Penalized Posterior with a Shrinking Prior (Supplementary Material B for: Posterior Consistency of Nonparametric Conditional Moment Restricted Models)

Yuan Liao
Princeton University

Wenxin Jiang
Northwestern University

September 16, 2011

Abstract

This is a supplementary material of Liao and Jiang (2011). The regularization is carried out by a truncated normal prior with shrinking variance, instead of a slowly-growing sieve dimension. The log-prior is then a regularization penalty attached to the log-likelihood. As a result, Conditions (3.10), (3.11) and Assumption 4.5 of the main paper are relaxed. A larger sieve dimension q_n can be allowed.

Keywords: partial identification, weakly compact, identified region, penalty, truncated shrinking prior, shrinkage prior, penalty function.

AMS 2000 subject classifications: Primary 62F15, 62G08, 62G20; secondary 62P20

1 Introduction

1.1 The model

We consider the nonparametric conditional moment restricted model:

$$E[\rho(Z, g_0)|W, g_0] = 0, \tag{1.1}$$

where g_0 is the parameter of interest, which is assumed to lie in the interior of an infinite dimensional Banach space $\mathcal{H}$ endowed with norm $\|.\|_s$. Depending on the conditional distribution $Z|W, g_0$, model (1.1) may only partially identify g_0. We define the identified region of g_0 to be:

$$\Theta_I = \{g_0 \in \mathcal{H} : E[\rho(Z, g_0)|W, g_0] = 0 \text{ a.s.}\}.$$

We construct the posterior distribution of g_0 on a sieve space $\mathcal{H}_n$, which is spanned by a set of basis functions $\{\phi_1, \phi_2, ...\phi_{q_n}\}$, where $q_n \to \infty$ as $n \to \infty$. Suppose for each $g_0 \in \Theta_I$, there exists $\{b_i^*\}_{i=1}^{\infty}$ such that if $g_{q_n}^* = \sum_{i=1}^{q_n} b_i^* \phi_i$, then $\|g_{q_n}^* - g_0\|_s = o(1)$ as $q_n \to \infty$.

Following the notation in the main paper, we construct the posterior to be:

$$P(g_b|X^n) \propto e^{-n\bar{G}(g_b)/2}\pi(b),$$

where $\pi(b)$ is the product of q_n independent identically distributed priors $\pi_i(b_i)$. Our goal is to show that, for this type of posterior, $\forall \epsilon > 0$,

$$P(g_b \in \Theta_I^{\epsilon}|X^n) \to^p 1$$

as $n \to \infty$, where $\Theta_I^{\epsilon} = \{g : \inf_{h \in \Theta_I} \|g - h\|_s < \epsilon\}$.

1.2 Regularization scheme

As is described in the main paper, achieving the posterior consistency using the traditional techniques in the Bayesian literature is difficult, because of the *ill-posed inverse problem*. Namely, let

$$G(g) = E_W[E(\rho(Z, g)|W)]^2,$$

then

$$\liminf_{n \to \infty} \inf_{g \in \mathcal{H}_n, g \notin \Theta_I^{\epsilon}} G(g) = 0.$$

Hence some techniques of regularization are needed.

Liao and Jiang (2011) used a slowly-growing-sieves method for regularization, by restricting q_n to grow slowly as $n \to \infty$, which is similar to Chen and Pouzo's small sieves technique. When q_n grows slowly, $\inf_{g\in\mathcal{H}_n, g\notin\Theta_I^\epsilon} G(g)$ may decay to zero at a not-too-fast rate. In the main paper, we found a lower-bound rate $\delta_n^* \to 0$, such that (See Liao and Jiang (2011), Theorem 3.3, 3.4) as long as

$$\delta_n^* = o(\inf_{g\in\mathcal{H}_n, g\notin\Theta_I^\epsilon} G(g)), \tag{1.2}$$

the posterior is consistent.

Chen and Pouzo (2009) verified condition (1.2) in the nonparametric instrumental variable model when g_0 is identified. They used the singular value system of a compact operator T. When g_0 is identified, and $\mathcal{H} = L^2(X) = \{g : Eg(X)^2 < \infty\}$, one of the two sets of the eigenfunctions of T forms an orthonormal basis of $\mathcal{H}$ (Kress 1999, ch 15). Realizing the limitation of the use of the eigenfunctions as the basis functions, in the main paper, we imposed Assumption 4.5 (of the main paper) to show that condition (1.2) holds for a general set of orthonormal basis functions, allowing g_0 to be only partially identified.

Instead of a small q_n, the regularization scheme can be also constructed using penalization. Chen and Pouzo (2009) defined an estimator as the solution to the following optimization problem:

$$\inf_{g\in\mathcal{H}_n} \bar{G}(g) + P_n(g) \tag{1.3}$$

for some penalty function P_n, where $\bar{G}(g)$ is a consistent estimator of $G(g)$. With some regularity conditions on $P_n(.)$, they showed that such an estimator is consistent. From the Bayesian point of view, we can construct a prior

$$\pi(b) \propto \exp(-nP_n(g_b));$$

then the log-prior $-nP_n(g_b)$ becomes a regularization penalty attached to the pseudo log-likelihood $-n\bar{G}$. Such a regularization technique is commonly used in the Bayesian literature. For example, in the parametric case, the penalty term in ridge regression can be viewed as a regularizing prior; see, e.g., Haitovsky and Wax (1980). In the nonparametric case, recently Florens and Simoni (2009b) specified a regularizing prior distribution that is an extension of Zellner's g-prior for the regularization.

In this supplementary material, we use a truncated normal prior whose variance decays to zero: define

$$\mathcal{F}_n = \{b = (b_1, ..., b_{q_n})^T : \max_{i\le q_n} |b_i| \le B_n\}$$

for some $B_n \to \infty$. Let

$$\pi(b) = \frac{I(\max_{i \le q_n} |b_i| \le B_n)}{P(\max_{i \le q_n} |Z_i| \le B_n)} e^{-na_n^2 \|b\|^2} \left(\frac{2na_n^2}{2\pi} \right)^{q_n/2}, \quad Z = (Z_1, ..., Z_{q_n}) \sim^{i.i.d.} N(0, (2na_n^2)^{-1}),$$

for some positive sequence $a_n \to 0$ and $na_n^2 \to \infty$. The prior is only supported on $\mathcal{F}_n$ and shrinks to the point mass zero. Note that the role of regularization is then played by the prior instead of the finite sieve dimension. As a result, we do not need to impose Conditions (3.10) or (3.11) in the main paper any more. This is parallel to the frequentist approach used in Chen and Pouzo (2009, Theorem 3.2).

Throughout this paper, for a q_n-dimensional vector b, let $\|b\| = \sum_{i \le q_n} b_i^2$.

2 General Posterior Consistency Theorem: Penalized Version

Suppose $\inf_{g \in \mathcal{H}} G(g) = 0$. We will show the following inequality in the Appendix:

Lemma 2.1. *Suppose the prior support can be partitioned as $\mathcal{F}_n \cup \mathcal{F}_n^c$. For any penalty function $P_n(.) \ge 0$, any $\tau \in (0,1)$, and any positive sequence δ_n,*

$$\begin{aligned} EP(G(g_b) + P_n(b) \ge \delta_n | D) &\le P(\sup_{g_b \in \mathcal{F}_n} |\bar{G}(g_b) - G(g_b)| > \frac{\tau}{4}\delta_n) \\ &+ \frac{\pi(\mathcal{F}_n^c) e^{3\tau n \delta_n/4} + e^{-(1-\tau)n\delta_n} E_\pi(e^{nP_n(b)})}{\pi(\mathcal{F}_n \cap G(g_b) < \tau\delta_n/2)}. \end{aligned} \tag{2.1}$$

We first present two general posterior consistency results, which are the penalized versions of those results in Section 2 of the main paper.

Theorem 2.1 (Penalized Risk Convergence)**.** *Suppose there exists $\tau \in (0,1)$ such that the following conditions hold with respect to a deterministic positive sequence δ_n:*

(i) Tail condition: as q_n and $n \to \infty$, $\pi(\mathcal{F}_n^c) = O(e^{-n\delta_n})$.

(ii) Approximation condition: $\pi(G(g_b) < \tau\delta_n/2,\ g_b \in \mathcal{F}_n) \succ e^{-3(1-\tau)n\delta_n/4}$.

(iii) Uniform convergence: $\sup_{g \in \mathcal{F}_n} |\bar{G}(g) - G(g)| = o_p(\delta_n)$.

(iv) Penalty condition: $E_\pi(e^{nP_n(b)}) = O(e^{(1-\tau)n\delta_n/4})$.

Then we have the penalized risk convergence result at rate δ_n

$$P(G(g_b) + P_n(g_b) < \delta_n | X^n) = 1 - o_p(1).$$

When the following condition is added, the penalized risk convergence leads to the estimation consistency.

Theorem 2.2 (Estimation consistency)**.** *Suppose there exists a sequence δ_n and a penalty function $P_n(.)$ such that the following conditions hold:*

(i)-(iv) in the previous theorem;

(v) (distinguishing ability) For any $\epsilon > 0$,

$$\inf_{g\in\mathcal{H}_n, g\notin\Theta_I^\epsilon} G(g) + P_n(g) \geq \delta_n.$$

Then for any $\epsilon > 0$, we have

$$P(g_b \in \Theta_I^\epsilon | X^n) \to^p 1. \tag{2.2}$$

We will then verify the conditions listed in these two theorems in the following section.

3 Posterior Consistency Using a Shrinking Prior

3.1 Prior specification and assumptions

Throughout this paper, we define $\mathcal{H} = L^2(X) = \{g : Eg(X)^2 < \infty\}$, and $\|g\|_s^2 = Eg(X)^2$. As in the main paper, we use the limited information likelihood function:

$$L(g_b) = \exp(-\frac{n}{2}\bar{G}(g_b)),$$

where

$$\bar{G}(g_b) = \bar{m}_n(g_b)^T \hat{V}^{-1} \bar{m}_n(g_b), \tag{3.1}$$

with $\bar{m}_n(g_b) = \frac{1}{n}\sum_{i=1}^n (\rho(Z_i, g_b) I_{(W_i \in R_1^n)}, ..., \rho(Z_i, g_b) I_{(W_i \in R^n_{k_n^d})})^T$, and $\hat{V} = \text{diag}\{\frac{1}{n}\sum_{i=1}^n I_{(W_i\in R_j^n)}\}_{j=1}^{k_n^d}$. The identified region Θ_I is defined as the set of $g \in \mathcal{H}$ such that $G(g) = 0$, where

$$G(g) = E_W[E(\rho(Z, g)|W, g_0)]^2.$$

We use a truncated normal prior with a shrinking variance: define

$$\mathcal{F}_n = \{b = (b_1, ..., b_q)^T : \max_{i\leq q_n} |b_i| \leq B_n\}$$

for some $B_n \to \infty$. Let

$$\pi(b) = \frac{I(\max_{i\leq q_n}|b_i| \leq B_n)}{P(\max_{i\leq q_n}|Z_i| \leq B_n)} e^{-na_n^2\|b\|^2} \left(\frac{2na_n^2}{2\pi}\right)^{q_n/2}, \quad Z_i \sim^{iid} N(0, (2na_n^2)^{-1}), \tag{3.2}$$

for some positive sequence $a_n \to 0$ and $na_n^2 \to \infty$. The prior is only supported on $\mathcal{F}_n$ and shrinks to the point mass zero.

Assumption 3.1. *The data $X^n = (X_1, ..., X_n)$ are independent and identically distributed.*

Assumption 3.2. *There exists a positive sequence $\lambda_n \to 0$ such that*

$$\sup_{g \in \mathcal{F}_n} |\bar{G}(g) - G(g)| = O_p(\lambda_n).$$

Assumption 3.3. *$\{\phi_1, \phi_2, ...\}$ forms an orthonormal basis of $\mathcal{H}$, and for any $g_0 = \sum_{i=1}^{\infty} b_i^* \phi_i \in \Theta_I$, its sieve approximation $g_{q_n}^* = \sum_{i=1}^{q_n} b_i^* \phi_i$ satisfies $\|g_{q_n}^* - g_0\|_s = o(1)$ as $q_n \to \infty$.*

Assumption 3.4. *There exists $C > 0$ such that $\forall g_1, g_2 \in \mathcal{H}$,*

$$E|\rho(Z, g_1) - \rho(Z, g_2)| \leq CE|g_1(X) - g_2(X)|.$$

3.2 Posterior consistency

As in the main paper, define

$$\gamma_n = \sup_{g \in \mathcal{F}_n, w \in [0,1]^d} |E(\rho(Z, g)|W = w)| + 1. \tag{3.3}$$

For $g_b = \sum_{i=1}^{q_n} b_i \phi_i$, define the penalty function to be

$$P_n(g_b) = M a_n^2 \|b\|^2$$

for some $M \in (0, 1)$. Applying Theorem 2.1, we obtain the following penalized risk consistency result first:

Theorem 3.1 (Penalized risk consistency)**.** *Suppose $a_n^2 \to 0$ is such that $\frac{q_n}{n} \log(n\gamma_n) = o(a_n^2)$. In addition, suppose $\delta_n = O(1)$ is such that, there exists $g_0 \in \Theta_I$,with $g_{q_n}^*$ being its sieve approximation on $\mathcal{H}_n$ satisfying:*
(i) There exist constants $C_1 > 1$, $C_2 > 0$ such that, for all large enough n,

$$\delta_n \geq (C_2 + C_1\|g_0\|)^2 a_n^2$$

(ii)

$$\max\{\lambda_n, G(g_{q_n}^*)\} = o(\delta_n).$$

Then under Assumptions 3.1-3.4, for all $M \in (0,1)$,

$$P(G(g_b) + Ma_n^2\|b\|^2 \geq \delta_n|X^n) = o_p(1).$$

To achieve the posterior consistency under $\|.\|_s$, we require the following additional assumptions.

Assumption 3.5. *$G(g)$ is convex and lower semicontinuous on $\{g : Eg(X)^2 \leq M\}$ for any $M > 0$.*

Since $\mathcal{H}$ is reflexive, and any l_2 ball in $\mathcal{H}$ is both weakly compact and weak sequentially compact, this assumption then ensures that for any $M > 0$, the functional $G(g) + Ma_n^2\|g\|_s^2$ has a minimum on any l_2-ball in $\mathcal{H}$, a technical fact to be used in the proof (see Zeidler (1985), Proposition 38.12).

Assumption 3.6. $\sup_{g\in\Theta_I}\|g\|_s < \infty$.

Comments on Assumption 3.6

1. When g_0 is partially identified, this assumption implicitly requires that the parameter space $\mathcal{H}$ be bounded. For example, in the NPIV model, suppose there exists $g_1 \neq 0$ such that $E(g_1(X)|W) = 0$ a.s., then for any $a \in \mathbb{R}$, and $g_0 \in \Theta_I$, if $g_0 + ag_1 \in \mathcal{H}$, then $g_0 + ag_1 \in \Theta_I$. But $\|g_0 + ag_1\|_s \geq a\|g_1\|_s - \|g_0\|_s$. As a can be made arbitrarily large, so is $\|g_0 + ag_1\|_s$, unless $\mathcal{H}$ is bounded.

2. Our method is general enough so that it covers both the point identified and partially identified cases. When $g_0 = \sum_{i=1}^{\infty} b_i^*\phi_i$ is point identified, i.e., $\Theta_I = \{g_0\}$, the parameter space is allowed to be unbounded (as opposed to Assumption 3.9 of Chen and Pouzo (2009)). In that case, Assumption 3.6 is naturally satisfied.

3. If, however, we only restrict to the case when g_0 is partially identified, i.e., Θ_I is not a singleton, a more refined approach is to assume that the parameter space $\mathcal{H}$ is a bounded subset (but still non-compact) of $L^2(X)$. In that case, by directly defining $\mathcal{F}_n = \mathcal{H}$, and using a shrinking normal prior truncated on $\mathcal{H}$, we can still achieve the posterior consistency.

With the shrinking truncated normal prior as the regularization scheme, we can obtain the following posterior consistency theorem without imposing Conditions (3.10) and (3.11) in the main paper. It also allows the use of a larger sieve dimension q_n.

Theorem 3.2 (Posterior consistency using shrinking prior). *Suppose $a_n^2 \to 0$ and there exists $g_0 \in \Theta_I$, $g_{q_n}^*$ being its sieve approximation such that*

$$\max\{\lambda_n, \frac{q_n}{n}\log(n\gamma_n), G(g_{q_n}^*)\} = o(a_n^2).$$

Then under Assumptions 3.1-3.6, for any $\epsilon > 0$,

$$P(d(g_b, \Theta_I) > \epsilon | X^n) = o_p(1).$$

3.3 Application: NPIV

The nonparametric instrumental variable regression model is given by

$$Y = g_0(X) + \epsilon,$$

where $E[\epsilon|W] = 0$; W is a d- dimensional instrumental variable.

The following assumptions are imposed in the main paper.

Assumption 3.7. *(i) $k_n^{-d} = O(\min_{j \leq k_n^d} P(W \in R_j^n))$,*
(ii) $\max_{j \leq k_n^d} P(W \in R_j^n) = O(k_n^{-d})$.

Assumption 3.8. *There exists $C > 0$ such that for all $i = 1, ..., q_n$*
(i) $\sup_w E(Y^2|W = w) < C$, $\sup_w E(\phi_i(X)^2|W = w) < C$;
(ii) $E(Y|W = w)$ is Lipschitz continuous with respect to w on $[0, 1]^d$;
(iii) For any $w_1, w_2 \in [0, 1]^d$,

$$|E(\phi_i(X)|W = w_1) - E(\phi_i(X)|W = w_2)| \leq C\|w_1 - w_2\|.$$

Assumption 3.9. *There exists $g_0 \in \Theta_I$, $g_{q_n}^* = \sum_{i=1}^{q_n} b_i^* \phi_i$ with $\sum_{i=1}^{\infty} b_i^{*2} < \infty$, and a positive sequence $\{\eta_j\}_{j=1}^{\infty}$ that strictly decreases to zero as $j \to \infty$ such that $\|g_{q_n}^* - g_0\|_s = O(\eta_{q_n})$ as $q_n \to \infty$.*

Since the regularization is carried out by the shrinking prior, there is no need to verify either Condition (3.10) or (3.11) of the main paper. As a result, Assumption 4.5 for the NPIV model in the main paper is relaxed. Theorem 3.2 immediately implies the following consistency result.

Theorem 3.3. *Under Assumptions 3.1, 3.6-3.9, suppose $a_n^2 \to 0$ and*

$$\max\left\{\frac{q_n^2 B_n^2 k_n^{3d/2}}{\sqrt{n}} + \frac{q_n^2 B_n^2}{k_n}, \eta_{q_n}^2\right\} = o(a_n^2), \tag{3.4}$$

then for any $\epsilon > 0$,

$$P(d(g_b, \Theta_I) > \epsilon | X^n) = o_p(1).$$

A Proofs for Section 2

Theorem 2.1 is implied directly from Lemma 2.1 given conditions (i)-(iv). The proof of Theorem 2.2 is similar to the proof of Theorem 2.2 in the main paper. Hence we only show Lemma 2.1.

Proof of Lemma 2.1

Proof. Note that

$$P(G(g_b) + P_n(b) \geq \delta_n | X^n) \leq \frac{\int_{\mathcal{F}_n^c} e^{-n\bar{G}(g_b)/2} d\pi + \int_{\mathcal{F}_n, G(g_b)+P_n(b)\geq\delta_n} e^{-n\bar{G}(g_b)/2} d\pi}{\int e^{-n\bar{G}(g_b)/2} d\pi}.$$

The terms on the right hand side of this inequality can be bounded respectively as follows:

$$\int_{\mathcal{F}_n, G(g_b)+P_n(b)\geq\delta_n} e^{-n\bar{G}(g_b)/2} d\pi \leq e^{n \sup_{\mathcal{F}_n} |\bar{G}-G|/2 - n\delta_n} \int e^{nP_n(b)} d\pi.$$

$$\int_{\mathcal{F}_n^c} e^{-n\bar{G}/2} d\pi \leq \pi(\mathcal{F}_n^c).$$

$$\int e^{-n\bar{G}/2} d\pi \geq \int_{\mathcal{F}_n, G(g_b)<\delta_n\tau/2} e^{-n\bar{G}/2} d\pi \geq e^{-n \sup_{\mathcal{F}_n} |\bar{G}-G|/2 - n\delta_n\tau/2} \pi(\mathcal{F}_n \cap G(g_b) < \delta_n\tau/2).$$

These imply the desired inequality. Q.E.D.

B Proofs for Section 3

B.1 Proof of penalized risk consistency

Lemma B.1. *Suppose* $g^*_{q_n}$ *is the sieve approximation of some* $g_0 \in \Theta_I$*, and* $G(g^*_{q_n}) = o(\delta_n)$*. Then for any* $\tau \in (0, 1)$,

$$\pi(\mathcal{F}_n \cap G(g_b) < \tau\delta_n/2) \geq$$

$$\frac{1}{P(\max_{i\leq q_n} |Z_i| \leq B_n)} \exp[-na_n^2(\|g^*_{q_n}\|^2 + O(\delta_n)) + \frac{q_n}{2}\log(Cna_n^2) - \frac{q_n}{2}\log q_n + \frac{q_n}{2}\log\frac{\delta_n^2}{\gamma_n^2}].$$

where $Z = (Z_1, ..., Z_q)^T \sim N_q(0, (2na_n^2)^{-1})$.

Proof. since $G(g^*_{q_n}) = o(\delta_n)$, and for any $g_b = \sum_{i\leq q_n} b_i\phi_i$,

$$|G(g_b) - G(g^*_{q_n})| \leq 2\gamma_n E|\rho(Z, g_b) - \rho(Z, g^*_{q_n})| \leq C\gamma_n E|g_b(X) - g^*_{q_n}(X)| \leq C\gamma_n \|b - b^*\|,$$

hence $\{G(g_b) < \tau\delta_n/2\} \supset \{\|b - b^*\| < \frac{\tau\delta_n}{4C\gamma_n}\}$. Then for some $\tilde{b} \in \{\|b - b^*\| \leq \frac{\tau\delta_n}{4C\gamma_n}\}$

$$\pi(\mathcal{F}_n \cap G(g_b) < \tau\delta_n/2) \geq \frac{e^{-na_n^2\|\tilde{b}\|^2}}{P(\max_i |Z_i| \leq B_n)} \left(\frac{na_n^2}{\pi}\right)^{q_n/2} \mu(\max_i |b_i| \leq B_n, \|b - b^*\| < \frac{\tau\delta_n}{4C\gamma_n}),$$

where $Z = (Z_1, ..., Z_q)^T \sim N_q(0, (2na_n^2)^{-1})$. Note that for large n, $\|b - b^*\| < \frac{\tau\delta_n}{4C\gamma_n}$ implies $\max_i |b_i| \leq B_n$. Hence

$$\mu(\max_i |b_i| \leq B_n, \|b - b^*\| < \frac{\tau\delta_n}{4C\gamma_n}) \geq (\frac{\tau\delta_n}{4C\sqrt{q_n}\gamma_n})^{q_n}.$$

In addition, $\|\tilde{b}\| \leq \|b^*\| + O(\delta_n)$. It then yields the result.

Lemma B.2. *For any $M \in (0, 1)$, let $P_n(b) = Ma_n^2\|b\|^2$. Then*

$$E_\pi e^{nP_n(b)} \leq \left(\frac{1}{1 - M}\right)^{q_n/2} \frac{1}{P(\max_{i\leq q_n} |Z_i| < B_n)}.$$

Proof. Note that for $H = (H_1, ..., H_{q_n}) \sim N_q(0, (2(1 - M)na_n^2)^{-1})$,

$$\frac{1}{P(\max_i |H_i| \leq B_n)} \int_{\max_i |b_i| \leq B_n} \left(\frac{2(1 - M)na_n^2}{2\pi}\right)^{q_n/2} e^{-(1-M)na_n^2\|b\|^2} db = 1.$$

The result then follows from direct calculation. Q.E.D.

Proof of Theorem 3.1 By Lemma 2.1, when the truncated normal prior is used, and $P_n(b) = Ma_n^2\|b\|^2$ for some $M \in (0, 1)$, then for any $\tau \in (0, 1)$, and any positive sequence δ_n

$$\begin{aligned} &EP(G(g_b) + Ma_n^2\|b\|^2 \geq \delta_n|D) \leq P(\sup_{g_b\in\mathcal{F}_n} |\bar{G}(g_b) - G(g_b)| > \frac{\tau}{4}\delta_n) \\ &+\frac{e^{-(1-\tau)n\delta_n} E_\pi(e^{nP_n(b)})}{\pi(\mathcal{F}_n \cap G(g_b) < \tau\delta_n/2)}. \end{aligned} \quad \text{(B.1)}$$

Suppose $g_0 \in \Theta_I$ is such that $\|g^*_{q_n} - g_0\| = o(1)$. By Lemma B.1 B.2, for any arbitrarily small $r > 0$, since $na_n^2\delta_n^2 \to \infty$,

$$\begin{aligned} &\frac{e^{-(1-\tau)n\delta_n} E_\pi(e^{nP_n(b)})}{\pi(\mathcal{F}_n \cap G(g_b) < \tau\delta_n/2)} \\ \leq\ & \exp(-(1 - \tau)n\delta_n + na_n^2(\|g_0\|^2 + o(1)) + \frac{q_n}{2} \log q_n\gamma_n^2 - \frac{q_n}{2} \log Cna_n^2\delta_n^2) \end{aligned}$$

$$\leq \exp(-(1-\tau)n\delta_n + na_n^2(\|g_0\|^2 + o(1)) + \frac{q_n}{2}\log q_n\gamma_n^2 + \frac{q_n}{2}\log\frac{C}{\delta_n^2}). \tag{B.2}$$

Since $n\delta_n \succ q_n \log n$, $n\delta_n \succ q_n \log \gamma_n$, then $n\delta_n \succ q_n \log \frac{C}{\delta_n^2} + q_n \log q_n\gamma_n$ for any $C > 0$. If $g_0 = 0$, then

$$(1-\tau)n\delta_n \geq 2(na_n^2(\|g_0\|^2 + o(1)) + \frac{q_n}{2}\log q_n\gamma_n^2 + \frac{q_n}{2}\log\frac{C}{\delta_n^2})$$

for all $\tau \in (0,1)$ and all large n. Thus the left hand side of (B.2) converges to zero.

If $\|g_0\| > 0$, $\forall\theta > 0$, for all large n, $\frac{q_n}{2}\log q_n\gamma_n^2 + \frac{q_n}{2}\log\frac{C}{\delta_n^2} < \theta n a_n^2\|g_0\|^2$, which implies that the right hand side of (B.2) is dominated by

$$\exp(-(1-\tau)n\delta_n + na_n^2((1+\theta)\|g_0\|^2 + o(1))). \tag{B.3}$$

We can find a small enough $\tau \in (0,1)$ and some $r_0 \in (0,1)$, such that $(1+\tau)^2 < r_0 C_1(1-\tau)$ as $C_1 > 1$. In addition, pick up $\theta < (\tau^2 + 2\tau)/2$. Hence

$$(1+\theta_n)\|g_0\|^2 + o(1) \leq (1+\tau)^2\|g_0\|^2 < C_1 r_0(1-\tau)\|g_0\|^2 \leq r_0(1-\tau)(C_1\|g_0\| + C_2)^2.$$

Hence we have shown that, either $\|g_0\| = 0$ or $\|g_0\| > 0$, there exist $\tau, r_0 \in (0,1)$,

$$a_n^2((1+\theta_n)\|g_0\|^2 + o(1)) < r_0 a_n^2(1-\tau)(C_1\|g_0\| + C_2)^2 \leq r_0\delta_n(1-\tau),$$

which implies

$$\frac{e^{-(1-\tau)n\delta_n}(1-M)^{-q/2}}{\pi(\mathcal{F}_n \cap G(g_b) < \tau\delta_n/2)} = o(1).$$

We can choose this τ for Lemma 2.1, and together with the fact that $P(\sup_{g_b\in\mathcal{F}_n} |\bar{G}(g_b) - G(g_b)| > \frac{\tau}{4}\delta_n) = o_p(1)$ to obtain the result. Q.E.D.

B.2 Proof of posterior consistency

Lemma B.3. *For any $\epsilon > 0$, and $M \in (0,1)$, if*

$$\delta_n \leq M a_n^2(\inf_{h\in\Theta_I} \|h\|_s^2 + \epsilon^2),$$

then we have:

$$\inf_{g_b\in\mathcal{H}_n, g_b\notin\Theta_I^\epsilon} G(g_b) + M a_n^2\|b\|^2 \geq \delta_n.$$

Proof. Let $\Omega_n = \mathcal{H}_n \cap (\Theta_I^\epsilon)^c$. For any $g_b \in \Omega_n$ and any constant $K > 0$, if $M a_n^2\|b\|^2 \geq$

$\delta_n + a_n^2 K$, then the result is proved. Hence we restrict ourselves to the set

$$A_n = \Omega_n \cap \{g_b \in \mathcal{H}_n, \|b\|^2 \leq \delta_n/(Ma_n^2) + K\}.$$

Write $\int g_1 g_2 = \int g_1(x)g_2(x)dF(x)$. Then for all $g \in L^2(X)$, and $g_b = \sum_{i \leq q_n} b_i \phi_i$, $\|b\|^2 - \|g\|_s^2 - 2\int g(g_b - g) = \|g_b - g\|_s^2$. Hence we have, for all $g_0 \in \Theta_I$,

$$\|b\|^2 = \|g_b - g_0\|_s^2 + 2\int g_0(g_b - g_0) + \|g_0\|_s^2,$$

which implies, by the definition of A_n,

$$\begin{aligned}
&\inf_{g_b \in A_n} G(g_b) + Ma_n^2\|b\|^2 = \inf_{g_b \in A_n}[G(g_b) + Ma_n^2 \sup_{g_0 \in \Theta_I}(\|g_b - g_0\|_s^2 \\
&+2\int g_0(g_b - g_0) + \|g_0\|_s^2)] \\
\geq\ & Ma_n^2\epsilon^2 + \inf_{g_b \in A_n}[G(g_b) + Ma_n^2 \sup_{g_0 \in \Theta_I}(2\int g_0(g_b - g_0) + \|g_0\|_s^2)] \\
\geq\ & Ma_n^2\epsilon^2 + \inf_{g_b \in A_n}[G(g_b) + Ma_n^2 \sup_{g_0 \in \Theta_I}(2\int g_0(g_b - g_0)] + Ma_n^2 \inf_{g_0 \in \Theta_I}\|g_0\|_s^2.
\end{aligned}$$

By the condition in the lemma, it suffices to show

$$\liminf_{n\to\infty} \inf_{g_b \in A_n}[\frac{G(g_b)}{a_n^2} + \sup_{g_0 \in \Theta_I} 2\int g_0(g_b - g_0)] \geq 0$$

In addition, $\delta_n/a_n^2 \leq M(\inf_{g_0 \in \Theta_I}\|g_0\|_s^2 + \epsilon^2)$, which yields, for each n,

$$A_n \subset \{g \in \mathcal{H} : \|g\|_s^2 \leq C_2 + K\}$$

for a large constant C_2. It suffices to show

$$\liminf_{n\to\infty} \inf_{g \in L^2(X): \|g\|_s^2 \leq C_2 + K}[\frac{G(g)}{a_n^2} + 2\sup_{g_0 \in \Theta_I}\int g_0(g - g_0)] \geq 0. \tag{B.4}$$

The $L^2(X)$ space is a reflexive Banach space, and the ball $\{g \in L^2(X) : \|g\|_s^2 \leq C_2 + K\}$ is both weakly compact and weak sequentially compact in $L^2(X)$, on which $G(g)$ is weak sequentially lower semicontinuous. Therefore, as a result of Proposition 38.12 of Zeidler (1985), for each n, the minimization problem

$$\inf_{g \in L^2(X): \|g\|_s^2 \leq C_2 + K}[\frac{G(g)}{a_n^2} + 2\sup_{g_0 \in \Theta_I}\int g_0(g - g_0)]$$

has a solution $g_n \in \{g \in L^2(X) : \|g\|_s^2 \le C_2 + K\}$. It then suffices to show

$$\liminf_{n\to\infty} \frac{G(g_n)}{a_n^2} + 2 \sup_{g_0\in\Theta_I} \int g_0(g_n - g_0) \ge 0. \tag{B.5}$$

Let g_{nk} be the sub-sequence of g_n such that $\liminf_{n\to\infty} \frac{G(g_n)}{a_n^2} + 2\sup_{g_0\in\Theta_I} \int g_0(g_n - g_0) = \lim_{n\to\infty} \frac{G(g_{nk})}{a_n^2} + 2\sup_{g_0\in\Theta_I} \int g_0(g_{nk} - g_0)$. Morever, g_{nk} has a further sub-sequence g_{nkl} that weakly converges to some $g_\infty \in \{g \in L^2(X) : \|g\|_s^2 \le C_2 + M\}$, which yields,

$$\liminf_{n\to\infty} \frac{G(g_n)}{a_n^2} + 2 \sup_{g_0\in\Theta_I} \int g_0(g_n - g_0) = \lim_{n\to\infty} \frac{G(g_{nkl})}{a_n^2} + 2 \sup_{g_0\in\Theta_I} \int g_0(g_{nkl} - g_0)$$
$$\ge \liminf_{n\to\infty} \frac{G(g_{nkl})}{a_n^2} + 2 \liminf_{n\to\infty} \sup_{g_0\in\Theta_I} \int g_0(g_{nkl} - g_0) \ge \liminf_{n\to\infty} \frac{G(g_{nkl})}{a_n^2} + 2 \sup_{g_0\in\Theta_I} \int g_0(g_\infty - g_0),$$

where in the last inequality, we use the fact that $\liminf \sup \ge \sup \liminf$, and the definition of weak convergence.

If $\liminf_{n\to\infty} \frac{G(g_{nkl})}{a_n^2} = \infty$, since $|\sup_{g_0\in\Theta_I} \int g_0(g_\infty - g_0)| < \infty$ according to the assumption that $\sup_{g_0\in\Theta_I} \|g_0\|_s < \infty$, we have the desired result (B.5).

If $\liminf_{n\to\infty} \frac{G(g_{nkl})}{a_n^2} < \infty$, as $a_n^2 = o(1)$, then $\liminf_{n\to\infty} G(g_{nkl}) = 0$. As G is weak sequentially lower semicontinuous, $G(g_\infty) = 0$, which implies $g_\infty \in \Theta_I$. Hence

$$\sup_{g_0\in\Theta_I} \int g_0(g_\infty - g_0) \ge \int g_\infty(g_\infty - g_\infty) = 0,$$

which also implies the result. Q.E.D.

Proof of Theorem 3.2 By Theorem 3.1 and Lemma B.3, it suffices to show that: $\forall \epsilon > 0$, there exists $M \in (0, 1)$, $C_1 > 1, C_2 > 0$, and $g_0 \in \Theta_I$, such that for all large n,

$$M(\inf_{h\in\Theta_I} \|h\|_s^2 + \epsilon^2) > (C_2 + C_1\|g_0\|_s)^2. \tag{B.6}$$

Because once this inequality holds, with a fixed $g_0 \in \Theta_I$, we can choose

$$\delta_n = \left[(C_2 + C_1\|g_0\|_s)^2 + \frac{M(\inf_{h\in\Theta_I} \|h\|^2 + \epsilon^2) - (C_2 + C_1\|g_0\|_s)^2}{2}\right] a_n^2$$

and apply Theorem 3.1 and Lemma B.3 to show the consistency.

Now we show (B.6): if $\inf_{h\in\Theta_I} \|h\|_s^2 = 0$, for any $M \in (0, 1)$, $C_1 > 1$, let $g_0 \in \Theta_I$ such that $\|g_0\|_s < \sqrt{M}\epsilon/(2C_1)$. Let $C_2 = \sqrt{M}\epsilon/2$, we have the result.

If $\inf_{h\in\Theta_I}\|h\|_s^2>0$, there exists $g_0\in\Theta_I$ such that $\inf_{h\in\Theta_I}\|h\|_s^2>\|g_0\|_s^2-\epsilon^2/2$. Hence

$$\inf_{h\in\Theta_I}\|h\|_s^2+\epsilon^2>\|g_0\|_s^2+\epsilon^2/2.$$

Pick up some $s\in(1,\frac{\epsilon^2}{6\|g_0\|_s^2}+1)$, and $M\in(1/s,1)$. Let $C_1,C_2>0$ be

$$\begin{aligned}&C_2<\min\{\frac{\sqrt{M}\epsilon^2}{12\sqrt{s}\|g_0\|_s},\frac{\sqrt{M}\epsilon}{\sqrt{6}}\}\\&C_1^2=Ms.\end{aligned}$$

Hence $C_1>1$, and $\max\{(C_1^2-M)\|g_0\|_s^2,C_2^2,2C_1C_2\|g_0\|_s\}<\frac{M\epsilon^2}{6}$, which then implies

$$M(\|g_0\|_s^2+\frac{\epsilon^2}{2})>(C_1\|g_0\|_s+C_2)^2.$$

Q.E.D.

References

Chen, X. and Pouzo, D. (2009). Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals. *Manuscript.* Yale University.

Florens, J. and Simoni, A, (2009b). Regularizing priors for linear inverse problems. *Manuscript.* Toulouse School of Economics.

Kress, R. (1999). *Linear integral equations*, Second Edition. Springer

Liao, Y. and Jiang, W. (2011). Posterior consistency of nonparametric conditional moment restricted models. *Manuscript.* Northwestern University.

Haitovsky, Y. and Wax, Y. (1980). Generalized ridge regression, least squares with stochastic prior information, and Bayesian estimators. *Appl. Math. Comput.* **7** 125-154.

Zeidler, E. (1985). *Nonlinear functional analysis and its applications III: variational methods and optimization*, Springer.