

# Efficient estimation using both direct and indirect observations

P.J. Bickel	Y. Ritov
Department of Statistics	Department of Statistics
University of California	The Hebrew University
Berkeley <sup>1</sup>	Jerusalem

September 11, 1992

Abstract:

The Ibragimov Hasminskii model postulates observing  $X_1, \dots, X_m$  independent, identically distributed according to an unknown distribution  $G$  and  $Y_1, \dots, Y_n$  independent and identically distributed according to  $\int k(\cdot, y)dG(y)$  where  $k$  is known, for example,  $Y$  is obtained from  $X$  by convolution with a Gaussian density. We exhibit sieve type estimates of  $G$  which are efficient under minimal conditions which include those of Vardi and Zhang (1992) for the special case,  $G$  on  $[0, \infty]$ ,  $k(x, y) = \frac{1}{y}1(x \leq y)$ .

---

<sup>0</sup>Research supported by NSF Grant DMS-9115577 and the U.S./Israel Binational Science Foundation

# 1 Introduction

Let  $(U, Y)$  be  $\mathcal{U} \times \mathcal{Y}$  valued random elements on a probability space  $(\Omega, \mathcal{A}, \mathcal{P})$  with  $U \sim G$  unknown and  $Y \sim Q$  where  $Q$  has density (with respect to some known measure  $\mu$  on the induced  $\sigma$  field on  $\mathcal{Y}$ ) given by

$$q(y) = \int q(y|u)dG(u), \quad (1.1)$$

and the kernel  $q(\cdot|u)$  is known. Let  $\Delta$  be a third variable independent of  $U$  and  $Y$  such that  $\Delta \in \{0, 1\}$ , and  $P(\Delta = 1) = \lambda \in (0, 1)$ . Suppose we observe a random sample of size  $n$  from  $(X, \Delta)$ ,  $X \equiv \Delta U + (1 - \Delta)Y$ . Let  $m$  be the number of observations such that  $\Delta_i = 1$  and let  $\lambda_n = m/n$ . Our task is the efficient estimation of  $G$ .

These models appear to have first formally been introduced and studied by Ibragimov and Hasminskii (1983), and we shall refer to them as (IH) models. These authors, in fact, considered  $q(y|v) \equiv q(y|u, \theta)$  where  $q$  is smoothly parametrized by the Euclidean parameter  $\theta$ . They obtained information bounds for estimating  $\theta$  in the presence of completely unknown  $G$  and exhibited an efficient estimate of  $\theta$ .

A convenient way of thinking about these models is that we have  $n$  independent pairs  $(U_i, Y_i)$  where  $U_i \sim G$  and  $Y_i$  given  $U_i = u$  has density  $q(\cdot|u)$  and  $X_i = \Delta_i U_i + (1 - \Delta_i)Y_i$ .  $U_i$  is sufficient for  $G$  in  $(U_i, Y_i)$  but  $\Delta_i = 0$  creates a missing  $U_i$  so that we must use  $Y_i$ . If there are no direct observations on  $G$  ( $\lambda = 0$ ) these models have been investigated under the name of mixture models by many authors including Jewell (1982), Lindsay (1983a,b,c), and Pfanzagl (1990). Typically, identifiable pieces of  $\theta$  are estimable at rate  $n^{-1/2}$  but  $G$  can only be estimated poorly. A prototypical example is the components of variance model considered by Neyman and Scott (1948) where  $Y_i = (Y_{i1}, Y_{i2})$ ,  $Y_{ij} = U_i + \sigma \epsilon_{ij}$  and the  $\epsilon_{ij}$  are independent  $\mathbf{N}(0, 1)$ .  $\sigma^2$  is efficiently estimated by  $\frac{1}{2n} \sum_{i=1}^n (Y_{i1} - Y_{i2})^2$  but the problem of estimating  $G$  is equivalent to the normal deconvolution problem which permits best rates of  $(\log n)^{-1/2}$  – see Fan (1991) for example. In the IH models both  $G$  and  $\theta$  can be estimated at rate  $n^{-1/2}$ .

Recently, Vardi and Zhang (1992) focused on the special IH model (with  $\theta$  known) where  $U > 0$  and  $q(y|u) = u^{-1}1(y \leq u)$ , that is,  $Y = VU$  where  $V$  is nonnegative and  $U$  is a  $U(0, 1)$  random variable independent of  $V$ . They showed, without further conditions that the nonparametric maximum

likelihood estimate (NPMLE) of  $G$  is efficient. In Bickel, Klaassen, Ritov, Wellner (1993) (BKRW) we show how to construct information bounds for estimation of “smooth” functions of  $G$  and exhibit readily calculable efficient estimates for  $(U, Y) \in R^2$ , under very strong conditions on the kernel  $q(\cdot|\cdot)$  and the smoothness and support of  $G$ . Our estimate  $\hat{G}$  maximizes  $\lambda_n \int \log g(u) p_{1n}^*(u) + (1 - \lambda_n) \int \log q(y) p_{2n}^*(y) dy$  where  $p_{1n}^*, p_{2n}^*$  are estimates of the densities  $g, q$  based on the “good” and “bad” observations respectively.

Recently, van der Vaart (1992) showed that the NPMLE of  $G$  is efficient in IH models under conditions similar but weaker than those of BKRW. However, it is not hard to give examples of such models where the NPMLE does not exist or is inconsistent. For instance, suppose  $U = (U', U'') \in R \times R^+$  and the conditional distribution of  $Y$  given  $U$  is  $\mathbf{N}(U', U'')$ . In that case the NPMLE will be the empirical distribution of the  $U$  sample plus the empirical distribution of the  $(Y_i, 0)$ . (The “maximum” of the likelihood will be achieved if we assume that all of the  $Y$ 's came from  $U' = Y$  and  $U'' = 0$ .)

As a second example, consider the case where given  $U$ ,  $Y = U$  with probability  $\alpha$  and, with probability  $1 - \alpha$ ,  $Y$  follows some distribution with Lebesgue density  $q(\cdot|U)$ . In that case we get that the NPMLE is the joint empirical distribution of the  $Y$ 's and the  $U$ 's, which, of course, isn't consistent. (Again, the likelihood will be maximized if we assume that all the  $Y$  sample came from its empirical distribution, and nothing came from  $\int q(\cdot|u) dG(u)$ .)

Our goal in this paper is to construct computable efficient estimates of  $G$  under minimal assumptions. We do so by,

- (i) Limiting ourselves to distributions concentrating on the observed “good” sample ( $U$ 's). This is *not* true of the NPMLE.
- (ii) Binning the “bad” sample ( $Y$ 's),

and then applying maximum likelihood or a variant thereof. Specifically, let  $\mathcal{Y}_{n1}, \dots, \mathcal{Y}_{nJ_n}$  be partitions of the range of  $\mathcal{Y}$ , where  $\mathcal{Y}$  is the range of  $Y$  and  $J_n$  is  $o(n^{1/3})$  and

$$\underline{\lim}_n \min_j J_n Q(\mathcal{Y}_{nj}) > 0. \tag{1.2}$$

Let  $G_{bn}(A) = m^{-1} \sum_i (1 + b_i) 1(u_i \in A)$  for any set  $A \in \text{supp}(G)$  and vector  $b \in \mathbf{R}^m$ . Our estimator will be written in that form and hence our object of estimation is the vector  $b$ . Let  $\tilde{Q}(A) = (n - m)^{-1} \sum_i 1(\Delta_i = 0, Y_i \in A)$ ,  $Q_{nj}(\cdot) \equiv Q(\mathcal{Y}_{nj}|\cdot)$ ,  $Q^*(\mathcal{Y}_{nj}) = m^{-1} \sum_i Q_{nj}(u_i)$ . More generally, for  $b \in \mathbf{R}^m$

let  $Q^*(\mathcal{Y}_{nj}|b) = Q^*(\mathcal{Y}_{nj}) + m^{-1} \sum_i b_i Q_{nj}(u_i)$ , where the  $u_i$ 's are the set of "good" observations.

**Convention** Without loss of generality we take  $\{i : \Delta_i = 1\} = \{1, \dots, m\}$ ,  $\{i : \Delta_i = 0\} = \{m+1, \dots, n\}$ . Unless specified otherwise  $\sum_i = \sum_{i=1}^m$ .

If we maximize the likelihood of  $\{U_1, \dots, U_m, 1(Y_i \in \mathcal{Y}_{nj}) \mid 1 \leq i \leq n-m, 1 \leq j \leq J_n\}$  over  $G_{bn}$  as above we obtain the self consistency equation.

$$\hat{p}_i = \frac{1}{n} \left( 1 + \sum_j (n-m) \tilde{Q}(\mathcal{Y}_{nj}) \frac{\hat{p}_i Q_{nj}(u_i)}{\sum_k p_k Q_{nj}(u_i)} \right). \quad (1.3)$$

where  $\hat{p}_i \equiv m^{-1}(1 + \hat{b}_i)$ . In section 2 we will show that this equation has a unique solution. Rewrite (1.3) as

$$mp_i = \lambda_n + m(1 - \lambda_n) \sum_j \tilde{Q}(\mathcal{Y}_{nj}) \frac{p_i Q_{nj}(u_i)}{\sum_k p_k Q_{nj}(u_i)} \quad (1.4)$$

or

$$0 = \lambda_n - \lambda_n mp_i + mp_i(1 - \lambda_n) \sum_j \left[ \frac{\tilde{Q}(\mathcal{Y}_{nj})}{\sum_k p_k Q_{nj}(u_i)} - 1 \right] Q_{nj}(u_i). \quad (1.5)$$

Substitute  $p_i = m^{-1}(1 + b_i)$  in (1.5) and divide by  $1 + b_i$  to obtain that  $\hat{b}$  is the solution of the  $\mathcal{L}(b) = 0$  where

$$\mathcal{L}_n(b)_i \equiv \lambda_n \frac{b_i}{1 + b_i} - (1 - \lambda_n) \sum_j \left[ \frac{\tilde{Q}(\mathcal{Y}_{nj})}{Q(\mathcal{Y}_{nj}|b)} - 1 \right] Q_{nj}(u_i). \quad (1.6)$$

Since asymptotically the values of the  $b_i$ 's would be small we can ignore terms of smaller order and obtain the estimating equation  $\mathbb{W}_n(b) = 0$  where  $\mathbb{W}_n : R^m \rightarrow R^m$  is defined by

$$\mathbb{W}_n(b)_i = \lambda_n b_i - (1 - \lambda_n) \sum_j \left[ \frac{\tilde{Q}(\mathcal{Y}_{nj})}{Q(\mathcal{Y}_{nj}|b)} - 1 \right] Q_{nj}(u_i). \quad (1.7)$$

It will be more convenient to analyze first the solution  $\tilde{b}$  of  $\mathbb{W}_n(b) = 0$ . We will show that it is efficient. The efficiency of the solution of  $\mathcal{L}_n(b)$  will be shown to follow from that.

In fact neither  $\hat{b}$  nor  $\tilde{b}$  are proper estimates since condition (1.2) makes the choice of the partition depend on the unknown  $G$ . To avoid this we have to make the partition data determined. For instance, if  $Y \in R$  we can take  $\mathcal{Y}_{nj} = [y_{(jk_n)}, y_{((j+1)k_n)})$ ,  $1 \leq j \leq J_n - 2$ ,  $\mathcal{Y}_{n0} = (-\infty, y_{(k_n)}]$ ,  $\mathcal{Y}_{nJ_n-1} = [y_{((J_n-1)k_n)}, \infty)$  where

$$y_{(1)} < y_{(2)} \cdots < y_{(n-m)}$$

are the order statistics of  $Y_i$ , and  $k_n J_n (n - m)^{-1} \geq \delta > 0$ . We sketch a more general version of this construction and the appropriate theorem for the resulting estimates  $\hat{b}^*$ ,  $\tilde{b}^*$  in section 2.

This paper is organized as follows. In section 2 we review necessary and sufficient conditions for efficiency in these models as discussed in BKRW and state our main results. In section 4 we state and prove the lemmas needed to prove theorem 1 of section 2 and complete the proof of the theorem. We also discuss models in which the conditional distributions of  $Y|U$  are not dominated and observation are missing at random (MAR) in the terminology of Little and Rubin (1987), as well as the situation considered by Ibragimov and Hasminskii.

## 2 Main Results

We will need the following assumptions in addition to (1.2).

R0:  $\lambda_n \geq \epsilon > 0$  for all  $n$ .

R1: (1.2) holds and  $J_n = o(n^{1/5})$

R2:  $\max_j \{Q(\mathcal{Y}_{nj}) : \mathcal{Y}_{nj} \text{ not an atom}\} \rightarrow 0$

### *Existence and computation of $\hat{b}$*

We begin by establishing the existence of  $\hat{b}$  and exhibiting an algorithm for its computation which converges for data in a set whose probability converges to 1. The same arguments serve for existence and computation of  $\tilde{b}$  and the estimates  $\tilde{\tilde{b}}$ ,  $\hat{\hat{b}}$  defined later in this in section.

Define on  $R^m$  the inner product

$$\langle x, z \rangle \equiv \frac{1}{m} \sum_i x_i z_i. \tag{2.1}$$

Then the matrix derivative  $\dot{W}_n(b)$  of  $W_n$  is defined by

$$\begin{aligned}
& \langle x, \dot{W}_n(b)z \rangle \\
&= \frac{1}{m} \lambda_n \sum_i x_i z_i \\
&+ (1 - \lambda_n) \sum_j \frac{\tilde{Q}(\mathcal{Y}_{nj})}{Q^2(\mathcal{Y}_{nj}|b)} \left( \frac{1}{m} \sum_i Q_{nj}(u_i) x_i \right) \\
&\left( \frac{1}{m} \sum_i Q_{nj}(u_i) z_i \right). \tag{2.2}
\end{aligned}$$

It follows that, for all  $x$ ,

$$\langle x, \dot{W}_n(b)x \rangle \geq \lambda_n \|x\|^2. \tag{2.3}$$

From (2.3) we deduce that  $W_n$  is the gradient of a strictly convex function which tends to  $\infty$  as  $\|b\| \rightarrow \infty$ . We deduce,

**Proposition 1.**  $\tilde{b}$  exists and is unique.

It also follows from (2.3) that  $\dot{W}_n$  is invertible and, in fact, for all  $b$ ,

$$\|\dot{W}_n^{-1}(b)\| \leq \lambda_n^{-1} \tag{2.4}$$

where  $\|\cdot\|$  is the operator norm induced by  $\langle \cdot, \cdot \rangle$ . Let  $b^{(0)} = 0$  and solve iteratively for

$$b^{(k+1)} = b^{(k)} - \dot{W}_{n0}^{-1} \mathbb{W}_n(b^{(k)}), \quad k = 0, 1, \dots \tag{2.5}$$

where  $\dot{W}_{n0} = \dot{W}_n(0)$ .

**Remark.** The Newton-Raphson iteration is the modification of the RHS of (2.5) given by  $b^{(k)} - \dot{W}_n^{-1} \mathbb{W}_n(b^{(k)})$ . The algorithm (2.5) is easier to analyse.

**Proposition 2.** Under R1, R2,

$$P[b^{(k)} \rightarrow b^{(\infty)}] \rightarrow 1,$$

where  $b^{(\infty)} = \tilde{b}$ .

The proof is given in section 4.

*Asymptotic theory of estimates of linear functionals of  $G$*

For  $h \in L_2(G)$ , it is natural to consider  $\int h dG_{\hat{b}_n}$  and  $\int h dG_{\tilde{b}_n}$  as estimates of  $\theta_h \equiv \int h dG$ .

If  $h \in L_2(G)$  we shall find that the influence function of these estimates is given by  $\Psi_h$  defined as follows. For functions  $a \in L_2(Q)$  and  $b \in L_2(G)$ , let  $E_U, E_Y$  be the conditional expectation operators given  $U, Y$  respectively. That is,

$$\begin{aligned} E_U(a)(U) &= \int a(y)q(y|U)d\mu(y), \\ E_Y(b)(Y) &= \frac{\int b(u)q(Y|u)dG(u)}{\int q(Y|u)dG(u)}, \end{aligned}$$

almost surely.

Note that the operator  $E_U E_Y : L_2(G) - L_2(G)$  is self adjoint and Hermitian. That is, for any two functions  $a, b \in L_2(G)$ ,  $\langle a, E_U E_Y b \rangle = \langle a, E_Y b \rangle = \langle E_Y a, E_Y b \rangle = \langle E_Y a, b \rangle = \langle E_U E_Y a, b \rangle$ , where the inner product is that of  $L_2(P)$ . Hence  $\lambda_n I + (1 - \lambda_n) E_U E_Y$  is selfadjoint, with all eigenvalues not smaller than  $\lambda_n$  and hence boundedly invertible, (Kantorovich and Akilov (1982)). Let  $f_h = (\lambda_n I + (1 - \lambda_n) E_U E_Y)^{-1} h \in L_2(G)$  and

$$\Psi_h(X) = \Delta f_h(U) + (1 - \Delta) E_Y(f_h)(Y). \quad (2.6)$$

It may be shown, see BKRW section 6.5 for example, that  $\Psi_h - \theta_h$  is the efficient influence function for estimation of  $\theta_h$ . Thus, the asymptotic variance of any regular estimator of  $\theta_h$  is no smaller than

$$\begin{aligned} \text{Var}(\Psi_h) &= \lambda_n \langle f_h, f_h \rangle + (1 - \lambda_n) \langle E_Y f_h, E_Y f_h \rangle \\ &= \lambda_n \langle f_h, f_h \rangle + (1 - \lambda_n) \langle f_h, E_U E_Y f_h \rangle \\ &= \langle f_h, (\lambda_n I + (1 - \lambda_n) E_U E_Y) f_h \rangle \\ &= \langle f_h, h \rangle \\ &= E\{f_h(U)h(U)\}. \end{aligned} \quad (2.8)$$

Define further,  $E_Y^{(n)} h \equiv E\{h|1(Y \in \mathcal{Y}_{nj}), 1 \leq j \leq J_n\}$ ,  $f_h^{(n)} = (\lambda_n I + (1 - \lambda_n) E_U E_Y^{(n)})^{-1} h$  and  $\Psi_h^{(n)} = \Delta f_h^{(n)} + (1 - \Delta) E_Y^{(n)}(f_h^{(n)})(Y)$ .

Define

$$\tilde{V}_n(h) = n^{1/2} \left( \int h d\hat{G}_{\tilde{b}_n} - \theta_h \right)$$

and  $\hat{V}_n$  correspondingly for  $\hat{b}$ . We also want to study weak convergence of  $\tilde{V}_n, V_n$  viewed as processes on sets  $\mathcal{H} \subset L_2(G)$ . We say that  $\mathcal{H}$  is *regular* if the following conditions hold. Let  $\Delta\mathcal{H} = \{h_1 - h_2 : h_j \in \mathcal{H}, 1 \leq j \leq 2\}$ . Let  $D_n(x, \delta)$  be the largest  $k$  such that there exist  $h_1, \dots, h_k \in \Delta\mathcal{H}$ ,  $x^2 \leq \frac{1}{m} \sum_i h_j^2(u_i) \leq \delta^2$  and define  $D_n(x)$  similarly with  $\mathcal{H}$  replacing  $\Delta\mathcal{H}$  and  $\delta^2 = \delta_n^2 \equiv \sup_{\mathcal{H}} \frac{1}{m} \sum h^2(u_i)$ . We require that,

$$\delta_n = O_P(1)$$

$$\int_0^{\delta_n} \sqrt{\log D_n(x)} dx = O_P(1) \quad (2.9)$$

$$\lim_n P[\sup\{\frac{1}{m} \sum_i h^2(u_i) : h \in \Delta\mathcal{H}, \|h\|_G \leq \delta\} \geq 2\delta] = 0$$

and

$$\lim_{\delta \rightarrow 0} \limsup_n P[\int_0^{\delta} \sqrt{\log D_n(x, \delta)} dx \geq \epsilon] = 0 \quad (2.10)$$

for every  $\epsilon > 0$ .

Note that any finite subset of  $L_2(G)$  is regular.

**Theorem 1:** *Suppose R0 and R1 hold.*

(i) *Then,*

$$\int hdG_{\hat{b}_n} = n^{-1} \sum_i \Psi_h^{(n)}(X_i) + o_p(n^{-1/2}) \quad (2.11)$$

and

$$\int hdG_{\tilde{b}_n} = n^{-1} \sum_i \Psi_h^{(n)}(X_i) + o_p(n^{-1/2}) \quad (2.12)$$

(ii) *Suppose that R2 holds also. Then*

$$E(\Psi_h(X) - \Psi_h^{(n)}(X))^2 \rightarrow 0, \quad (2.13)$$

so that  $\Psi_h^{(n)}$  may be replaced by  $\Psi_h$  in (2.11), (2.12).

(iii) *If  $\mathcal{H}$  is regular (2.11) and (2.12) hold uniformly for  $h \in \mathcal{H}$  and  $\tilde{W}_n, \hat{W}_n$  converge weakly (in the sense of Dudley-Pollard) to a mean 0 Gaussian process on  $\mathcal{H}$  with covariance given by*

$$c(h_1, h_2) = \text{cov}(\Psi_{h_1}(X_1), \Psi_{h_2}(X_1)).$$



Loosely, weak convergence holds for  $\tilde{W}_n, \hat{W}_n$  if it holds for the empirical processes based on  $u_1, \dots, u_m$ .

We, then, immediately obtain from Theorem 1 (iii),

**Corollary 1:** *If  $U \in R$ ,  $G_{\hat{b}_n}, G$  refer to distribution functions and R0-R2 hold then  $\tilde{V}_n(t) \equiv n^{1/2}(G_{\hat{b}_n}(t) - G(t))$  converges weakly on  $D[-\infty, \infty]$ .*

The same claim holds for  $G_{\hat{b}_n}$ . In particular,  $G_{\hat{b}_n}$  has the same behaviour as the NPMLE in the Vardi-Zhang model.

The proof of theorem 1 is involved and depends on a number of lemmas stated and proved in section 4. Here is its structure.

1. We establish Lemma 4.1 permitting us to replace  $Q_n^*, \tilde{Q}_n$  by  $Q$  when needed.
2. We show that  $W_n(0)$  is of order  $J_n n^{-1/2}$  (Lemma 4.2) and that  $\dot{W}_n(b) \simeq \dot{W}_n(0)$  in  $J_n n^{-1/2}$  neighbourhoods of 0 (Lemma 4.3). From this,
3. We deduce that  $\tilde{b}^{(1)} = -\dot{W}_{n0}^{-1} W_n(0)$  is of order  $J_n n^{-1/2}$  and

$$\tilde{b} = \tilde{b}^{(1)} + o_p(n^{-1/2}) \quad (2.14)$$

in  $\|\cdot\|$ . (Lemma 4.4).

4. We approximate  $W_n(0)$  by  $\bar{W}_n(0)$ , an average of independent random vectors. (Lemma 4.5), so that,

$$\tilde{b} = -\dot{W}_{n0}^{-1} \bar{W}_n(0) + o_p(n^{-1/2}). \quad (2.15)$$

5. We approximate  $\dot{W}_{n0}$  by a deterministic operator  $\tau_n \dot{\mathcal{W}}_{n0}$  where  $\dot{\mathcal{W}}_{n0}$  maps a lifting of  $L_2(G)$  onto itself and  $\tau_n$  is the evaluation map  $\tau h \equiv (h(u_1), \dots, h(u_m))$ . This approximation is made uniformly in  $\mathcal{H}$ . (Lemmas 4.6 and 4.7).

Then (2.15) becomes

$$\tilde{b} = -\tau_n \dot{\mathcal{W}}_{n0}^{-1} \bar{W}_n(0) + o_p(n^{-1/2}). \quad (2.16)$$

Finally, we show that (2.16) implies (2.11) (uniformly on  $\mathcal{H}$ ).

6. We show that  $\|\tilde{b}\|_\infty \equiv \max_j |\tilde{b}_j| = o_p(1)$  (Lemma 4.8) and hence deduce that  $\|\hat{b}\|_\infty = o_p(1)$  (Lemma 4.9). An analysis of  $\hat{b}$  analogous to 1. to 5. above now yields (2.12) See BKRW section 7.6 for a general presentation of the structure of such proofs.

7. Finally we show that  $\tilde{V}_n, \hat{V}_n$  are tight for regular  $\mathcal{H}$  (Lemmas 3.10, 3.11). Since uniformity in (2.11) and (2.12) has already been shown, part (iii) of Theorem 1 follows.

The estimates  $\hat{b}$  and  $\tilde{b}$  are unsatisfactory insofar as, through condition R1, the partition defining them depends on knowledge of  $Q$ . It is natural to define, if  $k = 1$ ,  $\frac{n-m}{J_n}$  is an integer  $r$  and  $y_{(0)} < \dots < y_{(n-m-1)}$  are the ordered  $y$ 's,

$$\mathcal{Y}_{nj} = [y_{(r(j-1))}, y_{(rj)})$$

for  $j = 1, \dots, J_n$ . (We make the natural minor changes in definition if  $\frac{n-m}{J_n}$  is not an integer or  $Q$  is not continuous.) If  $k > 1$  we can proceed lexicographically. First, if  $j_n^k = J_n$ ,  $r \equiv \frac{n-m}{j_n^k}$ , divide according to the  $rj$ ,  $j = 1, \dots, j_n^k$  order statistics of the first coordinate of  $y$ . Then subdivide each  $k$  dimensional strip according to the  $r$  order statistics of the second coordinate and then continue in this fashion until  $j_n^k$  boxes each containing approximately  $\frac{n-m}{J_n}$  observations are obtained as the  $\mathcal{Y}_{nj}$ . It is not hard to show that if  $\dot{W}_{n0}^{-1}h$  is bounded, where  $\dot{W}_{n0}$  is the operator which depends on the random partition  $\{\mathcal{Y}_{nj}\}$ , then theorem 1 carries over. For a clean general result, we resort to sample splitting although this is clearly just a technical device. Use the construction given above to determine  $\{\mathcal{Y}_{nj}\}$  but base it on the order statistics of the first  $\ell_n$   $y$ 's where  $\ell_n = o(n)$ . Then, base  $\tilde{b}, \hat{b}$  on the  $m$  observed  $u$ 's and the last  $n - m - \ell_n$  observed  $y$ 's. We state,

**Theorem 2:** *If  $\mathcal{Y} \equiv R^k$  is Euclidean and  $\hat{b}, \tilde{b}$  are determined by data determined partitions as given above conclusions (i) - (iii) of theorem 1 apply to  $\hat{b}$  and  $\tilde{b}$ .*

The same type of argument as that given for theorem 1 applies save that everything is done conditionally on the  $\{\mathcal{Y}_{nj}\}$ .

### 3 Extensions, open problems

*Estimation of the variance of  $\hat{\theta}_h$*

To set confidence bounds on  $\theta_h$  or to do more general inference we need an estimate of the asymptotic variance of  $\sqrt{n}\hat{\theta}_h$ , viz  $\text{Var}(\Psi_h(X))$ . Here is a

simple estimate which can be shown to work. Define

$$\hat{\Psi}_h = \Delta h + (1 - \Delta) \hat{E}_Y^{(n)} \hat{f}_h$$

where

$$\hat{E}_Y^{(n)} g(U) \equiv \sum_j 1(Y \in \mathcal{Y}_{nj}) \int_{\mathcal{Y}_{nj}} g(u) q(y|u) dG_{\hat{b}_n}(u) / Q^*(\mathcal{Y}_{nj}|\hat{b}) \quad (3.1)$$

$$\hat{f}_h = (\lambda_n I + (1 - \lambda_n) E_U \hat{E}_Y^{(n)})^{-1} h \quad (3.2)$$

and

$$\hat{\sigma}_h^2 = \frac{1}{n} \sum_{i=1}^n (\hat{\Psi}_h(X_i) - \hat{\theta}_h)^2. \quad (3.3)$$

Then it is easy to see that under our conditions  $\hat{\sigma}_h^2 \xrightarrow{P} \text{Var}(\Psi_h(X))$

*Inference in semiparametric IH models*

If  $q(y|u) = q(y|u, \theta)$ ,  $\theta \in R$ , say, it is natural to consider estimating equations for  $\theta$  of the form

$$\frac{1}{n - m} \sum_{i=m+1}^n \psi(Y_i, \theta, \hat{G}_\theta) = 0 \quad (3.4)$$

where  $\hat{G}_\theta$  is the estimate  $G_{\hat{b}_n}$  for  $q = q(\cdot|\cdot, \theta)$  and

$$\int \psi(y, \theta, G) q(y|u, \theta) dG(u) d\mu(y) = 0.$$

For efficiency one needs to use

$$\psi(y, \theta, G) = E_{(\theta, G)} \left\{ \frac{\partial}{\partial \theta} \log q(Y|U, \theta) | Y \right\},$$

Simpler estimating equations can be constructed by replacing, in (4.1),  $\hat{G}_\theta$  by  $G_{0n}$ , the empirical of the  $u_i$  but  $\psi$  needs to be chosen properly to obtain efficiency. We would need to take  $\psi$  to be the efficient score function for  $\theta$  rather than merely the score function for  $G$  known as in (3.2). This is essentially the approach of Ibragimov and Has'minskii (1983). See BKRW Section 7.7 for a discussion of these issues and the kind of conditions which need to be checked for these approaches to work.

*The undominated case*

Nothing in our discussion actually uses the fact that the family  $\{Q(\cdot|u)\}$  is dominated by  $\mu$  since our estimates are based on  $G_{0n}$ ,  $\tilde{Q}$ ,  $Q^*$  which are still well defined. Examples of such models are those for which  $Y \equiv Y(U)$ . A subclass of these models for which  $G$  is identifiable are those in which observations are missing at random (MAR) in the sense of Little and Rubin (1986), that is,  $P[\Delta = 1|U] = p(Y(u))$ . Let  $G_{0n}$  be the empirical of  $u_1, \dots, u_n$ ,  $N_j^{(1)} = \sum_{i=1}^m 1(Y_i \in \mathcal{Y}_{nj})$ ,  $N_j = \sum_{i=1}^n 1(Y_i \in \mathcal{Y}_{nj})$ , and  $j(u)$  be defined by  $u \in \mathcal{Y}_{nj(u)}$ . Then an explicit efficient estimate of  $G$  which concentrates on  $\{u_1, \dots, u_m\}$  is given by

$$d\hat{G}(u) = \frac{m}{n} dG_{0n}(u) \frac{N_{j(u)}}{N_j^{(1)}} \quad (3.5)$$

This is the maximum likelihood estimate if  $U$  is discrete. It is not hard to show that this estimate is efficient.

*What happens if  $\lambda_n \rightarrow 0$  with  $n$ ?*

This situation is of some interest since the fraction of good observations may be very small. We give a heuristic discussion of this situation. Rigorous treatment remains an open problem. If  $\lambda_n \equiv 0$  we know by BKRW section 6.5 that typically only  $\theta_h$  such that

$$h(u) = E(v(Y)|U) \quad (3.6)$$

can be estimated at rate  $n^{-1/2}$  by the obvious efficient estimate  $\hat{\theta}_n \equiv n^{-1} \sum_{i=1}^n v(Y_i)$ . In general the following heuristics suggest that the  $n(1 - \lambda_n)$  “bad” observations may still be of some use in estimation of  $\theta_h$  for  $h$  not obeying (3.6). Let  $\Psi_h$  be the efficient influence function for  $\theta_h$  given by (2.6) with  $\lambda_n \rightarrow \lambda \in (0, 1]$ . Note that  $E_U E_Y$  is a Hermitian operator on  $L_2(G)$ . By the spectral theorem,

$$E_U E_Y = \int_0^\infty \sigma I(\sigma) d\nu(\sigma)$$

where  $I(\sigma) d\nu(\sigma)$  is the spectral measure (e.g. Kantorovich and Akilov (1982)) and  $I(\sigma)$  is a projection on some subspace. Then, by (2.11)

$$\begin{aligned} \text{Var}(\Psi_h) &= ((\lambda I + (1 - \lambda) E_U E_Y)^{-1}(h), h)G \\ &= \int \frac{\|I(\sigma)h\|_G^2}{\lambda + (1 - \lambda)\sigma} d\nu(\sigma) \end{aligned} \quad (3.7)$$

where  $\|\cdot\|_G, (\cdot, \cdot)_G$  are the norm and inner product in  $L_2(G)$ . Then

$$\lambda \text{Var}(\Psi_h) \rightarrow \|I(0)h\|_G^2 d\nu(0) \quad \text{as } \lambda \rightarrow 0. \quad (3.8)$$

This suggests that if  $\|I(0)h\|_G > 0$  then estimation of  $\theta_h$  can only be achieved at rate  $m^{-1/2}$ . On the other hand, if  $I(0)h = 0$ , we may be able to obtain a better rate than  $m^{-1/2}$ !

However, in general, the  $\sqrt{n}$  rate cannot be saved. By (3.7),  $\lim_{\lambda \rightarrow 0} \text{Var}(\Psi_h) < \infty$  iff

$$I(0)h d\nu(0) = 0 \quad (3.9)$$

and

$$\int_{\sigma > 0} \frac{\|I(\sigma)h\|_G^2}{\sigma} d\nu(\sigma) < \infty. \quad (3.10)$$

It may be shown using Proposition A.1.6, part B, of BKRW that (3.9) and (3.10) hold iff there exists  $v(Y) \in L_2(P)$  such that  $h = E_U v(Y)$ . But this of course means that  $\theta_h$  is estimable at rate  $n^{-1/2}$  using the  $Y$ 's only, by  $(n-m)^{-1} \sum v(Y_j)(1 - \Delta_j)$ .

*The choice of  $\{\mathcal{Y}_{nj}\}, \{J_n\}$*

The choice of partition is a matter of convenience. However, selecting  $J_n$  can, to some extent, be made data dependent, for instance, by cross validation. Choose  $J_n$  which minimizes  $\sum (\hat{\theta}_{-iJ} - \bar{\theta}_J)^2$  where  $\hat{\theta}_{-iJ}$  is the estimate of  $\theta_h$  based on the partition  $\{\mathcal{Y}_{nj} : 1 \leq j \leq J\}$  and  $X_k, k \neq i$   $\bar{\theta}_J = \frac{1}{n} \sum_{k=1}^n \hat{\theta}_{-iJ}$ . Whether such a choice can be made which is consistent with (1.2) and works well is an open problem. Jin (1990) successfully carried through such a program for estimation of location and regression.

## 4 Lemmas for and proofs of theorems of section 2

The first lemma establishes the control which using coarse partition gives us.

**Lemma 4.1.** *For any partitions satisfying (1.2),*

$$\max_j \left| \frac{\tilde{Q}(\mathcal{Y}_{nj})}{Q(\mathcal{Y}_{nj})} - 1 \right| = O_P(n^{-1/2} J_n^{1/2} \log(J_n)) \quad (4.1)$$

$$\max_j \left| \frac{Q^*(\mathcal{Y}_{nj})}{Q(\mathcal{Y}_{nj})} - 1 \right| = O_P(n^{-1/2} J_n^{1/2} \log(J_n)) \quad (4.2)$$

**Proof.** Both  $\tilde{Q}_n$  and  $Q_n^*$  are averages of i.i.d. variables bounded by 1 and  $\text{Var}(\tilde{Q}_n) = n^{-1}Q(1-Q) \geq \text{Var}(Q_n^*)$ . A standard argument with Hoeffding's (1962) inequality completes the proof.  $\square$

**Lemma 4.2.** *Under R1,*

$$\|W_n(0)\| = O_P(J_n^{1/2} n^{-1/2}). \quad (4.3)$$

**Proof.** We compute,

$$\|W_n(0)\|^2 = \frac{1}{m} \sum_{i=1}^m \left\{ \sum_{j=1}^{J_n} \frac{(\tilde{Q}(\mathcal{Y}_{nj}) - Q^*(\mathcal{Y}_{nj}))}{Q^*(\mathcal{Y}_{nj})} Q_{nj}(u_i) \right\}^2. \quad (4.4)$$

It is enough to show that,

$$\begin{aligned} I &\equiv \frac{1}{m} \sum_{i=1}^m \left\{ \sum_{j=1}^{J_n} \frac{(\tilde{Q}(\mathcal{Y}_{nj}) - Q(\mathcal{Y}_{nj}))}{Q(\mathcal{Y}_{nj})} Q_{nj}(u_i) \right\}^2 \\ &= O_P(n^{-1} J_n) \end{aligned} \quad (4.5)$$

$$\begin{aligned} II &\equiv \frac{1}{m} \sum_{i=1}^m \left\{ \sum_{j=1}^{J_n} \frac{(\tilde{Q}^*(\mathcal{Y}_{nj}) - Q(\mathcal{Y}_{nj}))}{Q(\mathcal{Y}_{nj})} Q_{nj}(u_i) \right\}^2 \\ &= O_P(n^{-1} J_n) \end{aligned} \quad (4.6)$$

and

$$\begin{aligned} III &\equiv \frac{1}{m} \sum_{i=1}^m \left\{ \sum_{j=1}^{J_n} \left\{ \frac{(\tilde{Q}(\mathcal{Y}_{nj}) - Q(\mathcal{Y}_{nj}))(Q^*(\mathcal{Y}_{nj}) - Q(\mathcal{Y}_{nj}))}{Q(\mathcal{Y}_{nj})Q^*(\mathcal{Y}_{nj})} \right\} Q_{nj}(u_i) \right\}^2 \\ &= O_P(n^{-1} J_n). \end{aligned} \quad (4.7)$$

Now, after some algebra,

$$\begin{aligned}
E(I) &= E\left\{\sum_{j=1}^{J_n} (n-m)^{-1} \sum_{t=1}^{n-m} \frac{(1(Y_t \in \mathcal{Y}_{nj}) - Q(\mathcal{Y}_{nj}))}{Q(\mathcal{Y}_{nj})} Q_{nj}^2(u_1)\right\}^2 \\
&= (n-m)^{-1} E\left(\sum_{j=1}^{J_n} (1(y_1 \in \mathcal{Y}_{nj}) - Q(\mathcal{Y}_{nj})) \frac{Q_{nj}^2(u_1)}{Q(\mathcal{Y}_{nj})}\right)^2 \\
&\leq (n-m)^{-1} \sum_{j=1}^{J_n} \frac{E Q_{nj}^2(u_1)}{Q(\mathcal{Y}_{nj})}
\end{aligned} \tag{4.8}$$

Since  $E Q_{nj}^2(u_1) \leq Q(\mathcal{Y}_{nj})$ , (4.5) follows. Similarly,

$$\begin{aligned}
E(II) &= E\left\{\sum_{j=1}^{J_n} \frac{1}{m} \sum_{k=1}^m \frac{(Q_{nj}(u_i) - Q(\mathcal{Y}_{nj}))}{Q(\mathcal{Y}_{nj})} Q_{nj}^2(u_1)\right\}^2 \\
&= \frac{2}{m^2} [E\left\{\sum_{j=1}^{J_n} (Q_{nj}^2(u_1) - Q(\mathcal{Y}_{nj})) \frac{Q_{nj}^2(u_1)}{Q(\mathcal{Y}_{nj})}\right\}^2 \\
&\quad + (m-1) E\left\{\sum_{j=1}^{J_n} \frac{(Q_{nj}(u_2) - Q(\mathcal{Y}_{nj}))}{Q(\mathcal{Y}_{nj})} Q_{nj}(u_1)\right\}^2] \\
&= \frac{1}{m^2} (A + B) \text{ (say)}.
\end{aligned} \tag{4.9}$$

Now

$$\begin{aligned}
A &\leq E\left\{\sum_{j=1}^{J_n} \frac{Q_{nj}^2(u_1)}{Q(\mathcal{Y}_{nj})}\right\}^2 + 1 \\
&= O(\min_j Q^{-2}(\mathcal{Y}_{nj}) + 1) \\
&= O(J_n^2),
\end{aligned} \tag{4.10}$$

by (1.2). On the other hand,

$$\begin{aligned}
B &= (m-1) E\left\{\sum_{j=1}^{J_n} \frac{Q_{nj}(u_2) Q_{nj}(u_1)}{Q(\mathcal{Y}_{nj})} - 1\right\}^2 \\
&= (m-1) E\left\{\sum_{j=1}^{J_n} \frac{Q_{nj}(u_2) Q_{nj}(u_1)}{Q(\mathcal{Y}_{nj})}\right\}^2 - (m-1)
\end{aligned}$$

$$\begin{aligned}
&\leq (m-1) \sum_{j=1}^{J_n} \sum_{k=1}^{J_n} \frac{\{E(Q_{nj}(u_1)Q_{nk}(u_1))\}^2}{Q(\mathcal{Y}_{nj})Q(\mathcal{Y}_{nk})} \\
&\leq (m-1) \sum_{j=1}^{J_n} \sum_{k=1}^{J_n} \frac{EQ_{nj}(u_1)E(Q_{nj}(u_1)Q_{nk}^2(u_1))}{Q(\mathcal{Y}_{nj})Q(\mathcal{Y}_{nk})} \\
&= (m-1) \sum_{k=1}^{J_n} \frac{EQ_{nk}(u_1)}{Q(\mathcal{Y}_{nk})} \\
&\leq (m-1)J_n
\end{aligned} \tag{4.11}$$

and (4.6) follows. Finally, by lemma 4.1,

$$\begin{aligned}
III &= O_P((\max_j |\frac{Q_n^*(\mathcal{Y}_{nj})}{Q(\mathcal{Y}_{nj})} - 1|^2 \max_j |\frac{\tilde{Q}_n(\mathcal{Y}_{nj})}{Q(\mathcal{Y}_{nj})} - 1|^2) = O_P(n^{-2}J_n^2 \log^4 J_n) \\
&= O_P(n^{-1}J_n)
\end{aligned} \tag{4.12}$$

and (4.7) follows. The lemma follows from (4.5)-(4.7).  $\square$

**Lemma 4.3.** For all  $M < \infty$ ,

$$P(\inf\{\frac{Q(\mathcal{Y}_{nj}|b)}{Q(\mathcal{Y}_{nj})} : \|b\| < MJ_n^{1/2}n^{-1/2}\} < \frac{1}{2}) = o(1) \tag{4.13}$$

and

$$\sup\{\|\dot{W}_n(b) - \dot{W}_{n0}\| : \|b\| < MJ_n^{1/2}n^{-1/2}\} = O_P(J_n^{1/2}n^{-1/2}) \tag{4.14}$$

**Proof.** First note that if  $\|b\| < J_n^{1/2}Mn^{-1/2}$  then

$$\begin{aligned}
\frac{1}{m} \sum_i Q_{nj}(u_i)b_i &\leq \left( \|b\|^2 \frac{1}{m} \sum_i Q_{nj}^2(u_i) \right)^{1/2} \\
&\leq MJ_n n^{-1/2} Q^{*1/2}(\mathcal{Y}_{nj}) \\
&= Q(\mathcal{Y}_{nj}) O_P(J_n^{1/2}n^{-1/2})
\end{aligned} \tag{4.15}$$

and (4.13) follows.



For any vector  $x \in R^m$ :

$$\begin{aligned}
& \langle x, (\dot{\mathbb{W}}_n(b) - \dot{\mathbb{W}}(0))x \rangle \tag{4.16} \\
&= \sum_j \tilde{Q}(\mathcal{Y}_{nj}) \left( \frac{1}{Q^2(\mathcal{Y}_{nj}|b)} - \frac{1}{Q^2(\mathcal{Y}_{nj}|0)} \right) \left( \frac{1}{m} \sum_i Q_{nj}(u_i)x_i \right)^2 \\
&= -2 \sum_j \tilde{Q}(\mathcal{Y}_{nj}) \frac{m^{-1} \sum_i Q_{nj}(u_i)b_i}{Q^3(\mathcal{Y}_{nj}|b^*)} \left( \frac{1}{m} \sum_i Q_{nj}(u_i)x_i \right)^2
\end{aligned}$$

where  $b^* = tb$  for some  $t \in [0, 1]$ . Then, from (2.2) and lemma 4.1,

$$\begin{aligned}
& |\langle x, (\dot{\mathbb{W}}_n(b) - \dot{\mathbb{W}}(0))x \rangle| \\
&\leq \left( \sum_j \tilde{Q}(\mathcal{Y}_{nj}) \|b\| \left( \frac{1}{m} \sum_i Q_{nj}^2(u_i) \right)^{1/2} Q^{-1}(\mathcal{Y}_{nj}|b) \right) \\
&\quad \left( \frac{1}{mQ^*(\mathcal{Y}_{nj})} \sum_i Q_{nj}(u_i)x_i \right)^2 (1 + o(1)) \\
&\leq \sum_j \tilde{Q}(\mathcal{Y}_{nj}) \|b\| Q^{*-1/2}(\mathcal{Y}_{nj}) \frac{1}{mQ^*(\mathcal{Y}_{nj})} \sum_i Q_{nj}(u_i)x_i^2 \\
&= O_P(J_n^{3/2} \|b\| \|x\|^2). \tag{4.17}
\end{aligned}$$

by R1. Therefore, we conclude that

$$\begin{aligned}
& \sup\{\|\dot{W}_n(b) - \dot{W}_n(0)\| : \|b\| \leq MJ_n^{1/2}n^{-1/2}\} \\
&= O_P(J_n^2 n^{-1/2})
\end{aligned}$$

under R1, and the lemma follows.  $\square$

Let

$$T(b) = b - \dot{W}_{n0}^{-1}W_n(b). \tag{4.18}$$

**Lemma 4.4.** *Under R0, R1, for all  $M < \infty$ ,  $P[T$  is a contraction on  $\{b : \|b\| \leq MJ_n^{1/2}n^{-1/2}\} \rightarrow 1$ . Further,*

$$\sup_k \|b^{(k)}\| = O_P(J_n^{1/2}n^{-1/2}) \tag{4.19}$$

$$\sup_{k \geq 2} \|b^{(k)} - b^{(1)}\| = o_p(n^{-1/2}) \quad (4.20)$$

Note that, since  $b^{(k+1)} = T(b^{(k)})$  by definition, (2.6) and the claim that  $T$  is a contraction with probability tending to 1 establish proposition 2.

**Proof.** Write

$$\begin{aligned} \dot{T}(b) &= I - \dot{W}_{n0}^{-1} \dot{W}_n(b) \\ &= \dot{W}_{n0}^{-1} (\dot{W}_{n0} - \dot{W}_n(b)). \end{aligned} \quad (4.21)$$

It follows from (2.4) and lemma 4.3 that for all  $M$ ,

$$\zeta_n \equiv \sup\{\|\dot{T}(b)\| : \|b\| \leq MJ_n^{1/2}n^{-1/2}\} = O_P(J^2n^{-1/2}). \quad (4.22)$$

Further, by definition,

$$\begin{aligned} \|b^{(1)}\| &\leq \|\dot{W}_{n0}^{-1}\| \|W_n(0)\| \\ &= O_p(J_n^{1/2}n^{-1/2}) \end{aligned} \quad (4.23)$$

by lemma 4.2. By the usual induction argument

$$\|b^{(k)}\| \leq (1 - \xi_n)^{-1} \|b^{(1)}\|$$

and

$$\|b^{(k+1)} - b^{(k)}\| \leq \xi_n^k \|b^{(1)}\|.$$

Then, the lemma follows from (4.22) and (4.23).  $\square$

Define,

$$\bar{W}_n(0)_i \equiv -(1 - \lambda_n) \sum_j \frac{(\tilde{Q}(\mathcal{Y}_{nj}) - Q^*(\mathcal{Y}_{nj}))}{Q(\mathcal{Y}_{nj})} Q_{nj}(u_i) \quad (4.24)$$

and the evaluation map  $\tau_n : L_2(G) \rightarrow R^m$  by,

$$\tau_n h = (h(u), \dots, h(u_m))$$

(where we suppose that a representative  $h$  has been chosen in each equivalence class belonging to  $L_2(G)$ , a ‘‘lifting’’).

**Lemma 4.5.** *Under R0, R1*

$$\|W_n(0) - \bar{W}_n(0)\| = o_p(n^{-1/2}). \quad (4.25)$$

**Proof.** By definition

$$\begin{aligned} & \|W_n(0) - \bar{W}_n(0)\|^2 = \\ & m^{-1} \sum_i (1 - \lambda_n)^2 \left\{ \sum_j (\tilde{Q}(\mathcal{Y}_{nj}) - Q^*(\mathcal{Y}_{nj})) \frac{(Q(\mathcal{Y}_{nj}) - Q^*(\mathcal{Y}_{nj}))}{Q(\mathcal{Y}_{nj})Q^*(\mathcal{Y}_{nj})} Q_{nj}(u_i) \right\}^2. \end{aligned} \quad (4.26)$$

By lemma 4.1, (4.26) is  $O_P(n^{-2} J_n^2 \log^4 J_n) = o_p(n^{-1})$ .  $\square$

Define the operator  $\dot{W}_{n0} : L_2(G) \rightarrow L_2(G)$  by

$$\begin{aligned} (h_2, \dot{W}_{n0} h_1) &= \lambda_n \int h_1(u) h_2(u) dG(u) \\ &+ (1 - \lambda_n) \sum_j \frac{1}{Q(\mathcal{Y}_{nj})} \int Q_{nj}(u) h_2(u) dG(u) \int Q_{nj}(u) h_1(u) dG(u) \end{aligned} \quad (4.27)$$

where  $(\cdot, \cdot)_G$  is the inner product on  $L_2(G)$ . Note that  $\dot{W}_{n0}$  is invertible and,  $\|\dot{W}_{n0}^{-1}\|_G \leq \lambda_n^{-1}$ . Let  $\tilde{h}_n = \dot{W}_{n0}^{-1} h$ . For  $v \in L_2(G)$  let  $v_c \equiv v - Ev$ .

**Lemma 4.6.** *Under R0, R1, R3,*

$$\sup_{\mathcal{H}} \sum_j \left( \frac{1}{n} \sum_i (Q_{nj} \tilde{h}_n)_c(u_i) \right)^2 = O_P(J_n \log(J_n) n^{-1}). \quad (4.28)$$

**Proof.** Let  $v_{n1}, \dots, v_{nJ_n}$  be an orthonormal basis for  $V_k \equiv \text{span}\{Q_{nj}(\cdot), j = 1, \dots, J_n\}$ . For any  $h \in L_2(G)$  let  $h = h_1 + h_2$  where  $h_1 \in V_n$  and  $h_2 \perp V_n$ . Accordingly let  $\tilde{h}_{n1} = \dot{W}_n^{-1} h_1$  and  $\tilde{h}_{n2} = \dot{W}_n^{-1} h_2$ .

Now,  $\|\tilde{h}_{n1}\|_G \leq \lambda_n^{-1} \|h_1\|_G \leq c \equiv \sup_{h \in \mathcal{H}} \|h\|_G$ , and  $\tilde{h}_{n1} \in V_n$ . Then

$$E \sup_{\mathcal{H}} \sum_j \left( \frac{1}{m} \sum_i (Q_{nj} \tilde{h}_{n1})_c(u_i) \right)^2 \leq$$

$$\begin{aligned}
&\leq E \sup \left\{ \sum_j \left( \sum_\ell d_\ell \frac{1}{m} \sum_i (Q_{nj} v_{n\ell})_c(u_i) \right)^2 : \sum_\ell d_\ell^2 \leq d^2 \right\} \\
&\leq d^2 E \sum_j \sum_\ell \left( \frac{1}{m} \sum_{i=1}^m (Q_{nj} v_{n\ell})_c(u_i) \right)^2 \\
&\leq \frac{d^2}{m} \sum_j \sum_\ell E Q_{nj}^2 v_{n\ell}^2(u_1) \\
&\leq \frac{d^2}{m} \sum_\ell E v_{n\ell}^2(u_1) = \frac{d^2 J_n}{m}. \tag{4.29}
\end{aligned}$$

On the other hand, by the spectral theorem since  $E_U E_Y^{(n)}$  is a compact Hermitian operator  $\tilde{h}_{n2} = \int \frac{1}{\gamma} dE(\gamma)$  (h) where  $\lambda_n \leq \gamma \leq 1$  and  $E\gamma$  is a projection on a finite dimensional subspace of the range of  $E_U^{(n)} E_Y$  (which is contained in  $\mathcal{V}_n$ ). Therefore,  $\tilde{h}_{n2} = \lambda_n^{-1} h_2$ . If  $\mathcal{H}$  satisfies R3 so does each of the sets  $\mathcal{H}_j \equiv \{\lambda_n^{-1} Q_{nj} h(\cdot) : h \in \mathcal{H}\}$ . Therefore we can apply theorem (3.5) in Pollard (1990) to obtain,

$$\begin{aligned}
P \left[ \sup_{\mathcal{H}} \left| \frac{1}{m} \sum_i (Q_{nj} \tilde{h}_{n2})_c(u_i) \right| \geq M \left( \frac{\log J_n}{n} \right)^{1/2} \right] \\
\leq K_1 J_n^{-\lambda(M)}
\end{aligned}$$

where  $\lambda(M) \rightarrow \infty$  as  $M \rightarrow \infty$ . Hence

$$P \left[ \max_j \sup_{\mathcal{H}} \left| \frac{1}{m} \sum_i (Q_{nj} \tilde{h}_{n2})_c(u_i) \right| \geq M n^{-1/2} \right] = o(1). \tag{4.30}$$

Combining (4.29) and (4.30) the lemma follows.

**Lemma 4.7.** *If (1.2), R0, R1 hold and  $\mathcal{H}$  is regular,*

$$\sup_{\mathcal{H}} \|\dot{W}_{n0}^{-1} \tau_n h - \tau_n \dot{W}_{n0}^{-1} h\| = O_P(J_n^2 \log(J_n) n^{-1/2}). \tag{4.31}$$

**Proof.** For any  $v \in R^m$ ,

$$\begin{aligned}
&\langle v, (\dot{W}_{n0}^{-1} \tau_n - \tau_n \dot{W}_{n0}^{-1}) h \rangle \\
&= \langle v, \dot{W}_{n0}^{-1} (\tau_n \dot{W}_{n0} - \dot{W}_{n0} \tau_n) \dot{W}_{n0}^{-1} h \rangle \\
&= \langle \dot{W}_{n0}^{-1} v, (\tau_n \dot{W}_{n0} - \dot{W}_{n0} \tau_n) \dot{W}_{n0}^{-1} h \rangle. \tag{4.32}
\end{aligned}$$

Since further  $\|\dot{W}_{n0}^{-1}\| \leq \lambda_n$ , it is enough to show,

$$\sup_{\mathcal{H}} \|(\tau_n \dot{W}_{n0} - \dot{W}_{n0} \tau_n) \dot{W}_{n0}^{-1} h\| = O_P(J_n^2 \log J_n n^{-1/2}). \quad (4.33)$$

For any  $v \in R^m$ ,

$$\begin{aligned} & \langle v, (\tau_n \dot{W}_{n0} - \dot{W}_{n0} \tau_n) \tilde{h}_n \rangle \\ &= \sum_j \frac{1}{Q(\mathcal{Y}_{nj})} \left( \int Q_{nj}(u) \tilde{h}_n(u) dG(u) \right) \left( \frac{1}{m} \sum_i Q_{nj}(u_i) v_i \right) \\ & \quad - \sum_j \frac{\tilde{Q}(\mathcal{Y}_{nj})}{[Q^*]^2(\mathcal{Y}_{nj})} \left( \frac{1}{m} \sum_i Q_{nj}(u_i) \tilde{h}_n(u_i) \right) \left( \frac{1}{m} \sum_i Q_{nj}(u_i) v_i \right) \\ &= \sum_j \left( \frac{1}{Q(\mathcal{Y}_{nj})} - \frac{\tilde{Q}(\mathcal{Y}_{nj})}{[Q^*]^2(\mathcal{Y}_{nj})} \right) \left( \int Q_{nj}(u) \tilde{h}_n(u) dG(u) \right) \left( \frac{1}{m} \sum_i Q_{nj}(u_i) v_i \right) \\ & \quad - \sum_j \frac{\tilde{Q}(\mathcal{Y}_{nj})}{[Q^*]^2(\mathcal{Y}_{nj})} \left( \frac{1}{m} \sum_i Q_{nj}(u_i) v_i \right) \left( \frac{1}{m} \sum_i (Q_{nj} \tilde{h}_n(u_i) - \int Q_{nj}(u) \tilde{h}_n(u) dG(u)) \right) \\ & \quad = A_1(v) - A_2(v) \quad (\text{say}) \end{aligned} \quad (4.34)$$

By lemma 4.1,

$$\max_j \left| \frac{1}{Q(\mathcal{Y}_{nj})} - \frac{\tilde{Q}(\mathcal{Y}_{nj})}{[Q^*]^2(\mathcal{Y}_{nj})} \right| = O_P(Q^{-1}(\mathcal{Y}_{nj}) n^{-1/2} J_n^{1/2} \log(J_n)). \quad (4.35)$$

Further, by Cauchy-Schwartz,

$$\left| \frac{1}{m} \sum_i Q_{nj}(u_i) v_i \right| \leq \|v\| \left( \frac{1}{m} \sum_i Q_{nj}^2(u_i) \right)^{1/2} \|v\| Q^{*1/2}(\mathcal{Y}_{nj}). \quad (4.36)$$

Thus, by (1.2)

$$\sup_v \frac{|A_1(v)|}{\|v\|} = O_P(J_n n^{-1/2} \log(J_n)) \left( \int \tilde{h}_n^2(u) dG(u) \right)^{1/2}. \quad (4.37)$$

On the other hand

$$|A_2(v)| \leq \left( \sum_j \frac{\tilde{Q}^2(\mathcal{Y}_{nj})}{[Q^*]^4(\mathcal{Y}_{nj})} \left( \frac{1}{m} \sum_i Q_{nj}(u_i) v_i \right)^2 \right)^{1/2} \left( \sum_j \left( \frac{1}{m} \sum_i (Q_{nj} \tilde{h}_n)_c(u_i) \right)^2 \right) \quad (4.38)$$

The first term in (4.38) is  $O_P(J_n \|v\|)$  by lemma 4.1 and (4.35), while the second is  $O_P(J_n^{1/2} \log J_n n^{-1/2})$  by lemma 4.5.

Therefore,

$$\sup_v \frac{|A_2(v)|}{\|v\|} = O_P(J_n^{3/2} \log J_n n^{-1/2}). \quad (4.39)$$

Combining (4.37) and (4.39) we obtain the lemma.  $\square$

**Proof of (2.11):** We prove (2.11) as extended in Theorem 1 (iii). Note that by lemma 4.4,

$$\begin{aligned} \langle \tau_n h, \tilde{b} \rangle &= \langle \tau_n h, b^{(1)} \rangle + o_p(n^{-1/2}) \\ &= -\langle \tau_n h, \dot{W}_{n0}^{-1} W_n(0) \rangle + o_p(n^{-1/2}). \end{aligned} \quad (4.40)$$

uniformly on  $\mathcal{H}$  since  $\sup \|\tau_{nh}\| = O_P(1)$  by regularity of  $\mathcal{H}$ . Further, lemma 4.5 and  $\|\dot{W}_n^{-1}\| \leq \lambda_n^{-1}$  imply

$$-\langle \tau_n h, \dot{W}_{n0}^{-1} W_n(0) \rangle = -\langle \dot{W}_{n0}^{-1} \tau_n h, \bar{W}_n(0) \rangle + o_p(n^{-1/2}) \quad (4.41)$$

uniformly in  $\mathcal{H}$ . Apply lemmas 4.2, 4.5 and 4.7 to get from (4.40) and (4.41).

$$\begin{aligned} \langle \tau_n h, \tilde{b} \rangle &= -\langle \tau_n \tilde{h}_n, \bar{W}_n(0) \rangle + O_P(J_n^2 \log J_n n^{-1}) \\ &= (1 - \lambda_n) \sum_j \frac{(\tilde{Q}(\mathcal{Y}_{nj}) - Q^*(\mathcal{Y}_{nj}))}{Q(\mathcal{Y}_{nj})} \frac{1}{m} \sum_i Q_{nj}(u_i) \tilde{h}_n(u_i) + o_p(n^{-1/2}) \end{aligned} \quad (4.42)$$

uniformly in  $\mathcal{H}$ . Further, by lemma 4.6

$$\begin{aligned} &\sum_j \frac{(\tilde{Q}(\mathcal{Y}_{nj}) - Q^*(\mathcal{Y}_{nj}))}{Q(\mathcal{Y}_{nj})} \frac{1}{m} \sum_i Q_{nj}(u_i) \tilde{h}_n(u_i) \\ &= \sum_j \frac{(\tilde{Q}(\mathcal{Y}_{nj}) - Q^*(\mathcal{Y}_{nj}))}{Q(\mathcal{Y}_{nj})} \int Q_{nj}(u) \tilde{h}_n(u) dG(u) \\ &+ O_P \left( \left( \sum_j \frac{(\tilde{Q}(\mathcal{Y}_{nj}) - Q^*(\mathcal{Y}_{nj}))^2}{Q^2(\mathcal{Y}_{nj})} \right)^{1/2} \left( \sum_j \left( \frac{1}{m} \sum_i (Q_{nj}(u) \tilde{h}_n(u))_c(u_i) \right)^2 \right)^{1/2} \right) \\ &= - \sum_j \frac{(\tilde{Q}(\mathcal{Y}_{nj}) - Q^*(\mathcal{Y}_{nj}))}{Q(\mathcal{Y}_{nj})} \int Q_{nj}(u) \tilde{h}_n(u) dG(u) \end{aligned}$$

$$+O_P(n^{-1}J_n^{3/2}\log(J_n)), \quad (4.43)$$

uniformly on  $\mathcal{H}$ , by lemmas 4.1 and 4.6.

Note that

$$\dot{W}_{n0} = \lambda_n I + (1 - \lambda_n) E_U E_Y^{(n)}$$

so that

$$h - \lambda_n \tilde{h}_n = (1 - \lambda_n) E_U E_Y^{(n)} \tilde{h}_n \quad (4.44)$$

Therefore, by definition,

$$\begin{aligned} & (1 - \lambda_n) \sum_j \frac{Q^*(\mathcal{Y}_{nj})}{Q(\mathcal{Y}_{nj})} \int Q_{nj}(u) f_h^{(n)}(u) dG(u) \\ &= \frac{1 - \lambda_n}{m} \sum_i Q_{nj}(u_i) \frac{1}{Q(\mathcal{Y}_{nj})} \int Q_{nj}(u) f_h^{(n)}(u) dG(u) \\ &= \frac{(1 - \lambda_n)}{m} \sum_i E_U E_Y^{(n)}(f_h^{(n)})(u_i) \\ &= \frac{1}{m} \sum_i (h(u_i) - \lambda_n f_h^{(n)}(u_i)), \quad \text{by (4.44)} \\ &= \frac{1}{m} \sum_i h(u_i) - \frac{1}{n} \sum_{i=1}^m f_h^{(n)}(u_i). \end{aligned} \quad (4.45)$$

On the other hand:

$$\begin{aligned} & (1 - \lambda_n) \sum_j \frac{\tilde{Q}_n(\mathcal{Y}_{nj})}{Q(\mathcal{Y}_{nj})} \int Q_{nj}(u) f_h^{(n)}(u) dG(u) \\ &= \frac{1}{n} \sum_{i=m+1}^n E_Y^{(n)}(\tilde{h}_n(u_i) | y_i). \end{aligned}$$

Combine (4.42), (4.43), (4.44), (4.45) to conclude that

$$\int h(u) dG_{\tilde{b}_n}(u) = \frac{1}{n} \sum_i \Psi_h^{(n)}(X_i) + o_p(n^{-1/2})$$

uniformly on  $\mathcal{H}$ .  $\square$

**Lemma 4.8:** *Under R0, R1*

$$\|\tilde{b}\|_\infty = O_P(J_n n^{-1/2}) = o_p(1), \quad (4.46)$$

where  $\|x\|_\infty \equiv \max_i |x_i|$ .

**Proof.** By definition (1.7),

$$\tilde{b}_i = \frac{(1 - \lambda_n)}{\lambda_n} \sum_j \left[ \frac{\tilde{Q}(\mathcal{Y}_{nj})}{Q^*(\mathcal{Y}_{nj}|\tilde{b})} - 1 \right] Q_{nj}(u_i). \quad (4.47)$$

Now,

$$\begin{aligned} & |Q(\mathcal{Y}_{nj}|\tilde{b}) - Q^*(\mathcal{Y}_{nj})| \\ & \leq \|\tilde{b}\| \left[ \frac{1}{m} \sum_i Q^2(\mathcal{Y}_{nj}|u_i) \right]^{1/2} \\ & \leq \|\tilde{b}\| Q^{*1/2}(\mathcal{Y}_{nj}). \end{aligned} \quad (4.48)$$

Thus,

$$\max_j \left| \frac{Q(\mathcal{Y}_{nj}|\tilde{b})}{Q^*(\mathcal{Y}_{nj})} - 1 \right| = O_P(J_n n^{-1/2}) \quad (4.49)$$

by lemmas 4.4 and 4.1. Hence,

$$\begin{aligned} \|\tilde{b}\|_\infty & \leq \frac{(1 - \lambda_n)}{\lambda_n} \max_j \left| \frac{\tilde{Q}(\mathcal{Y}_{nj})}{Q^*(\mathcal{Y}_{nj}|\tilde{b})} - 1 \right| \\ & = O_P(J_n n^{-1/2}). \quad \square \end{aligned}$$

**Lemma 4.9:** Under  $R0, R1$

$$\|\hat{b}\|_\infty = O_P(J_n^{3/2} n^{-1/2}). \quad (4.50)$$

**Proof.** We show first that  $\|\hat{b}\|_\infty = o_p(1)$  or equivalently, for all  $c > 0$ ,

$$P[\hat{b} - \tilde{b} \in \mathcal{L}_c] \rightarrow 1 \quad (4.51)$$

where  $\mathcal{L}_c \equiv \{\Delta : \|\Delta\|_\infty \leq c\}$ . Since  $L_n$  is the gradient of a convex function, to establish (4.51) it is enough to show that, for  $c > 0$  sufficiently small,

$$P[\inf\{\langle L_n(\tilde{b} + \Delta), \Delta \rangle : \|\Delta\|_\infty = c\}]$$



$$> 0] \rightarrow 1. \quad (4.52)$$

Now,

$$\langle \Delta, L_n(\tilde{b} + \Delta) \rangle = \langle \Delta, W_n(\tilde{b} + \Delta) - W_n(\tilde{b}) \rangle + \langle \Delta, L_n(\tilde{b} + \Delta) - W_n(\tilde{b} + \Delta) \rangle. \quad (4.53)$$

Since  $\dot{W}_n(b)$  is well defined for  $\|b\|_\infty < 1$  and has minimal eigenvalue  $\geq \lambda_n$  we deduce, from (4.53), that

$$\begin{aligned} \langle \Delta, L_n(\tilde{b} + \Delta) \rangle &\geq \lambda_n \langle \Delta, \Delta \rangle \\ &\quad - \lambda_n m^{-1} \sum_{i=1}^m \Delta_i (\tilde{b}_i + \Delta_i)^2 (1 + \tilde{b}_i + \Delta_i)^{-1}. \end{aligned} \quad (4.54)$$

Now, if  $\|\Delta\|_\infty = c < \frac{1}{5}$  say,

$$\begin{aligned} \Delta_i^2 &- \Delta_i (\tilde{b}_i + \Delta_i)^2 (1 + \tilde{b}_i + \Delta_i)^{-1} \\ &= \Delta_i^2 \frac{(1 - \tilde{b}_i)}{(1 + \tilde{b}_i + \Delta_i)} - \frac{\tilde{b}_i^2 \Delta_i}{1 + \tilde{b}_i + \Delta_i} \\ &\geq \alpha \Delta_i^2 - 2\beta \tilde{b}_i^2 |\Delta_i| \\ &\geq -\alpha^{-1} \beta^2 \tilde{b}_i^4 \end{aligned} \quad (4.55)$$

for some  $\alpha, \beta > 0$  independent of  $\Delta, \tilde{b}$ . Since  $\|\Delta\|_\infty = c$  we deduce that,

$$\langle \Delta, m L_n(\tilde{b} + \Delta) \rangle \geq \lambda_n \{ \alpha c^2 - 2\beta \tilde{b}_i^2 c - \alpha^{-1} \beta^2 \sum_i \tilde{b}_i^4 \}. \quad (4.56)$$

By lemmas 4.4 and 4.8, under R1,  $\|\tilde{b}\|_\infty = o_p(1)$  and  $\sum_i \tilde{b}_i^4 = o_p(1)$ . Then, (4.52) and (4.51) follow. Next we note that by (1.6),

$$\frac{\hat{b}_i}{1 + \hat{b}_i} = \frac{1 - \lambda_n}{\lambda_n} \sum_j \left[ \frac{\tilde{Q}(\mathcal{Y}_{nj})}{Q^*(\mathcal{Y}_{nj}|\hat{b})} - 1 \right] Q_{nj}(u_i). \quad (4.57)$$

But if  $\|\hat{b}\|_\infty = o_p(1)$

$$\max_i \left| \frac{\hat{b}_i}{1 + \hat{b}_i} \right| = \|\hat{b}\|_\infty (1 + o(1)).$$

The argument of lemma 4.8 now yields (4.50).  $\square$

**Proof of (2.12):** Again, we prove the uniform extension given in (iii). Since,

$$\begin{aligned} L_n(0) &= W_n(0) \\ \dot{L}_n(0) &= \dot{W}_n(0) \end{aligned}$$

we can evidently prove (2.13) by the same argument as (2.11) once we have established the analogue of (4.14):

$$\sup\{\|\dot{L}_n(b) - \dot{L}_n(0)\| : \|b\| < MJ_n n^{-1/2}, \|b\|_\infty \leq MJ_n n^{-1/2}\} = O_P(J_n^2 n^{-1/2}). \quad (4.58)$$

But,

$$\begin{aligned} \|\dot{L}_n(b) - \dot{L}_n(0)\| &\leq \|\dot{W}_n(b) - \dot{W}_n(0)\| + \max_i \left| \frac{1}{(1+b_i)^2} - 1 \right| \\ &\leq \|\dot{W}_n(b) - \dot{W}_n(0)\| + 2(1+o(1))\|b\|_\infty. \end{aligned} \quad (4.59)$$

Thus, (4.55) follows from (4.14), (4.56) and lemma 4.7. The proof of theorem 1 (i) (uniformly on  $\mathcal{H}$ ) is complete.  $\square$

Let  $\hat{E}_U, \hat{E}_Y^{(n)}$  be the conditional expectation operators with respect to  $U$  and  $\sigma(\mathcal{Y}_{1n}, \dots, \mathcal{Y}_{J_n n})$  respectively, when  $U \sim G_{0n}$ , the empirical distribution of the  $u_i$  and, given  $U = u, Y \sim Q(\cdot|u)$ . Let  $\dot{W}_{n0}^* = \lambda_n I + (1 - \lambda_n)\hat{E}_U \hat{E}_Y^{(n)}$ . It is easy to see that  $\dot{W}_{n0}^*$  can be viewed as an  $m \times m$  matrix and

$$\langle x, \dot{W}_{n0}^* z \rangle = \lambda_n \langle x, z \rangle + (1 - \lambda_n) \sum_{j=1}^{J_n} \frac{\langle z, Q_{jn}(\cdot) \rangle \langle x, Q_{jn}(\cdot) \rangle}{Q^*(\mathcal{Y}_{jn})}. \quad (4.60)$$

where  $Q_{jn}(\cdot) \equiv (Q_{jn}(u_1), \dots, Q_{jn}(u_m))$ .

**Lemma 4.9:** *If  $\mathcal{H}$  is regular and  $\Delta(\mathcal{H}) \equiv \{h_1 - h_2 : h_j \in \mathcal{H}, j = 1, 2\}$  then*

$$\lim_{\delta \rightarrow 0} \limsup_n P[\sup\{n^{1/2} |\langle [\dot{W}_{n0}^*]^{-1} \tau_{nh}, \bar{W}_n(0) \rangle| : h \in \Delta(\mathcal{H}), \|h\| \leq \delta\} \geq \epsilon] = 0.$$

**Proof.** If  $a \in R^m$ ,

$$n^{1/2} \langle a, \bar{W}_n(0) \rangle = n^{1/2} \sum_{j=1}^{J_n} c_j(a) (\tilde{Q}_n(\mathcal{Y}_{nj}) - Q(\mathcal{Y}_{nj})) \quad (4.61)$$

where

$$c_j(a) \equiv \frac{1}{m} \sum_i a(u_i) \frac{Q_j(u_i)}{Q(\mathcal{Y}_{nj})}.$$

If we condition on  $u_1, \dots, u_m$  we can write

$$n^{1/2} \langle a, \bar{W}_n(0) \rangle = \frac{n^{1/2}}{n-m} \sum_k h(Y_k, a) \quad (4.62)$$

where  $h(y, a) = \sum_j (c_j - \bar{c})(a) 1(y \in \mathcal{Y}_{nj})$ ,  $\bar{c} = \sum_j c_j Q(\mathcal{Y}_{nj})$ . Now, uniformly in  $a$ ,

$$\begin{aligned} \frac{1}{n-m} \sum_k h^2(Y_k, a) &= \sum_j (c_j - \bar{c})^2 \tilde{Q}(\mathcal{Y}_{nj}) \\ &= \left\{ \sum_j (c_j - \bar{c})^2 Q(\mathcal{Y}_{nj}) \right\} (1 + o_p(1)) \\ &\leq \left\{ \sum_j c_j^2 Q(\mathcal{Y}_{nj}) \right\} (1 + o_p(1)) \\ &= \left\{ \sum_j \left( \frac{1}{m} \sum_i a(u_i) Q_{nj}(u_i) \right)^2 / Q(\mathcal{Y}_{nj}) \right\} (1 + o_p(1)) \\ &\leq \left\{ \sum_j \left( \frac{1}{m} \sum_i a^2(u_i) Q_{nj}(u_i) \right) Q^*(\mathcal{Y}_{nj}) / Q(\mathcal{Y}_{nj}) \right\} (1 + o_p(1)) \\ &= \|a\|^2 (1 + o_p(1)). \end{aligned} \quad (4.63)$$

We use the notation of Pollard (1990). If  $\mathcal{A} \subset R^m$   $\delta(\mathcal{A}) = \sup_{\mathcal{A}} \|a\|$  and

$$\int_0^{\delta(\mathcal{A})} \sqrt{\log D(x, \mathcal{A})} dx \leq M \quad (4.64)$$

then from lemma 3.4 of Pollard (1990)

$$\sqrt{m} \sup_{\mathcal{A}} \langle a, \sigma \rangle \leq 9M \quad (4.65)$$

where  $\sigma = (\sigma_1, \dots, \sigma_m)$  and the  $\sigma_i$  are independent  $\pm 1$  with probability  $\frac{1}{2}$ . By (4.63), if

$$\mathcal{B} = \{(h(Y_1, a), \dots, h(Y_{n-m}, a)) : a \in \mathcal{A}\} \subset R^{n-m}$$

then with probability tending to 1,

$$\int_0^{\delta(\mathcal{B})} \sqrt{\log D(x, \mathcal{B})} dx \leq \int_0^{2\delta(\mathcal{A})} \sqrt{\log D(x, \mathcal{A})} dx \leq 2M. \quad (4.66)$$

Finally, since  $\|[W_{n0}^*]^{-1}\| \leq \lambda_n^{-1}$ , if (4.64) holds and

$$\Gamma = \{[W_{n0}^*]^{-1}a : a \in \mathcal{A}\} \quad (4.67)$$

then

$$\int_0^{\delta(\Gamma)} \sqrt{\log D(x, \Gamma)} dx \leq \lambda_n^{-1}M. \quad (4.68)$$

Since  $\mathcal{H}$  is regular (4.64) and (4.68) imply (by standard arguments from (4.65)) that if  $\mathcal{A} = \{\tau_n h : h \in \Delta\mathcal{H}, \|\tau_n h\| \leq \delta\}$  and  $\Gamma$  is defined by (4.67) then

$$\limsup_n P[\sup\{n^{1/2}|\langle \gamma, \bar{W}_n(0) \rangle| : \gamma \in \Gamma\} \geq \epsilon] \leq C(\delta, \epsilon) \quad (4.69)$$

where  $C(\delta, \epsilon) \rightarrow 0$  as  $\delta \rightarrow 0$ .

Again by regularity,

$$\lim_n P[\|\tau_n h\| > 2\delta, \|h\|_G \leq \delta, h \in \Delta\mathcal{H}] = 0. \quad (4.70)$$

By combining (4.69) and (4.70) we obtain the lemma.  $\square$

**Proof of theorem 1(iii).** It is easy to argue as for lemma 4.7 that

$$\sup_{\mathcal{H}} \|[ \dot{W}_{n0}^* ]^{-1} \tau_n h - \dot{W}_{n0}^{-1} \tau_n h\| = O_P(n^{-1/2} J_n^{1/2} \log(J_n)). \quad (4.71)$$

In view of (4.40), to prove theorem 1 (iii) it is enough to show that

$$\lim_{\delta \rightarrow 0} \limsup_n P[\sup\{|n^{1/2} \langle \dot{W}_{n0}^{-1} \tau_n h, \bar{W}_n(0) \rangle| : h \in \Delta(\mathcal{H}), \|h\| \leq \delta\} \geq \epsilon] = 0 \quad (4.72)$$

for all  $\epsilon > 0$ . By (4.71) we can replace  $\dot{W}_{n0}^{-1} \tau_n h$  by  $[ \dot{W}_{n0}^* ]^{-1} \tau_n h$  in (4.72) and the theorem is now a consequence of lemma 4.9.  $\square$

## 5 References

Bickel, P., Klaassen, C.A.J., Ritov, Y. and Wellner, J. (1993). Efficient and adaptive estimation in semiparametric models. *Johns Hopkins Press* (To appear).

Fan, J. (1991). "On the optimal rates of convergence for nonparametric deconvolution problems," *Ann. Statist.*, 19, pp. 1257-1272.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* 58, 13-30.

Ibragimov, I.A. and Has'minskii, R.Z. (1983). "On asymptotic efficiency in the presence of an infinite - dimensional nuisance parameter," *Lecture Notes in Mathematics*, 1021, pp. 195-229, Springer - Verlag, New York. *Theor. Prob. Applic.*, 27, pp. 551-562.

Jewell, N.P. (1982). "Mixtures of exponential distributions," *Ann. Statist.*, 10, pp. 479-484.

Jin, K. (1990). "Empirical smoothing parameter selection in adaptive estimation," Ph.D. dissertation, University of California, Berkeley.

Kantorovich, L.V. and Akilov, G.P. (1982). "Functional Analysis," Second Edition, Pergamon Press, Oxford.

Lindsay, B.G. (1983a). "Efficiency of the conditional score in a mixture setting," *Ann. Statist.*, 11, pp. 486-497.

Lindsay, B.G. (1983b). "The geometry of mixture likelihoods: A general theory," *Ann. Statist.*, 11, pp. 86-94.

Little, R.J.A. and Rubin, D. (1987). Statistical analysis with missing data. *J. Wiley*.

Neyman, J. and Scott, E. (1948). "Consistent estimates based on partially consistent observations," *Econometrica*, 16, pp. 1-32.

Pfanzagl, J. (1990). "Estimation in Semiparametric Models: Some Recent

- Developments,” *Lecture Notes in Statistics*, 63, Springer-Verlag, New York.
- Pollard, D. (1990). “Empirical Processes: Theory and Applications,” *NSF - CBMS Regional Conference Series in Probability and Statistics*, 2, Institute of Mathematical Statistics and the American Statistical Association, Hayward.
- van der Vaart, A. (1992). “Maximum likelihood estimation with partially censored data,” Technical Report, Free University of Amsterdam.
- Vardi, Y. (1989). “Multiplicative censoring, renewal process, deconvolution, and decreasing density: Nonparametric estimation,” *Biometrika*, 76, 751-761.
- Vardi, Y. and Zhang, C.-H. (1992). “Large sample study of empirical distributions in a random - multiplicative censoring model,” *Ann. Statist.*, 20, pp. 1022-1039.