) - \inf_{g\in\mathcal{H}} G(g) > 5\delta_n, g_b \in \mathcal{F}_n|X^n)$$
$$+P(g_b \in \mathcal{F}_n^c|X^n)$$
$$\leq P(G(g_b) - \inf_{g\in\mathcal{H}} G(g) > 5\delta_n|g_b \in \mathcal{F}_n, X^n) + P(g_b \in \mathcal{F}_n^c|X^n),$$

which implies

$$P(G(g_b) - \inf_{g\in\mathcal{H}} G(g) > 5\delta_n|X^n) \leq EP(G(g_b) - \inf_{g\in\mathcal{H}} G(g) > 5\delta_n|g_b \in \mathcal{F}_n, X^n)$$
$$+EP(g_b \in \mathcal{F}_n^c|X^n).$$

AMS 2000 subject classifications: Primary 62F15, 62G08, 62G20; secondary 62P20
Keywords and phrases: identified region, limited information likelihood, sieve approximation, nonparametric instrumental variable, ill-posed problem, partial identification, Bayesian inference

For the original prior $\pi(.)$, define a truncated version as:

$$\tilde{\pi}(b) = \frac{\pi(b)I(g_b \in \mathcal{F}_n)}{\pi(g_b \in \mathcal{F}_n)}.$$

Apply Proposition 6 of Jiang and Tanner (2008) to

$$EP(G(g_b) - \inf_{g \in \mathcal{H}} G(g) > 5\delta_n | g_b \in \mathcal{F}_n, X^n)$$

using the truncated prior $\tilde{\pi}(.)$, and note that $\tilde{\pi}(\mathcal{F}_n^c) = 0$. Then we obtain

$$EP(G(g_b) - \inf_{g \in \mathcal{H}} G(g) > 5\delta_n | X^n) \le P(\sup_{g_b \in \mathcal{F}_n} |\bar{G}(g_b) - G(g_b)| \ge \delta_n)$$

$$+ e^{-2n\delta_n} / \tilde{\pi}(G(g_b) - \inf_{g \in \mathcal{H}} G(g) < \delta_n) + EP(g_b \in \mathcal{F}_n^c | X^n).$$

Note that $\tilde{\pi}(A) = \pi(A | g_b \in \mathcal{F}_n) \ge \pi(A, g_b \in \mathcal{F}_n)$. Then

$$EP(G(g_b) - \inf_{g \in \mathcal{H}} G(g) > 5\delta_n | X^n) \le P(\sup_{g_b \in \mathcal{F}_n} |\bar{G}(g_b) - G(g_b)| \ge \delta_n)$$

$$+ e^{-2n\delta_n} / \pi(G(g_b) - \inf_{g \in \mathcal{H}} G(g) < \delta_n, g_b \in \mathcal{F}_n) + EP(g_b \in \mathcal{F}_n^c | X^n).$$

To prove (A.2), Let $\Delta = \frac{1}{2} \sup_{g_b \in \mathcal{F}_n} |\bar{G}(g_b) - G(g_b)|$, and $M_n = \{b \in \mathbb{R}^{q_n} : g_b \in \mathcal{F}_n\}$. For any $\delta_n > 0$,

$$\int_{\mathbb{R}^{q_n}} \exp(-\frac{n}{2}\bar{G}(g_b)) db \ge \int_{M_n} \exp(-\frac{n}{2}\bar{G}(g_b)) db$$

$$\ge \exp(-n\Delta) \int_{M_n} I(G(g_b) < \delta_n) \exp(-\frac{n}{2} G(g_b)) \pi(db)$$

$$\ge \exp(-n\Delta - \frac{n}{2}\delta_n) \pi(G(g_b) < \delta_n, b \in M_n)$$

Therefore

$$P(g_b \in \mathcal{F}_n^c | X^n) = \frac{\int_{M_n^c} \exp(-\frac{n}{2}\bar{G}(g_b)) \pi(db)}{\int_{\mathbb{R}^{q_n}} \exp(-\frac{n}{2}\bar{G}(g_b)) \pi(db)} \le \frac{\pi(\mathcal{F}_n^c) e^{n\Delta + n\delta_n/2}}{\pi(G(g_b) < \delta_n, g_b \in \mathcal{F}_n)}$$

Now $EP(g_b \in \mathcal{F}_n^c | X^n) \le E[I(\Delta > \delta_n/2)] + E[P(g_b \in \mathcal{F}_n^c | X^n) I(\Delta < \delta_n/2)]$

$$\le P(\sup_{\mathcal{F}_n} |\bar{G}(g_b) - G(g_b)| > \delta_n) + \frac{\pi(\mathcal{F}_n^c) e^{2n\delta_n}}{\pi(G(g_b) < \delta_n, g_b \in \mathcal{F}_n)}.$$

Q.E.D.

APPENDIX B: PROOFS FOR SECTION 3

B.1. Lemmas.

LEMMA B.1. *Suppose $\pi(.)$ is continuous on $\mathcal{F}_n$. Suppose there exists a sequence $\delta_n = O(1)$, and $g_0 \in \Theta_I$ with $g^*_{q_n} = \sum_{i\leq q_n} b^*_i\phi_i$ being its sieve approximation such that $\|g^*_{q_n} - g_0\|_s = o(1)$ and $G(g^*_{q_n}) = o(\delta_n)$. Then there exists a constant $C > 0$, for any $g_b = \sum_{i\leq q_n} b_i\phi_i \in \mathcal{F}_n$, and large enough n,*

$$\pi(\mathcal{F}_n \cap G(g_b) < \delta_n) \geq \pi(\tilde{b})\left(\frac{C\delta_n}{\gamma_n\sqrt{q_n}}\right)^{q_n},$$

where γ_n is defined in equation (3.9) in the main paper, and $\tilde{b}$ is some point satisfying $\|b^* - \tilde{b}\| \leq C\delta_n/\gamma_n$.

PROOF. By the triangular inequality, and the fact that $G(g^*_{q_n}) = o(\delta_n)$, we have

$$\{G(g_b) < \delta_n\} \supset \{|G(g_b) - G(g^*_{q_n})| < \delta_n/2\}.$$

for all large n. In addition, $g^*_{q_n} \in \mathcal{F}_n$. By Assumption 3.2, for all $g_b \in \mathcal{F}_n$,

$$\begin{aligned}
&|G(g_b) - G(g^*_{q_n})| \leq \int \left|[E(\rho(Z, g_b)|w)]^2 - [E(\rho(Z, g^*_{q_n})|w)]^2\right| dF(w) \\
&\leq 2 \sup_{g\in\mathcal{F}_n, w\in[0,1]^d} |E(\rho(Z, g)|W = w)| \int |E(\rho(Z, g_b) - \rho(Z, g^*_{q_n})|W = w)| dF(w) \\
&\leq C \sup_{g\in\mathcal{F}_n, w\in[0,1]^d} |E(\rho(Z, g)|W = w)| E|g_b(X) - g^*_{q_n}(X)| \\
&\leq C \sup_{g\in\mathcal{F}_n, w\in[0,1]^d} |E(\rho(Z, g)|W = w)| \|b - b^*\|,
\end{aligned}$$

where we used the fact that $E|g_b(X) - g^*_{q_n}(X)| \leq \sqrt{E(|g_b(X) - g^*_{q_n}(X)|^2)} = \|b - b^*\|$ since the basis functions are orthonormal. Therefore,

$$\begin{aligned}
&\{\mathcal{F}_n \cap G(g_b) < \delta_n\} \\
\supset\ &\{\sup_{g\in\mathcal{F}_n, w\in[0,1]^d} |E(\rho(Z, g)|W = w)| \|b - b^*\| < C\delta_n, \max_{i\leq q_n} |b_i| \leq B_n\}. \\
\supset\ &\{\gamma_n\|b - b^*\| < C\delta_n, \max_{i\leq q_n} |b_i| \leq B_n\} \\
\supset\ &\{\gamma_n\|b - b^*\| < C\delta_n\}
\end{aligned}$$

for all large n. In the definition of γ_n, we add one to prevent it from being close to zero. Therefore, for all large n and continuous $\pi(.)$, we have, by the integral mean value theorem,

$$\pi(\mathcal{F}_n \cap G(g_b) < \delta_n) \quad \geq \quad \pi(\|b - b^*\| < C\delta_n/\gamma_n) = \pi(\tilde{b})\mu(\|b - b^*\| < C\delta_n/\gamma_n)$$

$$\begin{aligned} &\geq \pi(\tilde{b})\mu(\max_{i\leq q_n}|b_i - b_i^*| < \frac{C\delta_n}{\gamma_n\sqrt{q_n}}) \\ &= \pi(\tilde{b})\left(\frac{C\delta_n}{\gamma_n\sqrt{q_n}}\right)^{q_n}, \end{aligned}$$

where $\tilde{b}$ belongs to the l_2 ball $\{b : \|b - b^*\| \leq C\delta_n/\gamma_n\}$, and $\mu(.)$ denotes the Lebesgue measure.

LEMMA B.2. *For the thin tail prior defined in Section 3.2, if $\delta_n \leq CB_n^r/n$ for a large enough constant $C > 0$,*

$$\pi(g_b \in \mathcal{F}_n^c) \leq e^{-4n\delta_n}$$

PROOF. By the definition of the thin prior, it is straightforward to verify that,

$$\pi(b \notin \mathcal{F}_n) \leq \pi(\|b\| > B_n) \leq e^{-\beta^r B_n^r} \leq e^{-4n\delta_n}.$$

LEMMA B.3. *For all large enough n, suppose for a large enough constant $C > 0$:*
(i) for the uniform prior and truncated normal prior,

$$n\delta_n \geq q_n \log\left[\frac{B_n\sqrt{q_n}}{C\delta_n}(\sup_{g\in\mathcal{F}_n, w\in[0,1]^d}|E(\rho(Z,g)|W=w)| + 1)\right],$$

(ii) for the truncated normal prior,

$$n\delta_n \geq q_n \log\left[\frac{\sqrt{q_n}}{C\delta_n}(\sup_{g\in\mathcal{F}_n, w\in[0,1]^d}|E(\rho(Z,g)|W=w)| + 1)\right],$$

(iii) for the thin-tail prior,

$$n\delta_n \geq q_n \log\left[\frac{2}{C\delta_n}(\sup_{g\in\mathcal{F}_n, w\in[0,1]^d}|E(\rho(Z,g)|W=w)| + 1) + 1\right],$$

then

$$\pi(b \in \mathcal{F}_n, G(g_b) < \delta_n) \succ e^{-2n\delta_n}.$$

PROOF. (i) For the uniform prior, $\pi(b_i) = (2B_n)^{-1}I(|b_i| \leq B_n)$. In addition, $|\tilde{b}_i| \leq |b_i^*| + O(1) \leq B_n$ as $B_n \to \infty$. Hence $\pi(\tilde{b}) = (2B_n)^{-q_n}$. Lemma B.1 then immediately implies

$$\pi(g_b \in \mathcal{F}_n, G(g_b) \leq \delta_n) \geq \left(\frac{C_1\delta_n}{B_n\gamma_n\sqrt{q_n}}\right)^{q_n}.$$

The conclusion then follows.

(ii) For the truncated normal prior, $\pi(b_i) = I(|b_i| \leq B_n) f_Z(b_i)/P(|Z| \leq B_n)$, where Z follows $N(0, \sigma^2)$, for some $\sigma^2 > 0$, and $f_Z(.)$ denotes its density function. Since

$$\exp(-\|\tilde{b}\|^2/(2\sigma^2)) \geq \exp(-(\|b^*\|^2 + O(1))^2/(2\sigma^2)),$$

$\pi(\tilde{b}) \geq C^{q_n}$ for some $C > 0$. Hence

$$\pi(g_b \in \mathcal{F}_n, G(g_b) \leq \delta_n) \geq \left(\frac{C_1 \delta_n}{\gamma_n \sqrt{q_n}}\right)^{q_n}.$$

The conclusion then follows.

(iii) Suppose π is the density of the thin-tail prior defined in Section 3.2,

$$\pi(b) = r\|b\|^{r-q_n} \beta^r e^{-\beta^r \|b\|^r}/S_{q_n},$$

where S_{q_n} is the area of the $q_n - 1$ dimensional unit sphere in Euclidean norm. Since $\|\tilde{b}\| \leq \|b^*\| + O(1)$ for all large n, then as $q_n \to \infty$,

$$\pi(\tilde{b}) \geq C_1 C_2^{q_n}/S_{q_n}$$

for $C_1, C_2 > 0$. In addition, $\mu(\|b - b^*\| < C\delta_n/\gamma_n) = C_{q_n}(C\delta_n/\gamma_n)^{q_n}$, where C_{q_n} is the volume of the q_n dimensional unit ball in Euclidean norm.

Note that $S_{q_n} = q_n C_{q_n}$, a relation between the area of the sphere and the volume of the ball. Then by the proof of Lemma B.1,

$$\begin{aligned}
\pi(\mathcal{F}_n \cap G(g_b) < \delta_n) &\geq \pi(\tilde{b})\mu(\|b - b^*\| < C\delta_n/\gamma_n) \\
&\geq \frac{C_1}{q_n}\left(\frac{C_2\delta_n}{\gamma_n}\right)^{q_n} \\
&\geq \left(\frac{C_2\delta_n}{2\gamma_n}\right)^{q_n}.
\end{aligned}$$

The conclusion then follows.

B.2. Proofs of Theorems 3.1, 3.2, 3.3 and 3.4.

PROOF. Note that $\inf_{g \in \mathcal{H}} G(g) = 0$.

(i) When the truncated priors are used, for any sequence $c_n \succ q_n/n$, let $\delta_n^* = \frac{q_n}{n}\log(c_n n/q_n)$, and $\mu_n = \frac{q_n}{n}\log(\gamma_n n)$. If $\delta_n \succ \mu_n$, then for any $C > 1$, $\delta_n \succ \mu_n + \frac{q_n}{n}\log C$. For the uniform prior, let $c_n = B_n q_n \gamma_n / C$; for the truncated normal prior, let $c_n = C q_n \gamma_n / C$. Then $\delta_n \succ \delta_n^*$.

Let $x = n\delta_n^*/q_n$, $y = c_n n/q_n$, then $x = \log y$. Note that if for all large n, $y > e$. It follows that $x > \log(y/x)$, which is, $n\delta_n^* > q_n \log(c_n/\delta_n^*)$. Hence for all large n, when $\delta_n \succ \delta_n^*$ then $n\delta_n > q_n \log(c_n/\delta_n)$. Hence conditions (i)(ii) in Lemma B.3 are satisfied. Note that $q_n = o(n)$ and $B_n = o(n)$. Theorem 3.1 is then a straightforward application of Theorems 2.1 and Lemmas B.2, B.3.

In addition, under (3.10) of the main paper, there exists $\delta_n = O(1)$ such that

$$\max\{G(g_{q_n}^*), \lambda_n, \mu_n\} \prec \delta_n \prec \inf_{g \in \mathcal{H}_n, g \notin \Theta_I^\epsilon} G(g).$$

Hence $P(G(g_b) < \delta_n | X^n) = o_p(1)$ by Theorem 3.1, and δ_n also satisfies condition (iv) in Theorem 2.2, which then implies the posterior consistency. This proves Theorem 3.3.

(ii) Consider $\delta_n^* = O(1)$ that satisfies

$$\max\{G(g_{q_n}^*), \lambda(B_n^*), \gamma(B_n^*) e^{-n\lambda(B_n^*)/q_n}\} \prec \delta_n^* \prec B_n^{*r}/n, \tag{B.1}$$

then $\forall C > 0$, $q_n \log(C\gamma(B_n^*)/\delta_n^*) \leq n\lambda_n$ for all large n. Hence $q_n \log(C\gamma(B_n^*)/\delta_n^*) = o(n\delta_n^*)$. Hence by applying Theorems 2.1 and Lemma B.3, we have $P(G(g_b) > \delta_n^* | X^n) = o_p(1)$. Therefore $P(G(g_b) > \delta_n | X^n) = o_p(1)$ is proved if δ_n satisfies

$$\max\{G(g_{q_n}^*), \lambda(B_n^*), \gamma(B_n^*) e^{-n\lambda(B_n^*)/q_n}\} \prec \delta_n \prec B_n^{*r}/n.$$

This is because we can set $\delta_n^* = \delta_n$, and have just shown $P(G(g_b) > \delta_n^* | X^n) = o_p(1)$.

If $\delta_n \succ B_n^{*r}/n$, as there always exists δ_n^* such that (B.1) is satisfied. Thus $P(G > \delta_n | X^n) \leq P(G > \delta_n^* | X^n) = o_p(1)$. This proves Theorem 3.2.

Now we prove Theorem 3.4: the theorem's conditions imply that there exists $\delta_n^* = o(1)$ such that

$$\max\{G(g_{q_n}^*), \lambda(B_n^*), \gamma(B_n^*) e^{-n\lambda(B_n^*)/q_n}\} \prec \delta_n^* \prec \min\{\left(\frac{B_n^{*r}}{n}\right), \inf_{g \in \mathcal{H}_n, g \notin \Theta_I^\epsilon} G(g)\}.$$

Hence by Theorem 3.2, $P(G(g_b) > \delta_n^* | X^n) = o_p(1)$. The theorem is proved since $\delta_n^* = o(\inf_{g \in \mathcal{H}_n, g \notin \Theta_I^\epsilon} G(g))$. The result follows also from Theorem 2.2.

Q.E.D.

B.3. Proof of Corollary 3.1. By uniform continuity, for any $\delta > 0$, exists $\epsilon > 0$, when $\|g_1 - g_2\|_s < \epsilon$, $|h(g_1) - h(g_2)| < \delta$. By Theorems 3.3 and 3.4,

$$P(\exists g^* \in \Theta_I, \|g_b - g^*\|_s < \epsilon | X^n) \to^p 1.$$

Hence $P(\exists g^* \in \Theta_I, |h(g_b) - h(g^*)| < \delta | X^n) \to^p 1$, which implies

$$P(|h(g_b) - h(g_0)| < \delta | X^n) \to^p 1.$$

APPENDIX C: PROOFS FOR SECTION 4

C.1. Uniform convergence of the risk functional. Define

$$G_{k_n}(g) \equiv Em_n(g,Z)^T V^{-1} Em_n(g,Z) \tag{C.1}$$

where $V = \text{diag}\{P(W \in R_1^n), ..., P(W \in R_{k_n^d}^n)\}$.

Han and Phillips (2006) applied the standard results in Newey and McFadden (1994) to establish the uniform convergence of $|\bar{G}(g) - G_{k_n}|$ over a compact set, where they verified the equicontinuity conditions and the point-wise convergence of $|\bar{G}(g) - G_{k_n}|$ respectively. In this subsection, we establish the uniform convergence on the growing compact set $\mathcal{F}_n = \{\sum_{i=1}^{q_n} b_i\phi_i(x) \in \mathcal{H}_n : \max_{i\le q_n}|b_i| \le B_n\}$, by bounding $|\bar{G} - G_{k_n}|$ and $|G_{k_n}(g) - G(g)|$ directly.

LEMMA C.1. *As $k, q \to \infty$,*
(i) $\max_{j\le k_n^d} |\frac{1}{n}\sum_{i=1}^n I_{(W_i\in R_j^n)} - P(W \in R_j^n)| = O_p(\sqrt{\log k_n} k_n^{-d/2} n^{-1/2})$
(ii) $\max_{j\le k_n^d} |(\frac{1}{n}\sum_{i=1}^n I_{(W_i\in R_j^n)})^{-1} - P(W \in R_j^n)^{-1}| = O_p(\sqrt{\log k_n} k_n^{3d/2} n^{-1/2})$
(iii) $\max_{j\le k_n^d} |\frac{1}{n}\sum_{i=1}^n I_{(W_i\in R_j^n)}Y_i - EI_{(W\in R_j^n)}Y| = O_p(n^{-1/2})$
(iv) $\max_{j\le k_n^d, l\le q_n} |\frac{1}{n}\sum_{i=1}^n I_{(W_i\in R_j^n)}\phi_l(X_i) - EI_{(W\in R_j^n)}\phi_l(X)| = O_p(q_n^{1/2} n^{-1/2})$.

PROOF. (i) By Bernstein's inequality, for any $j \le k_n^d$, and $t > 0$,

$$P\Big(\Big|\frac{1}{n}\sum_{i=1}^n I_{(W_i\in R_j^n)} - P(W \in R_j^n)\Big| > t\Big) \le \exp\left(\frac{-nt^2/2}{P(W \in R_j^n) + t/3}\right)$$

Therefore, by the assumption that $\max_{j\le k_n^d} P(W \in R_j^n) = O(k_n^{-d})$,

$$P\Big(\max_{j\le k_n^d}\Big|\frac{1}{n}\sum_{i=1}^n I_{(W_i\in R_j^n)} - P(W \in R_j^n)\Big| \ge t\Big) \le k_n^d \exp\left(\frac{-nt^2/2}{P(W \in R_j^n) + t/3}\right),$$

which implies the result. Here we used the fact that $k_n^d = o(n^{1/2})$, which is implied by the assumption $q_n^2 B_n^2 k_n^{3d/2} = o(n)$ in Theorem 4.1.

(ii) Write $v_j = P(W \in R_j^n)$, $\hat{v}_j = \frac{1}{n}\sum_{i=1}^n I_{(W_i\in R_j^n)}$. By the assumption that $k_n^{-d} = O(\min_{j\le k_n^d} P(W \in R_j^n))$, and that $k_n^d = o(n^{1/2})$, we have

$$\begin{aligned}
\max_{j\le k_n^d} |v_j^{-1} - \hat{v}_j^{-1}| &\le \frac{1}{\min_j \hat{v}_j v_j}\max_j |\hat{v}_j - v_j| \\
&\le \frac{1}{\min_j v_j(\min_j v_j - \max_j|\hat{v}_j - v_j|)}\max_j |\hat{v}_j - v_j| \\
&= O_p(\sqrt{\log k_n} k_n^{3d/2} n^{-1/2}).
\end{aligned}$$

(iii) By Chebyshev inequality, for any $j \leq k_n^d$, and $t > 0$,

$$P(|\frac{1}{n}\sum_{i=1}^n I_{(W_i\in R_j^n)}Y_i - EI_{(W\in R_j^n)}Y| > t) \leq \frac{\text{var}(I_{(W\in R_j^n)}Y)}{nt^2} \leq \frac{EI_{(W\in R_j^n)}Y^2}{nt^2}.$$

Hence

$$\begin{aligned}&P(\max_{j\leq k_n^d}|\frac{1}{n}\sum_{i=1}^n I_{(W_i\in R_j^n)}Y_i - EI_{(W\in R_j^n)}Y| > t) \leq k_n^d \max_{j\leq k_n^d}\frac{EI_{(W\in R_j^n)}Y^2}{nt^2}\\ &\leq \frac{\sup_w E(Y^2|W=w)\max_j P(W\in R_j^n)k_n^d}{nt^2}.\end{aligned}$$

We obtain $\max_{j\leq k_n^d}|\frac{1}{n}\sum_{i=1}^n I_{(W_i\in R_j^n)}Y_i - EI_{(W\in R_j^n)}Y| = O_p(n^{-1/2})$.

Finally, part (iv) follows from similar arguments as those in part (iii), with the assumption that $\max_i \sup_w E(\phi_i^2|W=w) < \infty$, and the application of Chebyshev and Bonferroni inequalities. Q.E.D.

LEMMA C.2. *(i)* $\sup_{\mathcal{F}_n, j\leq k_n^d}|\frac{1}{n}\sum_{i=1}^n m_{nj}(g,X_i) - Em_{nj}(g,X)| = O_p(q_n^{3/2}B_n n^{-1/2})$.
(ii) $\sup_{\mathcal{F}_n, j\leq k_n^d}|(\frac{1}{n}\sum_{i=1}^n m_{nj}(g,X_i))^2 - (Em_{nj}(g,X))^2| = O_p(q_n^2B_n^2k_n^{-d/2}n^{-1/2} + q_n^3B_n^2n^{-1})$.

PROOF. (i) For any $g = \sum_{l=1}^{q_n} b_l\phi_l \in \mathcal{F}_n$, by Lemma C.1,

$$\begin{aligned}&\max_{j\leq k_n^d}\sup_{\mathcal{F}_n}|\frac{1}{n}\sum_{i=1}^n m_{nj}(g,X_i) - Em_{nj}(g,X)| \leq \max_{j\leq k_n^d}|\frac{1}{n}\sum_{i=1}^n Y_iI_{(W_i\in R_j^n)} - E(YI_{(W\in R_j^n)})|\\ &+ \sup_{|b_l|\leq B_n}\sum_{l=1}^{q_n}|b_l| \max_{j\leq k_n^d, l\leq q_n}|\frac{1}{n}\sum_{i=1}^n \phi_l(X_i)I_{(W_i\in R_j^n)} - E(\phi_l(X)I_{(W\in R_j^n)})|\\ =\ &O_p(q_n^{3/2}B_nn^{-1/2}).\end{aligned}$$

(ii) We have, by Cauchy-Schwarz inequality,

$$\begin{aligned}&\max_j\sup_{\mathcal{F}_n}|Em_{nj}(g,X)| \leq \sup_{g\in\mathcal{F}_n}\max_{j\leq k_n^d}|E(Y-g(X))I_{(W\in R_j^n)}|\\ &\leq \max_{j\leq k_n^d}\sqrt{E(Y^2)P(W\in R_j^n)} + \max_{j\leq k_n^d}\sup_{\mathcal{F}_n}\sqrt{E[g(X)^2]P(W\in R_j^n)}\\ &= O_p(q_n^{1/2}B_nk_n^{-d/2}).\end{aligned}$$

Therefore, by part (i),

$$\max_j\sup_{g\in\mathcal{F}_n}|\frac{1}{n}\sum_{i=1}^n m_{nj}(g,X_i)| \quad\leq\quad \max_j\sup_{g\in\mathcal{F}_n}|\frac{1}{n}\sum_{i=1}^n m_{nj}(g,X_i) - Em_{nj}(g,X)|$$

$$+\max_j \sup_{g\in\mathcal{F}_n} |Em_{nj}(g,X)|$$
$$= O_p(q_n^{1/2}B_n k_n^{-d/2} + q_n^{3/2}B_n n^{-1/2}).$$

The desired result then follow from the fact that

$$|(\frac{1}{n}\sum_{i=1}^n m_{nj}(g,X_i))^2 - (Em_{nj}(g,X))^2|$$
$$\leq |\frac{1}{n}\sum_{i=1}^n m_{nj}(g,X_i) - Em_{nj}(g,X)| \times |\frac{1}{n}\sum_{i=1}^n m_{nj}(g,X_i) + Em_{nj}(g,X)|.$$

Q.E.D.

LEMMA C.3. *Under Assumptions 3.1, 4.1, 4.2,*
(i)
$$\sup_{g\in\mathcal{F}_n} |\bar{G}(g) - G_{k_n}(g)| = O_p(q_n^2 B_n^2 k_n^{3d/2} n^{-1/2}).$$
(ii)
$$\sup_{g\in\mathcal{F}_n} |G(g) - G_{k_n}(g)| = O(q_n^2 B_n^2 k_n^{-1}).$$

PROOF. (i) For any $g \in L^2(X)$,

$$\bar{G}(g) = \bar{m}_n(g)^T \hat{V}^{-1} \bar{m}_n(g) = \sum_{j=1}^{k_n^d} (\frac{1}{n}\sum_{i=1}^n m_{nj}(g,X_i))^2 \hat{v}_j^{-1},$$
$$G_{k_n}(g) = Em_n(X,g)^T V^{-1} Em_n(X,g) = \sum_{j=1}^{k_n^d} [Em_{nj}(g,X)]^2 v_j^{-1}.$$

Hence $\bar{G} - G_{k_n} = P_1 + P_2$, where

$$P_1 = \sum_{j=1}^{k_n^d} [(\frac{1}{n}\sum_{i=1}^n m_{nj}(g,X_i))^2 - (Em_{nj}(g,X))^2]\hat{v}_j^{-1},$$
$$P_2 = \sum_{j=1}^{k_n^d} [Em_{nj}(g,X)]^2 (\hat{v}_j^{-1} - v_j^{-1}).$$

By Lemmas C.1 C.2, and $q_n k_n^{d/2} = o(\sqrt{n})$, which is implied by the assumption $q_n^2 B_n^2 k_n^{3d/2} = o(n)$ in Theorem 4.1,

$$\sup_{g\in\mathcal{F}_n} P_1 \leq \max_j \sup_{g\in\mathcal{F}_n} |(\frac{1}{n}\sum_{i=1}^n m_{nj}(g,X_i))^2 - (Em_{nj}(g,X))^2| \max_i (|\hat{v}_i^{-1} - v_i^{-1}| + |v_i^{-1}|) k_n^d,$$

$$= O_p(q_n^2 B_n^2 k_n^{3d/2} n^{-1/2}).$$

By Lemma C.1(ii),

$$\begin{aligned}
\sup_{g\in\mathcal{F}_n} P_2 &\le \max_j \sup_{\mathcal{F}_n} (Em_{nj}(g,X))^2 \max_i |\hat{v}_i^{-1} - v_i^{-1}| k_n^d \\
&= O_p(\sqrt{\log k_n} k_n^{3d/2} n^{-1/2}) k_n^d \max_{j\le k_n^d} \sup_{\max|b_l|\le B_n} |EYI_{(W\in R_j^n)} - \sum_{l=1}^{q_n} b_l E\phi_l(X) I_{(W\in R_j^n)}|^2 \\
&\le O_p(\sqrt{\log k_n} k_n^{5d/2} n^{-1/2}) \\
&\quad \times (\sup_{j\le k_n^d, w} |E(Y|W=w)| P(W\in R_j^n) + B_n q_n \max_{l,j} |E\phi_l(X) I_{(W\in R_j^n)}|)^2.
\end{aligned}$$

In addition,

$$\max_{l\le q_n, j\le k_n^d} |E\phi_l(X) I_{(W\in R_j^n)}| \le \max_{l\le q_n} \sup_{w\in[0,1]^d} |E(\phi_l(X)|W=w)| \max_{j\le k_n^d} P(W\in R_j^n).$$

By Assumption 4.2(i) and Cauchy-Schwarz inequality, $\sup_{w\in[0,1]^d}|E(Y|W = w) < \infty$, and $\max_{l\le q_n}\sup_{w\in[0,1]^d}|E(\phi_l(X)|W = w)| < \infty$. Hence it follows from Assumption 4.1(ii) that

$$\begin{aligned}
\sup_{g\in\mathcal{F}_n} P_2 &= O_p(\sqrt{\log k_n} k_n^{5d/2} n^{-1/2}) \times O_p((k_n^{-d} + B_n q_n k_n^{-d})^2) \\
&= O_p(\sqrt{\log k_n} k_n^{d/2} n^{-1/2} B_n^2 q_n^2).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\sup_{g\in\mathcal{F}_n} |\bar{G}(g) - G_{k_n}(g)| &\le \sup_{g\in\mathcal{F}_n} P_1 + \sup_{g\in\mathcal{F}_n} P_2 \\
&= O_p(q_n^2 B_n^2 k_n^{3d/2} n^{-1/2}).
\end{aligned}$$

(ii) By definition, $G_{k_n}(g) = Em_n(g,X)^T V^{-1} Em_n(g,X)$. Since V is diagonal, it is straightforward to show that

$$G_{k_n}(g) = \sum_{j=1}^{k_n^d} \frac{[E(\rho(Z,g)I_{(W\in R_j^n)})]^2}{P(W\in R_j^n)} = \sum_{j=1}^{k_n^d} \int_{R_j^n} \frac{[E(\rho(Z,g)I_{(W\in R_j^n)})]^2}{P(W\in R_j^n)^2} dF_W(w)$$

Also,

$$G(g) = \int_{[0,1]^d} [E(\rho(Z,g)|W=w)]^2 dF_W(w) = \sum_{j=1}^{k_n^d} \int_{R_j^n} [E(\rho(Z,g)|W=w)]^2 dF_W(w)$$

It follows that

$$\begin{aligned}
|G_{k_n}(g) - G(g)| &\leq \sum_{j=1}^{k_n^d} \int_{R_j^n} \left| \frac{[E(\rho(Z,g)I_{(W\in R_j^n)})]^2}{P(W\in R_j^n)^2} - [E(\rho(Z,g)|W=w)]^2 \right| dF_W(w) \\
&\leq \sup_{1\leq j\leq k_n^d} \sup_{w\in R_j^n} \left| \frac{[E(\rho(Z,g)I_{(W\in R_j^n)})]^2}{P(W\in R_j^n)^2} - [E(\rho(Z,g)|W=w)]^2 \right| \\
&\leq \sup_{1\leq j\leq k_n^d} \sup_{w\in R_j^n} \left| \frac{E(\rho(Z,g)I_{(W\in R_j^n)})}{P(W\in R_j^n)} + E(\rho(Z,g)|W=w) \right| \\
&\times \sup_{1\leq j\leq k_n^d} \sup_{w\in R_j^n} \left| \frac{E(\rho(Z,g)I_{(W\in R_j^n)})}{P(W\in R_j^n)} - E(\rho(Z,g)|W=w) \right| \\
&= A(g)\times B(g), \text{ say.}
\end{aligned}$$

$$\begin{aligned}
A(g) &\leq \sup_{1\leq j\leq k_n^d} \frac{|E(\rho(Z,g)I_{(W\in R_j^n)})|}{P(W\in R_j^n)} + \sup_{w\in[0,1]^d} |E(\rho(Z,g)|W=w)| \\
&\leq \sup_{1\leq j\leq k_n^d} \frac{|\int_{R_j^n} E(\rho(Z,g)|W=w)dF_W(w)|}{P(W\in R_j^n)} + \sup_{w\in[0,1]^d} |E(\rho(Z,g)|W=w)| \\
&\leq 2\sup_{w\in[0,1]^d} |E(\rho(Z,g)|W=w)|,
\end{aligned}$$

which implies

$$\sup_{g\in\mathcal{F}_n} A(g) = O(q_n B_n).$$

$$\begin{aligned}
B(g) &\leq \sup_{1\leq j\leq k_n^d} \sup_{w\in R_j^n} \left| \frac{E[\rho(Z,g)I_{(W\in R_j^n)}]}{P(W\in R_j^n)} - E(\rho(Z,g)|W=w) \right| \\
&= \sup_{1\leq j\leq k_n^d} \sup_{w\in R_j^n} P(W\in R_j^n)^{-1} \left| \int_{R_j^n} E(\rho(Z,g)|W=t)dF_W(t) - \right. \\
&\qquad \left. \int_{R_j^n} E(\rho(Z,g)|W=w)dF_W(t) \right| \\
&\leq \sup_{j\leq k_n^d} P(W\in R_j^n)^{-1} \int_{R_j^n} \sup_{w\in R_j^n} |K_g(t) - K_g(w)| dF_W(t) \\
&\leq \sup_{\|w-t\|\leq\sqrt{d}k_n^{-1}} |K_g(t) - K_g(w)|,
\end{aligned}$$

where

$$K_g(w) \equiv E[\rho(Z,g)|W=w] = E[Y - g(X)|W=w],$$

and we used the facts that $E[\rho(Z,g)I_{(W\in R_j^n)}]=\int_{R_j^n}E(\rho(Z,g)|W=t)dF_W(t)$, and $E(\rho(Z,g)|W=w)P(W\in R_j^n)=\int_{R_j^n}E(\rho(Z,g)|W=w)dF_W(t)$.

By Assumption 4.2, $E(Y|W=w)$ is Lipchitz continuous and $\{E(\phi_i(X)|W=w):i\leq q_n\}$ is Lipchitz equicontinuous with respect to w. Therefore,

$$\sup_{g\in\mathcal{F}_n}\sup_{\|w-t\|\leq\sqrt{d}k_n^{-1}}|K_g(t)-K_g(w)|=O(q_nB_nk_n^{-1}),$$

which yields the result.

Proof of Theorem 4.1

PROOF. The theorem follows from Lemma C.3. Q.E.D.

C.2. Proof of Theorem 4.2-4.3.

LEMMA C.4. *(i) For any $g_0\in\Theta_I$, let $g_{q_n}^*$ be its sieve approximation, then $G(g_{q_n}^*)\leq\|g_{q_n}^*-g_0\|_s^2$.*
(ii) $\gamma_n=O(q_nB_n)$.

PROOF. (i) The proof is simply a straightforward calculation:

$$\begin{aligned}G(g_{q_n}^*) &= E_W\{[E(g_{q_n}^*-g_0|W)]^2\}\leq E_W\{E[(g_{q_n}^*-g_0)^2]|W\}\\ &= E[(g_{q_n}^*-g_0)^2]=\|g_{q_n}^*-g_0\|_s^2.\end{aligned}$$

(ii) By definition,

$$\begin{aligned}\gamma_n &= \sup_{g\in\mathcal{F}_n,w\in[0,1]^d}|E(\rho(Z,g)|W=w)|+1\\ &\leq \sqrt{\sup_w E(Y^2|W=w)}+1+\sup_{|b_i|\leq B_n}\sum_{i\leq q_n}|b_i|\sqrt{\sup_w E(\phi_i^2|W=w)}\\ &= O(q_nB_n).\end{aligned}$$

Proof of Theorem 4.2

PROOF. Write $\beta_n=k_n^{3d/2}/\sqrt{n}+k_n^{-1}$. Then $\beta_n\geq n^{-1/(3d+2)}$.
(i) For the truncated prior, note that for all large n, $\log B_n^2q_n\leq\log n$, hence $q_n^2B_n^2\beta_n=o(\delta_n)$ implies $\frac{q_n}{n}\log n=o(\delta_n)$, which then implies $\frac{q_n}{n}\log(B_n\gamma_nn)=o(\delta_n)$. The theorem is thus a straightforward application of Lemma C.4 and Theorem 3.1.
(ii) The assumption $n^{2/(r-2)}q_n^{2r/(r-2)}\beta_n^{r/(r-2)}=o(\delta_n)$ implies $(n\delta_n)^{1/r}=o(\sqrt{\frac{\delta_n}{q_n^2\beta_n}})$ and $q_n^2\beta_n=o(\delta_n)$. Hence there exists $B_n^*\to\infty$ such that

$$(n\delta_n)^{1/r}\prec B_n^*\prec\sqrt{\frac{\delta_n}{q_n^2\beta_n}}.$$

Hence

$$\delta_n = o(\frac{B_n^{*r}}{n}), \tag{C.2}$$

$$B_n^{*2} q_n^2 \beta_n = o(\delta_n). \tag{C.3}$$

Let $\gamma_n^* = q_n B_n^*$ and $\lambda_n(B_n^*) = B_n^{*2} q_n^2 \beta_n$. Since $B_n q_n \beta_n \geq n^{-1/2}$, and $\sqrt{n} e^{-\sqrt{n}} = o(1)$ we have $\gamma_n^* e^{-n\lambda_n(B_n^*)/q_n} = o(B_n^{*2} q_n^2 \beta_n)$. Hence (C.2) and (C.3) verify the conditions in Theorem 3.2, and therefore imply the desired result. Q.E.D.

For any $g_0 = \sum_{j=1}^{\infty} b_j^* \phi_j \in \Theta_I$, let $g_{q_n}^* = \sum_{j \leq q_n} b_j^* \phi_j$ denote the projection of g_0 onto the sieve space $\mathcal{H}_n$. Then $\forall g_b \in \mathcal{H}_n$, $\sum_{j \leq q_n} |\langle g_b - g_0, \phi_j \rangle_X|^2 = \|g_b - g_{q_n}^*\|_s^2$ as $\{\phi_1, ..., \phi_q\}$ are orthonormal.

LEMMA C.5. *Under Assumption 4.3 and 4.5(i), for any $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that when $n > N$,*

$$\inf_{g \in \mathcal{H}_n / \Theta_I^{\epsilon}} G(g) \geq \varphi(\eta_{q_n}^2) C \epsilon^2,$$

where $C > 0$ is a constant independent of n and ϵ.

PROOF.

$$\begin{aligned}
\inf_{g \in \mathcal{H}_n, g \notin \Theta_I^{\epsilon}} G(g) &= \inf_{g \in \mathcal{H}_n, g \notin \Theta_I^{\epsilon}} \|g - g_0\|_w^2 \geq \inf_{g \in \mathcal{H}_n, g \notin \Theta_I^{\epsilon}} C_1 \sum_{j=1}^{q_n} \varphi(\eta_j^2) |\langle g - g_0, \phi_j \rangle_X|^2 \\
&\geq \varphi(\eta_{q_n}^2) \inf_{g \in \mathcal{H}_n, g \notin \Theta_I^{\epsilon}} C_1 \sum_{j=1}^{q_n} |\langle g - g_0, \phi_j \rangle_X|^2 \\
&= \varphi(\eta_{q_n}^2) \inf_{g \in \mathcal{H}_n, g \notin \Theta_I^{\epsilon}} C_1 \|g - g_{q_n}^*\|_s^2 \\
&\geq \varphi(\eta_{q_n}^2) \inf_{g \in \mathcal{H}_n, \|g - g_{q_n}^*\|_s \geq \epsilon - \|g_{q_n}^* - g_0\|_s} C_1 \|g - g_{q_n}^*\|_s^2 \\
&\geq \varphi(\eta_{q_n}^2) C_1 [\epsilon - \|g_{q_n}^* - g_0\|_s]^2. \qquad \text{(C.4)}
\end{aligned}$$

Note that by Assumption 4.3, there exists $N \in \mathbb{N}$ such that when $n > N$, $\|g_{q_n}^* - g_0\|_s < \epsilon/2$, which guarantees that (C.4) holds, and that the right-hand-side of (C.4) is no less than $\varphi(\eta_{q_n}^2) C_1 \epsilon^2 / 4$.

Q.E.D.

LEMMA C.6. *Under Assumptions 4.3 and 4.5, for any $\epsilon > 0$, $G(g_{q_n}^*) = o(\inf_{g \in \mathcal{H}_n / \Theta_I^{\epsilon}} G(g))$.*

PROOF. Under Assumption 4.5(ii),

$$\begin{aligned} G(g_{q_n}^*) = \|g_{q_n}^* - g_0\|_w^2 &\leq C_2 \sum_{j \leq q_n} \varphi(\eta_j^2)(b_j^* - b_j^*)^2 + C_2 \sum_{j > q_n} \varphi(\eta_j^2) b_j^{*2} \\ &\leq C_2 \varphi(\eta_{q_n}^2) \|g_{q_n}^* - g_0\|_s^2 = O(\varphi(\eta_{q_n}^2)\eta_{q_n}^2). \end{aligned}$$

But for any $\epsilon > 0$, $\inf_{g \in \mathcal{H}_n / \Theta_I^\epsilon} G(g) \geq C\varphi(\eta_{q_n}^2)$, as proved in Lemma C.5.
Q.E.D.

Proof of Theorem 4.3

PROOF. The proof proceeds by verifying conditions in Theorem 3.3 and 3.4, which is similar to the proof of Theorem 4.2, and is completed by Lemma C.5 and C.6.
Q.E.D.

C.3. Proof of Posterior Consistency Using Gaussian Prior. It can be easily shown that the posterior has a form

$$b|X^n \sim N((\hat{\Xi}^T \hat{V}^{-1} \hat{\Xi} + \frac{1}{n\sigma^2} I_{q_n})^{-1} \hat{\Xi}^T \hat{V}^{-1} \hat{\xi}, (n\hat{\Xi}^T \hat{V}^{-1} \hat{\Xi} + \frac{1}{\sigma^2} I_{q_n})^{-1}),$$

where

$$\hat{\xi} = [n^{-1} \sum_{i=1}^n I_{(W_i \in R_j^n)} Y_i]_{j \leq k_n^d}, \qquad \hat{V} = \text{diag}[n^{-1} \sum_{i=1}^n I_{(W_i \in R_j^n)}]_{j \leq k_n^d},$$
$$\hat{\Xi} = [n^{-1} \sum_{i=1}^n I_{(W_i \in R_j^n)} \phi_l(X_i)]_{j \leq k_n^d; l \leq q_n},$$

and I_{q_n} denotes the $q_n \times q_n$ identity matrix.

Define the population versions as
$\xi = [EI(W \in R_j)Y]_{j \leq k_n^d}$,
$V = \text{diag}[EI(W \in R_j)]_{j \leq k_n^d} = \text{diag}(v_1, ..., v_{k_n^d})$,
$\Xi = [EI(W \in R_j)\phi_l(X)]_{j \leq k_n^d; l \leq q_n}$.

We proceed by checking the conditions in Theorems 2.1 and 2.2 as before. Note that, for Condition (i) in Theorem 2.1, we check $EP(g_b \in \mathcal{F}_n^c | X^n) = o(1)$ instead of the tail condition on the prior.

LEMMA C.7. *If* $\max\{q_n \log q_n, q_n \log \delta_n^{-1}, nG(g_{q_n}^*)\} = o(n\delta_n)$, $q_n \log B_n < \frac{1}{2} n\delta_n$, *and* $\delta_n = o(1)$, *then*

$$\pi(G(g_b) < \delta_n, g_b \in \mathcal{F}_n) \succ e^{-2n\delta_n}.$$

PROOF. By Lemma B.1,

$$\pi(G(g_b) < \delta_n, g_b \in \mathcal{F}_n) \geq \pi(\tilde{b}) \left(\frac{C\delta_n}{q_n^{3/2} B_n} \right)^{q_n},$$

where $\tilde{b}$ lies in the ball $\{b \in \mathbb{R}^{q_n} : \|b - b^*\| \leq C\delta_n/(q_n B_n)\}$ for some b^* such that $g^*_{q_n} = \sum_{i \leq q_n} b^*_i \phi_i$ is the sieve approximation of some $g_0 \in \Theta_I$. Hence

$$\begin{aligned}\pi(G(g_b) < \delta_n, g_b \in \mathcal{F}_n) &\geq \left(\frac{1}{2\pi\sigma^2}\right)^{q_n/2} e^{-\frac{1}{2}(\|g_0\|_s^2+1)} \left(\frac{C\delta_n}{q_n^{3/2} B_n}\right)^{q_n} \\ &\geq C_1 \left(\frac{C_2\delta_n}{q_n^{3/2} B_n}\right)^{q_n} \\ &\succ e^{-n\delta_n},\end{aligned}$$

given the lemma conditions, where in the second inequality above, $C_2 = C(2\pi\sigma^2)^{-1/2}$, and $C_1 = \exp(-\frac{1}{2}(\|g_0\|_s^2 + 1))$.

LEMMA C.8. *If* $q_n^{3/2} k_n^d/\sqrt{n} + q_n/k = o(\varphi(\eta_{q_n}^2))$ *and* $\sqrt{q_n} B_n^{-1} = o(\varphi(\eta_{q_n}^2))$, *then for* $\mathcal{F}_n^c = \{\sum_{i=1}^{q_n} b_i\phi_i : \max_{i \leq q_n} |b_i| > B_n\}$,

$$P(g_b \in \mathcal{F}_n^c | X^n) = o_p(1).$$

PROOF. We use the following inequality:

$$P(g_b \in \mathcal{F}_n^c | X^n) \leq P(\|b\|^2 \geq B_n^2 | X^n) \leq \frac{E(\|b\|^2 | X^n)}{B_n^2}.$$

We bound $E(\|b\|^2|X^n) \leq 2E[\|b - E(b|X^n)\|^2|X^n] + 2\|E(b|X^n)\|^2$ as following: Denote by $\{\lambda_i(A)\}_{i=1}^{q_n}$ as the eigenvalues of a $q_n \times q_n$ square matrix A, and by $\lambda_{\max}(A)$, $\lambda_{\min}(A)$ as the maximum and minimum eigenvalues of A.

$$\begin{aligned}&E[\|b - E(b|X^n)\|^2|X^n] = \mathrm{tr}(\mathrm{cov}(b|X^n)) \\ &= \sum_{i=1}^{q_n} \lambda_i^{-1}(n\hat{\Xi}^T\hat{V}^{-1}\hat{\Xi} + \frac{1}{\sigma^2} I_{q_n}) \leq \sigma^2 q_n \\ &= O_p(q_n).\end{aligned}$$

For the second term, it is bounded by:

$$\begin{aligned}\|E(b|X^n)\|^2 &\leq \lambda_{\max}^2((\hat{\Xi}^T\hat{V}^{-1}\hat{\Xi} + (n\sigma^2)^{-1} I_{q_n})^{-1})\|\hat{\Xi}^T\hat{V}^{-1}\hat{\xi}\|^2 \\ &\leq \lambda_{\max}^2((\hat{\Xi}^T\hat{V}^{-1}\hat{\Xi})^{-1})\|\hat{\Xi}^T\hat{V}^{-1}\hat{\xi}\|^2 \\ &= \lambda_{\min}^{-2}(\hat{\Xi}^T\hat{V}^{-1}\hat{\Xi})\|\hat{\Xi}^T\hat{V}^{-1}\hat{\xi}\|^2\end{aligned}$$

$$\leq \quad \lambda_{\min}^{-2}(\hat{\Xi}^T\hat{V}^{-1}\hat{\Xi})\|\hat{\Xi}\|_F^2\lambda_{\min}^{-2}(\hat{V})\|\hat{\xi}\|^2,$$

where $\|\hat{\Xi}\|_F^2 = \mathrm{tr}(\hat{\Xi}\hat{\Xi}^T)$. We bound all the terms on the right hand side of the last inequality separately. By Lemma C.1,

$$\begin{aligned}
\|\hat{\Xi}\|_F^2 &\leq 2\|\hat{\Xi}-\Xi\|^2 + 2\|\Xi\|^2 = q_n k_n^d (q_n/n + \max_{j,l} |EI_{(W\in R_j^n)}\phi_l(X)|^2) \\
&\leq q_n k_n^d (q_n/n + \max_{j,l} [E(I_{(W\in R_j^n)} E(|\phi_l(X)||W))]^2) \\
&\leq q_n k_n^d (q_n/n + \max_{j,l} [E(I_{(W\in R_j^n)} \sqrt{E(\phi_l(X)^2|W)})]^2) \\
&\leq q_n k_n^d (q_n/n + \sup_{w,l} E(\phi_l(X)^2|W=w) \max_j [EI_{(W\in R_j^n)}]^2) \\
&\leq q_n k_n^d (q_n/n + O(k_n^{-2d})) = O(q_n/k_n^d). \qquad \text{(C.5)}
\end{aligned}$$

Since $\max_j |\hat{v}_j - v_j| = o(\min_j v_j)$, we have

$$\lambda_{\min}^{-2}(\hat{V}) = O_p(\lambda_{\min}^{-2}(V)) = O_p((\min_j EI_{(W\in R_j^n)})^{-2}) = O_p(k_n^{2d}). \qquad \text{(C.6)}$$

Again by Lemma C.1,

$$\begin{aligned}
\|\hat{\xi}\|^2 &\leq 2\|\hat{\xi}-\xi\|^2 + 2\|\xi\|^2 = O_p(k_n^d/n) + k_n^d \max_j |EI_{(W\in R_j^n)} Y|^2 \\
&\leq O_p(k_n^d/n + k_n^{-d}) = O_p(k_n^{-d}). \qquad \text{(C.7)}
\end{aligned}$$

$$\|\hat{\Xi}^T\hat{V}^{-1}\hat{\Xi} - \Xi^T V^{-1}\Xi\|_F = O_p(q_n^{3/2}k_n^d n^{-1/2}). \qquad \text{(C.8)}$$

To bound $\lambda_{\min}^{-2}(\hat{\Xi}^T\hat{V}^{-1}\hat{\Xi})$, define

$$A^* = E[E(\phi_i(X)|W)E(\phi_j(X)|W)]_{i,j\leq q_n}, \quad A_0 = \Xi^T V^{-1}\Xi = (a_{ij})_{i,j\leq q_n}.$$

Then $a_{ij} = \sum_{l=1}^{k_n^d} EI(W\in R_l^n)E(\phi_i(X)|W\in R_l^n)E(\phi_j(X)|W\in R_l^n)$.

Using the same technique as in the proof of Lemma C.3(ii), we can show that

$$\begin{aligned}
&\max_{ij} |a_{ij} - A_{ij}^*| \leq \max_{l,i,j} \sup_w \left|\frac{E\phi_i I(W\in R_l^n)E\phi_j I(W\in R_l^n)}{P(W\in R_l^n)^2} - E(\phi_i|w)E(\phi_j|w)\right| \\
\leq\ & \max_{l,i,j} \sup_w \left|\frac{E\phi_i I(W\in R_l^n)}{P(W\in R_l^n)} - E(\phi_i|w)\right| \left(\frac{|E\phi_j I(W\in R_l^n)|}{P(W\in R_l^n)} + |E(\phi_j|w)|\right) \\
\leq\ & C \max_{i,l} P(W\in R_l^n)^{-1} \sup_w \int_{R_l^n} |(E\phi_i|t) - (E\phi_i|w)|dt \\
=\ & O(k_n^{-1}).
\end{aligned}$$

Hence $\|A^* - A_0\|_F = O(q_n/k_n)$, which then implies

$$\|\hat{\Xi}^T\hat{V}^{-1}\hat{\Xi} - A^*\|_F = q_n^{3/2}k_n^d n^{-1/2} + q_n/k_n. \qquad \text{(C.9)}$$

Now for any $b^T=(b_1,...,b_q)\in\mathbb{R}^{q_n}$, let $g(x)=\sum_{l=1}^{q_n}b_l\phi_l(x)$.

$$G(g+g_0)=\|g\|_w^2=E[(Eg(X)|W)^2]=b^TA^*b.$$

Since g_0 is point identified, Assumption 4.5(i) implies

$$\|g\|_w^2\geq C\sum_{j=1}^{q_n}\varphi(\eta_j^2)|\langle g,\phi_j\rangle_X|^2\geq C\varphi(\eta_{q_n}^2)b^Tb. \tag{C.10}$$

Thus $\lambda_{\min}(A^*)\geq C\varphi(\eta_{q_n}^2)$. By (C.9), when $q_n^{3/2}k_n^dn^{-1/2}+q_n/k=o(\varphi(\eta_{q_n}^2))$, we have $\lambda_{\min}(\hat{\Xi}^T\hat{V}^{-1}\hat{\Xi})\geq C'\varphi(\eta_{q_n}^2)$. It then follows that $\|E(b|X^n)\|^2\leq C\frac{q_n}{\varphi(\eta_{q_n}^2)^2}$. Hence

$$E(\|b\|^2|X^n)\leq Cq_n(1+\varphi(\eta_{q_n}^2)^{-2}).$$

Therefore, as long as $\sqrt{q_n}\varphi(\eta_{q_n}^2)^{-1}=o(B_n)$,

$$P(g_b\in\mathcal{F}_n^c|X^n)\leq P(\|b\|^2\geq B_n^2|X^n)\leq\frac{E(\|b\|^2|X^n)}{B_n^2}=o_p(1).$$

Q.E.D.

LEMMA C.9. *If* $\max\{q_n\log n,q_n\log(\varphi(\eta_{q_n}^2)^{-1}),nG(g_{q_n}^*),nq_n^3\beta_n\varphi(\eta_{q_n}^2)^{-2}\}=o_p(n\delta_n)$,*and* $q_n^{3/2}k_n^d/\sqrt{n}+q_n/k=o(\varphi(\eta_{q_n}^2))$, *then*

$$P(G(g_b)<\delta_n|X^n)\to^p 1.$$

PROOF. $\max\{q_n\log n,q_n\log\frac{1}{\varphi(\eta_{q_n}^2)}\}=o(n\delta_n)$ implies $\max\{q_n\log q_n,q_n\log(n/q_n)\}=o(n\delta_n)$, and

$$\log\frac{q_n}{\varphi(\eta_{q_n}^2)^2}=o(\frac{n\delta_n}{2q_n}).$$

Hence $\frac{q_n}{\varphi(\eta_{q_n}^2)^2}=o(\exp(\frac{n\delta_n}{2q_n}))$, which together with $\frac{q_n}{\varphi(\eta_{q_n}^2)^2}=o(\frac{\delta_n}{q_n^2\beta_n})$ implies that there exists $B_n^*\to\infty$ such that $\frac{q_n}{\varphi(\eta_{q_n}^2)^2}\prec B_n^{*2}\prec\min\{\frac{\delta_n}{q_n^2\beta_n},\exp(\frac{n\delta_n}{2q_n})\}$. Therefore

$$q_n\log B_n^*<n\delta_n/2 \tag{C.11}$$

$$q\varphi(\eta_{q_n}^2)^{-2}=o(B_n^{*2}) \tag{C.12}$$

$$B_n^{*2}q_n^2\beta_n=o(\delta_n). \tag{C.13}$$

In addition, let $\delta^*=n^{-1}q_n\log(n/q_n)$, $x=n\delta^*/q_n$ and $y=n/q_n$. Then $x=\log y$ Hence $y>e$ for all large n. It follows that $x>\log(y/x)$, which is $n\delta^*>q_n\log(1/\delta^*)$. Hence when $\delta_n\succ\delta^*$, $n\delta_n\succ n\delta^*>q_n\log(\delta_n^{-1})$, which is

$$q_n\log(\delta_n^{-1})=o(n\delta_n). \tag{C.14}$$

Hence all the conditions in Lemma C.7 and C.8 are satisfied. The lemma then follows from Theorem 2.1. Q.E.D.

Proof of Theorem 4.4

We check the conditions in Lemma C.9. Since $q_n\beta_n^{1/3} = o(\varphi(\eta_{q_n}^2))$, and $\beta_n^{1/3} > n^{-1}\log n$, we have $(q_n \log n)/n = o(\varphi(\eta_{q_n}^2))$. It then implies $q_n \log(\varphi(\eta_{q_n}^2)^{-1}) = o(n\varphi(\eta_{q_n}^2))$. In addition, $q_n^{3/2}k_n^d/\sqrt{n}+q_n/k_n = o(q_n\beta_n^{1/3})$ as $q_n^2\beta_n = o(1)$. Hence

$$\max\{\frac{q_n^3\beta_n}{\varphi(\eta_{q_n}^2)^2}, \frac{q_n}{n}\log n, \frac{q_n}{n}\log\frac{1}{\varphi(\eta_{q_n}^2)}, q_n^{3/2}\frac{k_n^d}{\sqrt{n}} + \frac{q_n}{k_n}\} = O(\varphi(\eta_{q_n}^2)).$$

It then implies the existence of $\delta^* = o(1)$ such that

$$\max\{\frac{q_n^3\beta_n}{\varphi(\eta_{q_n}^2)^2}, \frac{q_n}{n}\log n, \frac{q_n}{n}\log\frac{1}{\varphi(\eta_{q_n}^2)}, q_n^{3/2}\frac{k_n^d}{\sqrt{n}} + \frac{q_n}{k_n}\} \prec \delta^* \prec \varphi(\eta_{q_n}^2).$$

Therefore, all the conditions in Lemma C.9 are checked. The theorem follows from Lemmas C.5, C.6 C.9, and Theorem 2.2. Q.E.D.

C.4. Proof of Corollaries 4.1, 4.2, 4.3.

PROOF. When $\eta_{q_n} \sim q_n^{-v}$ for some $v > 0$, then $\varphi(\eta_{q_n}^2) = \varphi(q_n^{-2v})$.

- In the mildly ill-posed case, $\varphi(\eta_{q_n}^2) = q_n^{-2v\alpha}$.
- In the severely ill-posed case, $\varphi(\eta_{q_n}^2) = \exp(-q_n^{2v\alpha})$.

The corollaries then follows immediately from Theorems 4.3, 4.4 and straightforward calculations. Q.E.D.

APPENDIX D: PROOFS FOR SECTION 5

Proof of Theorem 5.1

PROOF. We will also apply Theorems 2.1 and 2.2 to prove the consistency results. Here the parameter is (q, b) instead of b. The prior proposes a parameter (q, b), which corresponds to a function $g = \sum_{i=1}^q b_i\phi_i(x)$. The dimension q specifies that the vector b, if regarded as infinite dimensional, has all 0 components for $(b_j)_{j>q}$. We define $\mathcal{F}_n = \{\sum_{i=1}^{Mq_n} b_i\phi_i(x) : b \in \tilde{F}_n\}$. The restriction $b \in \tilde{F}_n$ can take the form of either $\|b\| \leq B_n$, or $\|b\|_\infty \leq B_n$, in Sections 3 and 4.

We will discuss the conditions of Theorems 2.1 and 2.2, with δ_n chosen to be the same as the one specified in a deterministic case of Sections 3 and 4.

Uniform Convergence Condition (iii):

The uniform convergence condition (iii) is the same as in the deterministic case before, with a sieve dimension $q = Mq_n$, which is satisfied under the conditions placed on q_n and $\tilde{F}_n$.

Tail Condition (i):

There are two versions of the tail Condition. For the version in the prior tail,

$$\pi(g_b \notin \mathcal{F}_n) \leq \pi(q > Mq_n) + \sum_{q \leq Mq_n} \pi(b \notin \tilde{F}_n|q)\pi(q)$$

$$\leq \pi(q > Mq_n) + \max_{q \leq Mq_n} \pi(b \notin \tilde{F}_n|q).$$

The first term is zero (one can allow it to have a very thin nonzero tail also). For the second term, in all situations considered, one can easily verify that

$$\max_{q \leq Mq_n} \pi(b \notin \tilde{F}_n|q) = \pi(b \notin \tilde{F}_n|q = Mq_n),$$

which is the tail probability in the deterministic case with sieve dimension $q = Mq_n$, which satisfies the required upperbound in the tail condition. Therefore, $\pi(g_b \notin \mathcal{F}_n)$ also satisfies the same upperbound of the tail condition.

For the version of the posterior tail condition, the posterior instead of the prior will be used in the previous argument. For each q_n, an upperbound of $P(b \notin \tilde{F}_n|q, X^n)$ can be derived using the Chebyshev inequality as in the proof Lemma C.8, and the largest upperbound occurs at the largest sieve dimension q_n: $q = Mq_n$. The condition placed on q_n in the deterministic case then ensures that the posterior tail satisfies the tail condition (i).

Approximation Condition (ii):

Due to the uniform prior on q_n,

$$\pi(G(g_b) < \delta_n, g_b \in \mathcal{F}_n) \geq (Mq_n)^{-1}\pi(G(g_b) < \delta_n, g_b \in \mathcal{F}_n|q = Mq_n).$$

Therefore, if (*) $1/q_n = e^{-o(n\delta_n)}$ (this is implied by $\delta_n \succ (\log q_n)/n$) and (**) $\pi(G(g) < \delta_n, g \in \mathcal{F}_n|q = Mq_n) \succ e^{-1.9n\delta_n}$, then $\pi(G(g) < \delta_n, g \in \mathcal{F}_n) \succ e^{-2n\delta_n}$.

So the approximation condition (ii) is satisfied.

In the derivation of all but two theorem in Sections 3 and 4, neither of the requirements (*) and (**) generate new conditions. (For example, $\delta_n \succ (\log q_n)/n$ is implied by $\mu_n = o(\delta_n)$ in Theorem 3.1.)

The two exceptions are for analogs of Theorems 3.2 and 3.4, where by following the same ideas as in the proofs of Theorems 3.2 and 3.4, we can derive consistency results by making the additional assumptions $(\log q_n)/n = o(\delta_n)$ and $(\log q_n)/n = o(\inf_{g \in \mathcal{H}_n, g \notin \Theta_I^\epsilon} G(g))$, respectively.

Distinguishing Ability Condition (iv):

Regarding the quantity $\inf_{g\in\mathcal{F}_n, g\notin\Theta_I^\epsilon} G(g)$, it is the same as in a deterministic case with $q = Mq_n$, since $\mathcal{F}_n = \{\sum_{1=1}^q b_j\phi_j : q = Mq_n, b \in \tilde{F}_n\}$. Then the same derivations for the results in Sections 3 and 4 imply that $\inf_{g\in\mathcal{F}_n, g\notin\Theta_I^\epsilon} G(g) \succ \delta_n$.

Since all conditions (i) to (iv) are satisfied, we have verified the risk consistency and estimation consistency of the posterior, for all situations considered in Sections 3 and 4.

Q.E.D.

REFERENCES

[1] CHEN, X. and POUZO, D. (2009a). Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals. *Manuscript*. Yale University.

[2] HAN, C. and PHILLIPS, P. (2006). GMM with many moment conditions. *Econometrica*. **74** 147-192

[3] JIANG, W. and TANNER, M. (2008), Gibbs posterior for variable selection in high dimensional classification and data mining. *Ann. Statist.* **36** 2207-2231.

[4] LIAO, Y. and JIANG, W. (2011), Posterior consistency of nonparametric conditional moment restricted models. *Manuscript.* Northwestern University

[5] NEWEY, W. (1991). Uniform convergence in probability and stochastic equicontinuity. *Econometrica*. **59** 1161-1167.

[6] NEWEY, W. and MCFADDEN, D. (1994), Large sample estimation and hypothesis testing. *Handbook of Econometrics*. **IV** ch 36

DEPARTMENT OF OPERATIONS RESEARCH
AND FINANCIAL ENGINEERING
PRINCETON UNIVERSITY
SHERRERD HALL
PRINCETON, NJ, 08544
E-MAIL: yuanliao2008@u.northwestern.edu

DEPARTMENT OF STATISTICS
NORTHWESTERN UNIVERSITY
2006 SHERIDAN RD
EVANSTON, IL, 60208
E-MAIL: wjiang@northwestern.edu