C Additional Estimation Results (PRELIMINARY AND INCOMPLETE!!!) (TO BE MADE AVAILABLE UPON REQUEST

C.1 Asymptotic Linearity of $\hat{P}(Z)$:

Using Theorem 3 of Heckman, Ichimura and Todd, for any $0 \leq p \leq \overline{p}_z \leq \tilde{p}$, we get[26]

$$\left[\hat{P}(z) - P(z)\right]\hat{I}_1(x,z) = \frac{1}{N}\sum_{j=1}^{N}\psi_{NP}(X_i, Z_i, D_i; x, z) + \hat{b}_P(z) + \hat{R}_P(z),$$

where $N^{-1/2}\sum_{i=1}^{N}\hat{R}_P(X_i,Z_i) = o_p(1)$, $plim_{N\to\infty}N^{-1/2}\sum_{i=1}^{N}\hat{b}_P(X_i,Z_i) = b_P < \infty$, $E[\psi_{NP}(X_i,Z_i,D_i;X,Z|X=x,Z=z] = 0$. The particular forms of ψ_{NP} and $\hat{b}_P$ are given by

$$\psi_{NP}(X_i, Z_i, D_i; x, z) = \tfrac{1}{h_{NP}^{d_z}}e_1[M_{p,N}^P(z)]^{-1}\left[\left(\tfrac{Z_i - z}{h_{NP}}\right)^{Q_p}\right]' K^P\left(\tfrac{Z_i - z}{h_{NP}}\right)\varepsilon_i^D I_1(x,z)$$

$$\hat{b}_P(x,z) = h_{NP}^{\overline{p}_z}e_1[M_{pN}^P(z)]^{-1}\hat{I}_1(x,z)\sum_{s=p+1}^{\overline{p}_z}\left[\int u^{Q(0)}\cdot u^{Q(s)}P^{(s)}(z)'\cdot u^{Q(\overline{p}_z - s)}K(u)\mathrm{d}u,\right.$$
$$\left. \ldots, \int u^{Q(p)}\cdot u^{Q(s)}P^{(s)}(z)'\cdot u^{Q(\overline{p}_z - s)}K(u)\mathrm{d}u\right]f_Z^{(\overline{p}_z - s)}(z)'$$

where $P^{(s)}$ denotes the s-th order derivative of P and $d_z = dim(Z)$. And if $p = \overline{p}_z$, the estimator has the same form, but with $\hat{b}_P(z) = o(h_{NP}^{\overline{p}_z})$.

C.2 Estimating $h_0(x, P(z))$

This appendix has two goals: first to show that the local polynomial regression estimator of h_0 is asymptotically linear with trimming; second, to show that its derivative with respect to P is uniformly consistent for the derivative of h_0 with respect to P. To show that local polynomial regression estimator of h_0 is asymptotically linear with trimming, we follow arguments similar to those in the proof of theorem 3 of Heckman, Ichimura and Todd.

Write $Y = m + \varepsilon = X_{\overline{p}_h}(x_0)\beta^*_{\overline{p}_h}(x_0) + r_{\overline{p}_h}(X,x_0) + \varepsilon$, where $\varepsilon = Y - E(Y|X)$. In our case $-(1-D)Y$ will play the role of Y in HIT, and the vector $(P(Z), X)$ will play the role of X in HIT. In the

[26] If we were only estimating $E(D|Z)$, i.e. if we did not have the later steps of our estimation procedure, we would use a trimming function that is based on the estimated density of Z, and we would only need that $E(D|Z)$ is $\overline{p}_z$ smooth with $\overline{p}_z > d_z$. But we need to employ another trimming function at a later step, and for that trimming function to be well behaved it must be that $f_{X,Z}$ and $E(D|Z)$ are both $\tilde{p}$-smooth with $\tilde{p} > \dim(X,Z)$. We could, in principle, state the result of this section in terms of $\tilde{p}$ only, but to have a bias term that is $O(h_{NP}^{\tilde{p}})$ we would need to assume more moments of K^P are 0.

first part of this section, we will use $\tilde{Y}$ and S to denote $(1-D)Y$ and $(P(Z),X)$. Let $d_s = d_x + 1$[27],

$$s_0 = (P(z_0), x_0),\ S_p(s_0) = \begin{pmatrix} (S_1 - s_0)^{Q_p} \\ \vdots \\ (S_N - s_0)^{Q_p} \end{pmatrix},\ H = diag(1, h_{Nh}^{-1}\iota_{d_s}, \dots, h_{Nh}^{-p}\iota_{(p+d_s-1)!/[p!(d_s-1)!]}),$$

$$W(s_0) = h_{Nh}^{-d_s} diag\ K^h\ \frac{S_1 - s_0}{h_{Nh}}\ , \dots K^h\ \frac{S_N - s_0}{h_{Nh}} \qquad \hat{M}_{pN}^h(s_0) = N^{-1}H'S_p(s_0)'W(s_0)S_p(s_0)H$$

and $M_{pN}^h(s_0) = E[\hat{M}_{pN}^h(s_0)]$. Just as in Heckman, Ichimura and Todd, we will consider the case, where p, the order of the polynomial terms included, is less than $\overline{p}_h \leq \tilde{p}$[28]. To do that partition, $S_{\overline{p}_h}(s_0) = [S_p(s_0), \overline{S}_{\overline{p}_h}(s_0)]$ and $\beta^*_{\overline{p}_h}(s_0) = [\beta^*_p(s_0)', \overline{\beta}^*_{\overline{p}_h}(s_0)']'$. Then,

$$\begin{aligned} [\hat{\beta}_p(s_0) - \beta^*_p(s_0)]\hat{I}_{10} &= H[\hat{M}_{pN}^h(s_0)]^{-1}N^{-1}H'S_p(s_0)'W(s_0)\varepsilon^h\hat{I}_{10} \\ &+ H[\hat{M}_{pN}^h(s_0)]^{-1}N^{-1}H'S_p(s_0)'W(s_0)\overline{S}_{\overline{p}_h}(s_0)\overline{\beta}^*_{\overline{p}_h}(s_0)\hat{I}_{10} \\ &+ H[\hat{M}_{pN}^h(s_0)]^{-1}N^{-1}H'S_p(s_0)'W(s_0)r_{\overline{p}_h}(s_0)\hat{I}_{10} \end{aligned}$$

We need to show that $e_2[\hat{\beta}_p(s_0) - \beta^*_p(s_0)]\hat{I}_{10}$ is asymptotically linear.

C.2.1 First Step

As our first step, we claim that

$$e_2H[\hat{M}_{pN}^h(s_0)]^{-1}N^{-1}H'S_p(s_0)'W(s_0)\varepsilon^h\hat{I}_{10} = e_2H[M_{pN}^h(s_0)]^{-1}N^{-1}H'S_p(s_0)'W(s_0)\varepsilon^h I_0 + \hat{R}_1(s_0)$$

where $e_2 = (0,1,0,...,0)$ and $\sum_{i=1}^N \hat{R}_1(P(Z_i),X_i)/\sqrt{N} = o_p(1)$. Note that $e_2H = \frac{1}{h_{Nh}}e_2$. Let $\gamma_{N0}(P(Z_j),X_j) = e_2[M_{pN}^h(P(Z_j),X_j)]^{-1}$, $\hat{\gamma}_N(P(Z_j),X_j) = e_2[\hat{M}_{pN}^h(P(Z_j),X_j)]^{-1}$, and

$$\Gamma_N = \{\gamma_N : \sup_{(x,z)\in\overline{A}_1} |\gamma_N(x,z) - e_2[M_{pN}(x,z)]^{-1}| \leq \epsilon_\gamma\}$$

Let

$$\mathcal{G}_{1N} = \left\{ g_N : g_N(\varepsilon_i^h, X_i, Z_i, X_j, Z_j) = N^{-3/2}\gamma_N(P(Z_j),X_j)\frac{\varepsilon_i^h}{h_{Nh}^{d_x+2}}\left[\ \frac{(P(Z_i),X_i)-(P(Z_j),X_j)}{h_{Nh}}\ ^{Q_p}\right]' K^h\ \frac{(P(Z_i),X_i)-(P(Z_j),X_j)}{h_{Nh}}\ \tilde{I}_{1i}\tilde{I}_{1j} \right\}$$

where $\tilde{I}_1 \in \mathcal{I}_1$, which defined in Appendix (C.4).

Also let g_{N0} be the same as g_N except with γ_N replaced by γ_{N0}, and $\tilde{I}_{1i}$ and $\tilde{I}_{1j}$ replaced by I_{1i} and I_{1j}. And define $\hat{g}_N$ similarly with $\hat{\gamma}_N$, $\hat{I}_{1i}$ and $\hat{I}_{1j}$ replacing γ_N,$\tilde{I}_{1i}$ and $\tilde{I}_{1j}$, respectively. With

[27]In this section we pretend that the estimation is done as if the actual values of $P(Z_i)$ are observed. In a later section we are going to show that this does not affect the asymptotic variance of our estimator as long as $\hat{P}(Z_i) - P(Z_i) = o_p(h_{Nh})$, which is only true if $\hat{P}$ is estimated on a region where the density of (X,Z) is bounded away from 0. This is the reason we trim every observation in defining $\hat{W}$.

[28]We introduce $\overline{p}_h$ for the same reason we introduced $\overline{p}_P$. For additional information please refer to footnote (C.1).

this new notation $1/\sqrt{N}\sum_i \hat{R}_1(X_j, Z_j) = \sum_i \sum_j [\hat{g}_N(\varepsilon_i^h, X_i, Z_i, X_j, Z_j) - g_{N0}(\varepsilon_i^h, X_i, Z_i, X_j, Z_j)]$. To show that this sum is $o_p(1)$, we first need to show that $\sum_j \sum_i g_N(\varepsilon_i^h, X_i, Z_i, X_j, Z_j)$ is equicontinuous over $\mathcal{G}_{1N}$ in a neighborhood of $g_{N0}(\varepsilon_i^h, X_i, Z_i, X_j, Z_j)$ and that with probability approaching to 1, $\hat{g}_N(\varepsilon_i^h, X_i, Z_i, X_j, Z_j)$ lies within the neighborhood over which equicontinuity is established. For the first step, we try using this lemma. To apply that lemma, we need to have a degenerate U-process, and $\sum_i \sum_j g_N(\varepsilon_i^h, X_i, Z_i, X_j, Z_j)$ is not degenerate. To deal with this issue, we first split the $\sum_i \sum_j g_N(\varepsilon_i^h, X_i, Z_i, X_j, Z_j)$ process into two sums: $\sum_i \sum_{j\neq i} g_N(\varepsilon_i^h, X_i, Z_i, X_j, Z_j)$ and $\sum_i g_N(\epsilon_i^h, X_i, Z_i)$. The latter process is symmetric. To see that it is also degenerate, we observe that $\left(\frac{(P(Z_i),X_i)-(P(Z_i),X_i)}{h_{Nh}}\right)^{Q_p}$ is a row vector whose first component equals 1 and all other components equal 0.

$$
\begin{aligned}
g_N(\varepsilon_i^h, X_i, Z_i) &= N^{-3/2}\gamma_N(P(Z_i), X_i)e_1'\varepsilon_i^h \left(\frac{1}{h_{Nh}}\right)^{d_x+2} K(0) \\
E[g_N(\varepsilon_i^h, X_i, Z_i)] &= E[\gamma_N(P(Z_i), X_i)e_1' E(\varepsilon_i^h | X_i, P(Z_i)] \frac{K(0)}{N^{3/2}h_{Nh}^{d_x+2}} = 0
\end{aligned}
$$

Thus, $\sum_i g_N(\varepsilon_i^h, X_i, Z_i)$ is degenerate.

Next, define $g_N^0 := \frac{g_N(\varepsilon_i^h,X_i,Z_i,X_j,Z_j)+g_N(\varepsilon_j^h,X_j,Z_j,X_i,Z_i)}{2}$, $L_i := (\varepsilon_i^h, X_i, Z_i)$, $\phi_N(L_i) := E[g_N^0(L_i, l)|L_i] = E[g_N^0(l, L_i)|L_i]$, and $\tilde{g}_N^0(L_i, L_j) = g_N^0(L_i, L_j) - \phi_N(L_i) - \phi_N(L_j)$ as in Heckman, Ichimura and Todd, so that $\sum_i \sum_{j\neq i} g_N^0(L_i, L_j) = \sum_i \sum_{j\neq i} \tilde{g}_N^0(L_i, L_j) + \sum_{i=1}^N 2(N-1)\phi_N(L_i)$. To show equicontinuity of our original process we need to show that that each of the processes $\sum_{i=1}^N g_N(\epsilon_i^h, X_i, Z_i)$, $\sum_i \sum_{j\neq i} \tilde{g}_N^0(L_i, L_j)$ and $\sum_{i=1}^N 2(N-1)\phi_N(L_i)$ are degenerate. We already verified that the first of these is degenerate. To show that the latter two are degenerate, one could use law of iterated expectations and the independence of (ϵ, U) from (X, Z) in a similar fashion to argue that

$$
\phi_N(L_i) = \frac{1}{2}\frac{\varepsilon_i^h}{h_{Nh}^{d_x+2}} N^{-3/2}\tilde{I}(X_i, Z_i) E\left[\gamma_N(P(Z_j), X_j) \quad \frac{(P(Z_i),X_i)-(P(Z_j),X_j)}{h_{Nh}} \quad {}^{Q_p}\Big)' \varepsilon_j^h K \quad \frac{(P(Z_i),X_i)-(P(Z_j),X_j)}{h_{Nh}} \quad |\varepsilon_i^h, X_i, Z_i\right].
$$

Thus, $\phi_N(\varepsilon_i^h, X_i, Z_i) = \frac{\varepsilon_i^h \tilde{I}(X_i,Z_i)t(X_iZ_i)}{2N^{3/2}h_{Nh}^{d_x+2}}$ for some measurable function $t(X_i, Z_i)$. Now we can go through the same arguments as above and show that $E[\phi_N(\varepsilon_i^h, X_i, Z_i)] = 0$. Thus, all the processes are degenerate, and lemma 3 is applicable to each of them. So the next step is the verification of the conditions of the equicontinuity lemma.

Let $\overline{I}_{1i}$ be as in Appendix (C.4). Then $|g_N(\varepsilon_i^h, X_i, Z_i)| \leq N^{-3/2}C|e_1'||\varepsilon_i^h|K(0)\overline{I}_{1i}$, and

$$
\sum_{i=1}^N E\left[N^{-3}C^2(\varepsilon_i^h)^2 \frac{1}{h_{Nh}^{2(d_x+2)}} K(0)^2 \overline{I}_{1i}\right] \leq C^2 K(0)^2 E((\varepsilon^h)^2)\left(\frac{1}{Nh_{Nh}^{(d_x+2)}}\right)^2 < \infty
$$

This shows that condition (i) of the equicontinuity lemma holds for the $\sum_i g_n(\varepsilon_i^h, X_i, Z_i)$ process if $Nh_{Nh}^{d_x+2} \to \infty$. Condition (ii) holds under the same assumption by the dominated convergence theorem.

Next, we recall that $K(\cdot)$ is zero outside a compact set, so that when $\left|\left|\frac{(P(Z_i),X_i)-(P(Z_j),X_j)}{h_{Nh}}\right|\right|$ is "too large" $K\left(\frac{(P(Z_i),X_i)-(P(Z_j),X_j)}{h_{Nh}}\right)=0$. This implies that there exist C_1, C_2 such that any element of $\left[\left(\frac{(P(Z_i),X_i)-(P(Z_j),X_j)}{h_{Nh}}\right)^{Q_p}\right]' K\left(\frac{(P(Z_i),X_i)-(P(Z_j),X_j)}{h_{Nh}}\right)$ is bounded by $C_1\left(\frac{1}{h_{Nh}}\right)^{d_x+1} 1\{||(P(Z_i),X_i)-(P(Z_j),X_j)|| \leq C_2 h_{Nh}\}$. Then

$$|g_N(\varepsilon_i^h,X_i,Z_i,X_j,Z_j)| \leq N^{-3/2}CC_1 1\{||(P(Z_i),X_i)-(P(Z_j),X_j)|| \leq C_2 h_{Nh}\} \; \frac{1}{h_{Nh}}^{d_x+2} |\varepsilon_i^h| \overline{I}_1(X_j,Z_j)\overline{I}_1(X_i,Z_i)$$

Thus, as long as $Nh_{Nh}^{(d_x+2)} \to \infty$, conditions (i) and (ii) are satisfied for the process $\sum_i \sum_{j\neq i} g_N(\varepsilon_i^h, X_i, Z_i, X_j, Z_j)$ as well.

To verify the same conditions for the $2N\phi(\cdot)$ process, note that $|2N\phi_N(\varepsilon_i^h, X_i, Z_i)| \leq 2N^{-1/2}C|\varepsilon_i^h|$. Then the expression on the right hand side of this inequality provides an envelope for this process. Since $\sum_{i=1}^N 4N^{-1}C^2E\left((\varepsilon_i^h)^2\right) = 4C^2E\left((\varepsilon_i^h)^2\right) < \infty$, the first condition of Lemma 3 of HIT holds for the $2N\phi$ process. On the other hand, since $E\left((\varepsilon_i^h)^2\right) < \infty$, $(\varepsilon_i^h)^2 1\{|\varepsilon_i^h| > \sqrt{n}\frac{\delta}{2C}\} \to 0$ as $n\to\infty$, almost everywhere. Moreover, $(\varepsilon_i^h)^2 1\{|\varepsilon_i^h| > \sqrt{n}\frac{\delta}{2C}\} \leq (\varepsilon_i^h)^2$. Therefore, we could apply the Dominated Convergence Theorem to get that

$$\forall\delta>0, \quad \lim_{n\to\infty}\sum_{i=1}^n 4n^{-1}C^2E\left((\varepsilon_i^h)^2 1\left\{|\varepsilon_i^h| > \sqrt{n}\frac{\delta}{2C}\right\}\right) = 0$$

Now we move on to verifying condition (iii) of Lemma 3 of HIT for our three processes. First, let g_N^1 and g_N^2 be any two elements of $\mathcal{G}_{1N}$. Then

$$\begin{aligned}
|g_N^1(\varepsilon_i^h,X_i,Z_i,X_j,Z_j) - g_N^2(\varepsilon_i^h,X_i,Z_i,X_j,Z_j)| \;&\leq\; N^{-3/2}C_1 1\{||(P(Z_i),X_i)-(P(Z_j),X_j)|| \leq C_2 h_{Nh}\} \\
&\times\; \frac{|\varepsilon_i^h|}{h_{Nh}^{d_x+2}}\Big(|I_{1j}^1 - I_{1j}^2|\;||\gamma_n^1(S_j)||\overline{I}_{1i} \\
&+\; |I_{1i}^1 - I_{1i}^2|\;||\gamma_n^1(S_j)||\overline{I}_{1j} + |\gamma_n^1(S_j)-\gamma_n^2(S_j)|\overline{I}_{1j}\overline{I}_{1i}\Big)
\end{aligned}$$

Since $Nh_{Nh}^{d_x+2} \to \infty$, the $\mathcal{L}^2$ covering number of $\mathcal{G}_{1N}$ family is bounded by the product of the covering numbers of $\mathcal{I}_1$ and Γ_N. By the Kolmogorov-Tihomirov lemma and the results given in Appendix (C.4), we know that the third condition of the equicontinuity lemma is satisfied since $\tilde{p} > d_x+1$ and $Nh_{Nh}^{d_x+2}/\log N \to \infty$.

The arguments so far showed equicontinuity of the process $\sum_{i=1}^N\sum_{j=1}^N g_N(\epsilon_i^h, X_i, Z_i, X_j, Z_j)$ over $\mathcal{G}_{1N}$ in a neighborhood of $g_{N0}(\epsilon_i^h, X_i, Z_i, X_j, Z_j)$.

C.2.2 Second Step:

Next, we move on to the term that will contain the bias:

$$e_2H[\hat{M}_{pN}^h(s_0)]^{-1}N^{-1}H'S_p'(s_0)W(s_0)\overline{S}_{\overline{p}}(s_0)\overline{\beta}_{\overline{p}}^*(s_0)\hat{I}_{10} = \frac{\hat{I}_{10}}{Nh_{Nh}^{d_x+2}}e_2[\hat{M}_{pN}^h(P(z_0),x_0)]^{-1}\times$$

$$\sum_{k=p+1}^{\overline{p}_h}\sum_{i=1}^{\overline{p}_h} \frac{(P(Z_i),X_i)-(P(z_0),x_0)}{h_{Nh}}^{Q_p\;\prime} ((P(Z_i),X_i)-(P(z_0),x_0))^{Q(k)}[m^{(k)}(P(z_0),x_0)]'K^h \; \frac{(P(Z_i),X_i)-(P(z_0),x_0)}{h_{Nh}}$$

We add and subtract

$$e_2[\hat{M}_{pN}^h(s_0)]^{-1}\sum_{k=p+1}^{\overline{p}_h}\frac{\hat{I}_{10}}{h_{Nh}^{d_x+2}}E\left\{\left[\left(\frac{S_i-S_j}{h_{Nh}}\right)^{Q_p}\right]'(S_i-S_j)^{Q(k)}K^h\left(\frac{S_i-S_j}{h_{Nh}}\right)|S_j=s_0\right\}[m^{(k)}(s_0)]'$$

This gives us three terms. The difference of the first two terms can be shown to be $o_p(1)$ by appealing to the equicontinuity lemma. In particular, we take γ_N, Γ_N, and $\mathcal{I}_1$ as before and define

$$g_N(X_i,Z_i,X_j,Z_j)=\frac{\gamma_{Nj}\tilde{I}_{1j}}{N^{3/2}h_{Nh}^{d_x+2}}\left[\quad\frac{(P(Z_i),X_i)-(P(Z_j),X_j)}{h_{Nh}}\quad^{Q_p}\right]'((P(Z_i),X_i)-(P(Z_j),X_j))^{Q(k)}K^h\quad\frac{(P(Z_i),X_i)-(P(Z_j),X_j)}{h_{Nh}}\quad[m_j^{(k)}]'$$
$$-\frac{\tilde{I}_{1j}\gamma_{Nj}}{N^{3/2}h_{Nh}^{d_x+2}}E\left[\left[\quad\frac{(P(Z_i),X_i)-(P(Z_j),X_j)}{h_{Nh}}\quad^{Q_p}\right]'((P(Z_i),X_i)-(P(Z_j),X_j))^{Q(k)}K^h\quad\frac{(P(Z_i),X_i)-(P(Z_j),X_j)}{h_{Nh}}\quad|X_j,P(Z_j)\right][m_j^{(k)}]'$$

Let$\hat{g}_N(X_i,Z_i,X_j,Z_j)$ and $g_{N0}(X_i,Z_i,X_j,Z_j)$ be defined in the same way as before. Moreover, let $\mathcal{G}_{2N}:=\{g_N(X_i,Z_i,X_j,Z_j)|\gamma_N(X_j,Z_j)\in\Gamma_N\}$. Then going through the same steps as in **Step 1**, we can show that $1/\sqrt{N}\sum_{j=1}^N\hat{R}_{21}(X_j,Z_j)=\sum_{i=1}^N\sum_{j=1}^N[\hat{g}_N(X_i,Z_i,X_j,Z_j)-g_{N0}(X_i,Z_i,X_j,Z_j)]=o_p(1)$. Thus, we are left with the term

$$e_2[\hat{M}_{pN}^h(s_0)]^{-1}\sum_{k=p+1}^{\overline{p}_h}\frac{1}{h_{Nh}^{d_x+2}}E\left\{\left[\left(\frac{S_i-S_j}{h_{Nh}}\right)^{Q_p}\right]'(S_i-S_j)^{Q(k)}K^h\left(\frac{S_i-S_j}{h_{Nh}}\right)|S_j=s_0\right\}[m^{(k)}(s_0)]'\hat{I}_{10}$$

The last expression equals

$$h_N^{\overline{p}-1}e_2[M_p(s_0)]^{-1}\times$$
$$\sum_{k=p+1}^{\overline{p}}\left[\int u^{Q(0)}\cdot u^{Q(k)}m^{(k)}(s_0)'\cdot u^{Q(\overline{p}-1)}K(u)\mathrm{d}u,...,\int u^{Q(p)}\cdot u^{Q(k)}m^{(k)}(s_0)'\cdot u^{Q(\overline{p}-k)}K(u)\mathrm{d}u\right]$$
$$\times f^{(\overline{p}-k)}(s_0)'\hat{I}_{10}$$

We need

$$plim_{N\to\infty}1/\sqrt{N}\sum_{i=1}^N h_N^{\overline{p}-1}e_2[M_p(S_i)]^{-1}\times$$
$$\sum_{k=p+1}^{\overline{p}}\left[\int u^{Q(0)}\cdot u^{Q(k)}m^{(k)}(S_i)'\cdot u^{Q(\overline{p}-1)}K(u)\mathrm{d}u,...,\int u^{Q(p)}\cdot u^{Q(k)}m^{(k)}(S_i)'\cdot u^{Q(\overline{p}-1)}K(u)\mathrm{d}u\right]$$
$$\times f^{(\overline{p}-k)}(S_i)'\hat{I}(X_i,Z_i)=b_{h_0}<\infty$$

All the terms involving (X_i,Z_i) are bounded with probability 1. Since $Nh_{Nh}^{2(\overline{p}_h-1)}\to a<\infty$ this is indeed true.

C.2.3 Third Step:

Here we focus on the $e_2H[\hat{M}_{pN}^h(s_0)]^{-1}N^{-1}H'S_p'(s_0)W(s_0)r_{\overline{p}_h}(s_0)\hat{I}_{10}$ term. Since $e_2H=1/h_{Nh}e_2$, this term equals

$$\frac{1}{h_{Nh}}e_2[\hat{M}_{pN}^h(s_0)]^{-1}N^{-1}H'S_p'(s_0)W(s_0)r_{\overline{p}_h}(s_0)\hat{I}_{10}$$

Following the same steps as in the proof of lemma 8 of Heckman, Ichimura and Todd, we can show that $N^{-1}H'X_p'(s_0)W(s_0)r_{\overline{p}_h}(s_0) = o_p(h_{Nh}^{\overline{p}_h+1})$.
Combining all these results, we conclude that under the assumptions given in the Appendix B,

$$\left[\hat{h}_0(p,x) - h_0(p,x)\right]\hat{I}_1(x,z) = \frac{1}{N}\sum_{i=1}^{N}\psi_{Nh_0}(P(Z_i), X_i, -(1-D_i)Y_i; x, z) + \hat{b}_{h_0}(p,x,z) + \hat{R}_{h_0}(p,x,z)$$

where $N^{-1/2}\sum_{i=1}^{N}\hat{R}_{h_0}(P(Z_i, X_i, Z_i) = o_p(1)$, $\text{plim}_{N\to\infty}N^{-1/2}\sum_{i=1}^{N}\hat{b}_{h_0}(P(Z_i), X_i, Z_i) = b_{h_0} < \infty$, and $E[\psi_{Nh_0}(P(Z_i), X_i, D_iY_i; P(Z_i), X_i, Z_i|P(Z_i) = p, X_i = x, Z_i = x] = 0$, $p = P(z)$. Moreover,

$$\psi_{Nh_0}(P(Z_i),X_i,-(1-D_i)Y_i;p,x,z)=\frac{1}{h_{Nh}^{d_x+2}}e_2[M_{p,N}^h(P(z),x)]^{-1} \quad \frac{(P(Z_i),X_i)-(P(z),x)}{h_{Nh}} \quad ^{Q_p} \quad ' K^h \quad \frac{(P(Z_i),X_i)-(P(z),x)}{h_{Nh}} \quad \varepsilon_i^{h_0} I_1(x,z)$$

C.2.4 Asymptotic linearity of $\hat{h}_0(\hat{P}(z), x)$:

To show this, we need to use Lemma 1 of Heckman, Ichimura and Todd. Recall that

Lemma C.1 *(Heckman, Ichimura and Todd (1998)) Suppose that:*

1. *Both $\hat{P}(z)$ and $\hat{g}(p,t)$ are asymptotically linear with trimming where*

$$\left[\hat{P}(z) - P(z)\right]I\big((x,z) \in \hat{A}_1\big) = N^{-1}\sum_{j=1}^{N}\psi_{NP}(D_j, Z_j; x, z) + \hat{b}_P(x,z) + \hat{R}_P(x,z)$$

$$\left[\hat{g}(p,t) - g(p,t)\right]I\big((x,z) \in \hat{A}_1\big) = N^{-1}\sum_{j=1}^{N}\psi_{Ng}(Y_j, T_j, P(Z_j); p, t, z) + \hat{b}_g(p,t,z) + \hat{R}_g(p,t,z);$$

2. *$\partial\hat{g}(p,t)/\partial p$ and $\hat{P}(z)$ are uniformly consistent and converge to $\partial g(p,t)/\partial p$ and $P(z)$, respectively and $\partial g(p,t)/\partial p$ is continuous;*

3. $\text{plim}_{N\to\infty}N^{-1/2}\sum_{i=1}^{N}\hat{b}_g(P(Z_i), T_i, Z_i) = b_g$ *and*
$\text{plim}_{N\to\infty}N^{-1/2}\sum_{i=1}^{N}\frac{\partial g(P(Z_i),T_i)}{\partial p}\hat{b}_P(P(Z_i), T_i, Z_i) = b_{gP}$;

4. $\text{plim}_{N\to\infty}N^{-1/2}\sum_{i=1}^{N}\left[\frac{\partial\hat{g}(\overline{P}_{T_i}(Z_i),T_i)}{\partial p} - \frac{\partial g(P(Z_i),T_i)}{\partial p}\right]\hat{R}_P(P(Z_i), T_i, Z_i) = 0$, *and*
$\text{plim}_{N\to\infty}N^{-1/2}\sum_{i=1}^{N}\left[\frac{\partial\hat{g}(\overline{P}_{T_i}(Z_i),T_i)}{\partial p} - \frac{\partial g(P(Z_i),T_i)}{\partial p}\right]\hat{b}_P(P(Z_i), T_i, Z_i) = 0$;

5. $\text{plim}_{N\to\infty}N^{-3/2}\sum_{i=1}^{N}\left[\frac{\partial\hat{g}(\overline{P}_{T_i}(Z_i),T_i)}{\partial p} - \frac{\partial g(P(Z_i),T_i)}{\partial p}\right]\psi_{NP}(D_j, Z_j; T_i, Z_i) = 0.$

then $\hat{g}(\hat{P}(z),t)$ *is also asymptotically linear with trimming where*

$$\begin{aligned}\big[\hat{g}(\hat{P}(z),t)-g(P(z),t)\big]I\big((x,z)\in\hat{A}_1\big) &= N^{-1}\sum_{j=1}^{N}\big[\psi_{Ng}(Y_j,T_j,P(Z_j),Z_j;P(z),t,z)\\ &+ \partial g(t,P(z))/\partial p\ \cdot I_{1j}\psi_{NP}(D_j,Z_j,X_j;x,z)\big]\\ &+ \hat{b}^g(x,z)+\hat{R}^g(x,z),\end{aligned}$$

and $\operatorname{plim}_{N\to\infty}\sum_{i=1}^{N}\hat{b}^g(X_i,Z_i)=b_g+b_{gP}$.

In our case, $g(p,x)=\frac{\partial}{\partial p}E\big(-(1-D)Y|P(Z)=p,X=x\big)$. The verification of the conditions for Lemma 1 of HIT for the case where g itself is the derivative of some conditional expectation with respect to one of the conditioning variables is not really different from what Heckman, Ichimura and Todd have. The only potential difference is in the proof of theorem 4, but even there, their argument holds for the entire $\nabla\hat{\beta}$ vector, not just the first component. Therefore,

$$\begin{aligned}\big[\hat{h}_0(\hat{P}(z),x)-h_0(P(z),x)\big]I\big((x,z)\in\hat{A}_1\big) &= N^{-1}\textstyle\sum_{j=1}^{N}\big[\psi_{Nh_0}(-(1-D_j)Y_j,P(Z_j),X_j;P(z),x,z)+\frac{\partial h_0(P(z),x)}{\partial p}I_{1j}\psi_{nP}(D_j,Z_j;x,z)\big]\\ &+ \hat{b}_{\hat{h}_0}(x,z)+\hat{R}_{\hat{h}_0}(x,z)\end{aligned}$$

with $\operatorname{plim}_{N\to\infty}\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\hat{b}_{\hat{h}_0}(X_j,Z_j)=b_{h_0}+b_{h_0P}<\infty$, and $\operatorname{plim}_{N\to\infty}\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\hat{R}_{\hat{h}_0}(X_j,Z_j)=0$.

C.3 Estimating $q(h_0(x,P(z)),P(z))$

We need to estimate $E(Y|D=1,h_1(X,P(Z)),P(Z))=\frac{E(DY|h_1(X,P(Z)),P(Z))}{P(Z)}$. We are going to use local polynomial regression to estimate $E(DY|h_1(X,P(Z)),P(Z))$. As a result, the analysis here is very similar to the proof of their theorem 3. The only difference is that we evaluate this estimator at the value of the random vector $(h_0(X_i,P(Z_i)),P(Z_i))$, which is different from the random vector we condition on. As long as the support of $h_0(X_i,P(Z_i))$ overlaps with the support of $h_1(X_i,P(Z_i))$ this is well defined.

To simplify the following expressions, define $T_{1i}:=(h_1(X_i,P(Z_i)),P(Z_i))$, and $T_{0i}:=(h_0(X_i,P(Z_i)),P(Z_i))$. Let t_1 and t_0 denote a value in the interior of the support of T_1 and T_0, respectively. And let p_0 denote that point in the interior of the support of $P(Z)$ that corresponds to t_0. Let $\hat{I}_{1i}$ and I_{1i} be as before. Let $\hat{I}_{2i}:=1\{\hat{f}_{\hat{h}_1,\hat{P}}(\hat{h}_0(X_i,\hat{P}(Z_i)),\hat{P}(Z_i))\geq q_{02}\}$, and $I_{2i}:=1\{f_{h_1,P}(h_0(X_i,P(Z_i)),P(Z_i))\geq q_{02}\}$. Our goal is to derive the asymptotic distribution of

$$\tfrac{1}{\sqrt{N}}\textstyle\sum_{i=1}^{N}\big[D_i\big(\hat{q}(h_{0i},P_i)-q(h_{0i},P_i)\big)\big]\hat{I}_1(X_i,Z_i)\hat{I}_2(X_i,Z_i)$$

Let $T_p(t_0):=\begin{pmatrix}(T_{11}-t_0)^{Q_p}\\ \vdots\\ (T_{1N}-t_0)^{Q_p}\end{pmatrix}$, $\varepsilon_i^q:=D_iY_i-E(D_iY_i|T_{1i})$, and $M_{pN}^q(t_0)=E[\hat{M}_{pN}^q(t_0)]$, where

$$W^q(t_0):=h_{Nq}^{-2}diag\ \ K^q\ \ \tfrac{T_{11}-t_0}{h_{Nq}}\ \ ,\dots,K^q\ \ \tfrac{T_{1N}-t_0}{h_{Nq}}\ \ ,\qquad \hat{M}_{pN}^q(t_0)=N^{-1}H^{q\prime}T_p(t_0)'W^q(t_0)T_p(t_0)H^q.$$

where H^q is defined in the same way as H in section (C.2) with h_{Nq} replacing h_{Nh}. Then

$$\frac{1}{\sqrt{N}}\sum_{j=1}^{N} D_j[\hat{q}(T_{0j})-q(T_{0j})]\hat{I}_{1j}\hat{I}_{2j} = \frac{1}{N\sqrt{N}}\sum_{j=1}^{N}\frac{D_j}{P(Z_j)}e_1[\hat{M}^q_{pN}(T_{0j})]^{-1}H^{q\prime}T_p'(T_{0j})W^q(T_{0j})\varepsilon^q\hat{I}_{1j}\hat{I}_{2j}$$
$$+\frac{1}{N\sqrt{N}}\sum_{j=1}^{N}\frac{D_j}{P(Z_j)}e_1[\hat{M}^q_{pN}(T_{0j})]^{-1}H^{q\prime}T_p(T_{0j})'(T_{0j})W^q(T_{0j})\overline{T}_{\overline{p}}(T_{0j})\overline{\beta}^*_{\overline{p}}(T_{0j})\hat{I}_{1j}\hat{I}_{2j}$$
$$+\frac{1}{N\sqrt{N}}\sum_{j=1}^{N}\frac{D_j}{P(Z_j)}e_1[\hat{M}^q_{pN}(T_{0j})]^{-1}H^{q\prime}T_p'(T_{0j})W^q(T_{0j})r_{\overline{p}}(T_{0j})\hat{I}_{1j}\hat{I}_{2j}.$$

C.3.1 First Term:

Our goal in this section is to show that

$$\frac{1}{N\sqrt{N}}\sum_{j=1}^{N}\sum_{i=1}^{N_1}\frac{D_j}{P(Z_j)}e_1[\hat{M}^q_{pN}(T_{0j})]^{-1}\left[\left(\frac{T_{1i}-T_{0j}}{h_{Nq}}\right)^{Q_p}\right]' h_{Nq}^{-2}\varepsilon_i^q K^q\left(\frac{T_{1i}-T_{0j}}{h_{Nq}}\right)\hat{I}_{1j}\hat{I}_{2j}$$
$$-\frac{1}{N\sqrt{N}}\sum_{j=1}^{N}\sum_{i=1}^{N_1}\frac{D_j}{P(Z_j)}e_1[M^q_{pN}(T_{0j})]^{-1}\left[\left(\frac{T_{1i}-T_{0j}}{h_{Nq}}\right)^{Q_p}\right]' h_{Nq}^{-2}\varepsilon_i^q K^q\left(\frac{T_{1i}-T_{0j}}{h_{Nq}}\right)I_{1j}I_{2j} = o_p(1).$$

Define $\gamma^q_{N0}(T_{0j}) = e_1[M^q_{pN}(T_{0j})]^{-1}$, $\hat{\gamma}^q_N(T_{0j}) = e_1[\hat{M}^q_{pN}(T_{0j})]^{-1}$. Let

$$\Gamma^q_N \quad = \quad \{\gamma^q_N : \sup\{||\gamma^q_N(x,z)-e_1[M^q_{pN}(x,z)]^{-1}|| \le \epsilon_{\gamma q}\} : (x,z)\in\overline{A}_1\}.$$

Next, we define

$$\mathcal{G}^q_{1N} := \Bigg\{ g_N : g_N(\varepsilon_i^q, X_i, Z_i; D_j, X_j, Z_j)$$
$$= N^{-3/2}\gamma^q_N(T_{0j})\frac{D_j}{P(Z_j)}\left[\left(\frac{T_{1i}-T_{0j}}{h_{Nq}}\right)^{Q_p}\right]' h_{Nq}^{-2}\varepsilon_i^q K^q\left(\frac{T_{1i}-T_{0j}}{h_{Nq}}\right)\tilde{I}_{1j}\tilde{I}_{2j}\Bigg\}$$

$$g_{N0}(\varepsilon_i^q, X_i, Z_i; D_j, X_j, Z_j) = N^{-3/2}\gamma^q_{N0}(T_{0j})\frac{D_j}{P(Z_j)}\left[\left(\frac{T_{1i}-T_{0j}}{h_{Nq}}\right)^{Q_p}\right]' h_{Nq}^{-2}\varepsilon_i^q K^q\left(\frac{T_{1i}-T_{0j}}{h_{Nq}}\right)I_{1j}I_{2j}$$

with $\tilde{I}_1\in\mathcal{I}_1$, $\tilde{I}_2\in\mathcal{I}_2$ and $\gamma^q_N\in\Gamma^q_N$, and $\mathcal{I}_1,\mathcal{I}_2$ as in Appendix C.4. We are going to try to show that the process $\sum_{i=1}^{N}\sum_{j=1}^{N} g_N(\varepsilon_i^q, X_i, Z_i; D_j, X_j, Z_j)$ is equicontinuous over $\mathcal{G}^q_{1N}$ in a neighborhood of $g_{N0}(\varepsilon_i^q, X_i, Z_i; D_j, X_j, Z_j)$, and that $\hat{g}_N(\varepsilon_i^q, X_i, Z_i; D_j, X_j, Z_j)$ lies in the neighborhood over which we establish equicontinuity with probability approaching to 1.

The equicontinuity lemma is applicable to symmetric, degenerate U-processes only, and $\sum_{i=1}^{N}\sum_{j=1}^{N} g_N(\varepsilon_i^q, X_i, Z_i; D_j, X_j, Z_j)$ is neither symmetric nor degenerate. But we can analyze this process in multiple steps. First, note that $\sum_i\sum_j g_N(\varepsilon_i^q, X_i, Z_i; D_j, X_j, Z_j) = \sum_i g_N(\varepsilon_i^q, D_i, X_i, Z_i)$ $+\sum_i\sum_{j\neq i} g_N(\varepsilon_i^q, X_i, Z_i; D_j, X_j, Z_j)$.

The first piece is a symmetric, but not degenerate U-process. To remedy this consider first

$$\begin{aligned}\tilde{g}_N(\varepsilon_i^q, D_i, X_i, Z_i) &:= N^{-3/2}\gamma_N^q(T_{0i})\frac{D_i}{P(Z_i)}\left[\left(\frac{T_{1i}-T_{0i}}{h_{Nq}}\right)^{Q_p}\right]' h_{Nq}^{-2}\varepsilon_i^q K^q\left(\frac{T_{1i}-T_{0i}}{h_{Nq}}\right)\tilde{I}_{1i}\tilde{I}_{2i} \\ &- E\left\{N^{-3/2}\gamma_N^q(T_{0i})\frac{D_i}{P(Z_i)}\left[\left(\frac{T_{1i}-T_{0i}}{h_{Nq}}\right)^{Q_p}\right]' h_{Nq}^{-2}\varepsilon_i^q K^q\left(\frac{T_{1i}-T_{0i}}{h_{Nq}}\right)\tilde{I}_{1i}\tilde{I}_{2i}\right\}\end{aligned}$$

This process is degenerate, and satisfies all the conditions of the lemma 3 of Heckman, Ichimura and Todd. But it is only one piece of the $\sum_{i=1}^N g_N(\varepsilon_i^q, D_i, X_i, Z_i)$ process. The other piece is

$$\begin{aligned}&\sum_{i=1}^N E\left\{N^{-3/2}\gamma_N^q(T_{0i})\frac{D_i}{P(Z_i)}\left[\left(\frac{T_{1i}-T_{0i}}{h_{Nq}}\right)^{Q_p}\right]' h_{Nq}^{-2}\varepsilon_i^q K^q\left(\frac{T_{1i}-T_{0i}}{h_{Nq}}\right)\tilde{I}_{1i}\tilde{I}_{2i}\right\} \\ &= N^{-1/2}E\left\{\gamma_N^q(T_{0i})\frac{D_i}{P(Z_i)}\left[\left(\frac{T_{1i}-T_{0i}}{h_{Nq}}\right)^{Q_p}\right]' h_{Nq}^{-2}\varepsilon_i^q K^q\left(\frac{T_{1i}-T_{0i}}{h_{Nq}}\right)\tilde{I}_{1i}\tilde{I}_{2i}\right\}\end{aligned}$$

We have to make sure that the limit of this is 0. We know that any element of $\left[\frac{T_{1i}-T_{0i}}{h_{Nq}}^{Q_p}\right]' K^q((T_{1i}-T_{0i})/(h_{Nq}))$ is bounded by $C_1 h_{Nq}^{-2} I\{||T_{1i}-T_{0i}|| \leq C_2 h_{Nq}\}$ for some finite C_1 and C_2. On the other hand, $|D_i| \leq 1$, $|\tilde{I}_{1i}\tilde{I}_{2i}| \leq 1$ $E|\varepsilon_i^q| < \infty$, and $P(Z_i)$ is almost surely bounded away from 0. Combining these facts with $Nh_{Nq}^4 \to \infty$, we get that the desired limit is in fact 0.

Next, we focus on the part containing different indices. Let $S_i := (\varepsilon_i^q, D_i, X_i, Z_i)$. Define

$$\begin{aligned}g_N^0(S_i, S_j) &= \frac{1}{2}N^{-3/2}\gamma_N^q(T_{0j})\frac{D_j}{P(Z_j)}\left[\left(\frac{T_{1i}-T_{0j}}{h_{Nq}}\right)^{Q_p}\right]' h_{Nq}^{-2}\varepsilon_i^q K^q\left(\frac{T_{1i}-T_{0j}}{h_{Nq}}\right)\tilde{I}_{1j}\tilde{I}_{2j} \\ &+ \frac{1}{2}N^{-3/2}\gamma_N^q(T_{0i})\frac{D_i}{P(Z_i)}\left[\left(\frac{T_{1j}-T_{0i}}{h_{Nq}}\right)^{Q_p}\right]' h_{Nq}^{-2}\varepsilon_j^q K^q\left(\frac{T_{1j}-T_{0i}}{h_{Nq}}\right)\tilde{I}_{1i}\tilde{I}_{2i}\end{aligned}$$

Define $\phi_N(S_i) = E[g_N^0(S_i, S_j)|S_i]$. Then,

$$\phi_N(S_i) = \tfrac{1}{2}N^{-3/2}h_{Nq}^{-2}E\left\{\gamma_N(T_{0j})\tfrac{D_j}{P(Z_j)}\left[\left(\tfrac{T_{1i}-T_{0j}}{h_{Nq}}\right)^{Q_p}\right]'\varepsilon_i^q K^q\left(\tfrac{T_{1i}-T_{0j}}{h_{Nq}}\right)\tilde{I}_{1j}\tilde{I}_{2j}|\varepsilon_i^q, D_i, X_i, Z_i\right\}$$

This is of the form $\varphi_N(T_{1i})\varepsilon_i^q$, and $E\big(\varphi_N(T_{1i})\varepsilon_i^q\big) = E\Big[\varphi_N(T_{1i})E(\varepsilon_i^q|T_{1i})\Big] = 0$. Thus we can define $\tilde{g}_N^0(S_i, S_j) := g_N^0(S_i, S_j) - \phi_N(S_i) - \phi_N(S_j)$. The process $\sum_i\sum_{j\neq i}\tilde{g}_N^0(S_i, S_j)$ is a degenerate U-process of order two. On the other hand, the above calculations show that $\sum_i 2(N-1)\phi_N(S_i)$ is a degenerate order one process. Since $|D_i| \leq 1$ and $P(Z_i)$ is bounded away from 0, and $\tilde{I}_{1i}\tilde{I}_{2i} \leq \tilde{I}_{1i} \leq I_{1i}^* := 1\{f_{X,Z}(X_i, Z_i) \geq q_{01} - \epsilon_{f1}\}$ the same steps as on p. 287 of HIT prove that each of these processes satisfies the first two conditions of the equicontinuity lemma. For the third condition, take any $g_N^{(1)}, g_N^{(2)} \in \mathcal{G}_{1N}$.

$$|g_N^{(1)} - g_N^{(2)}| = \left| N^{-3/2} \frac{D_j}{P(Z_j)} h_{Nq}^{-2} \varepsilon_i^q K\left(\frac{T_{1i}-T_{0j}}{h_{Nq}}\right)\right|$$
$$\times \left| \gamma_N^{(1)}(T_{0j}) \left[\left(\frac{T_{1i}-T_{0j}}{h_{Nq}}\right)^{Q_p}\right]' \tilde{I}_{1j}^{(1)} \tilde{I}_{2j}^{(1)} - \gamma_N^{(2)}(T_{0j}) \frac{D_j}{P(Z_j)} \left[\left(\frac{T_{1i}-T_{0j}}{h_{Nq}}\right)^{Q_p}\right]' \tilde{I}_{1j}^{(2)} \tilde{I}_{2j}^{(2)} \right|$$
$$\leq \left| N^{-3/2} \frac{D_j}{P(Z_j)} h_{Nq}^{-2} \varepsilon_i^q K\left(\frac{T_{1i}-T_{0j}}{h_{Nq}}\right)\right| \left| \left(\gamma_N^{(1)}(T_{0j}) - \gamma_N^{(2)}(T_{0j})\right) \left[\left(\frac{T_{1i}-T_{0j}}{h_{Nq}}\right)^{Q_p}\right]' \tilde{I}_{1j}^{(1)} \tilde{I}_{2j}^{(1)} \right|$$
$$+ \left| N^{-3/2} \frac{D_j}{P(Z_j)} h_{Nq}^{-2} \varepsilon_i^q K\left(\frac{T_{1i}-T_{0j}}{h_{Nq}}\right)\right| \left| \gamma_N^{(2)}(T_{0j}) \left[\left(\frac{T_{1i}-T_{0j}}{h_{Nq}}\right)^{Q_p}\right]' \tilde{I}_{2j}^{(1)} \right| \left| \tilde{I}_{1j}^{(1)} - \tilde{I}_{1j}^{(2)} \right|$$
$$+ \left| N^{-3/2} \frac{D_j}{P(Z_j)} h_{Nq}^{-2} \varepsilon_i^q K\left(\frac{T_{1i}-T_{0j}}{h_{Nq}}\right)\right| \left| \gamma_N^{(2)}(T_{0j}) \left[\left(\frac{T_{1i}-T_{0j}}{h_{Nq}}\right)^{Q_p}\right]' \tilde{I}_{1j}^{(2)} \right| \left| \tilde{I}_{2j}^{(1)} - \tilde{I}_{2j}^{(2)} \right|$$

Therefore, the third condition of the equicontinuity lemma will hold, if each of the families $\Gamma_N, \mathcal{I}_1, \mathcal{I}_2$ satisfy it. By Assumption (4.3(b)) and 4 of HIT, Γ_N satisfies this condition. On the other hand, in section (C.4), we verify this condition for $\mathcal{I}_1$.

Combining all these results, we conclude that the process $\sum_{j=1}^N \sum_{i=1}^N g_N(\varepsilon_i^q, T_{1i}, T_{0j})$ is equicontinuous over $\mathcal{G}_{1N}$ in a neighborhood of $g_{N0}(\varepsilon_i^q, T_{1i}, D_j, T_{0j}, X_j, Z_j)$.

Lemma 5 and 6 of HIT can be used to show that $\sup_{(x,z)\in A_1\cap A_2} ||\hat{M}_{pN}(h_0(x,P(z)),P(z)) - M_{pN}(h_0(x,P(z)),P(z))|| \to 0$. This result combined with the arguments at the beginning shows that $\hat{g}_N(\varepsilon_i^q, T_{1i}, T_{0j})$ lies in the neighborhood of $g_{N0}(\varepsilon_i^q, T_{1i}, T_{0j}, X_j, Z_j)$ over which equicontinuity was shown.

C.3.2 Second Term:

Next, we look at

$$\frac{1}{\sqrt{N}} \sum_{j=1}^N D_j e_1 [\hat{M}_{pN}(T_{0j})]^{-1} N^{-1} H' T_p(T_{0j})'(T_{0j}) W(T_{0j}) \overline{T}_{\overline{p}}(T_{0j}) \overline{\beta}_{\overline{p}}^*(T_{0j}) \hat{I}_{1j} \hat{I}_{2j}$$

Fix the evaluation point (d_0, x_0, z_0) such that $(x_0, z_0) \in A_1 \cap A_2$. Let $P_0 = P(z_0)$, $t_0 = (h_0(x_0, P(z_0)), P(z_0))$. Then each term in this sum equals:

$$e_1 H [\hat{M}_{pN}(t_0)]^{-1} N^{-1} \frac{d_0}{P_0} H' T_p'(t_0) W(t_0) \overline{T}_{\overline{p}}(t_0) \overline{\beta}_{\overline{p}}^*(t_0) \hat{I}_{10} \hat{I}_{20}$$
$$= e_1 [\hat{M}_{pN}(t_0)]^{-1} \sum_{s=p+1}^{\overline{p}} N^{-1} \frac{1}{h_{Nq}^2} \sum_{i=1}^{\overline{p}} \frac{d_0}{P_0} \left[\left(\frac{T_{1i}-t_0}{h_{Nq}}\right)^{Q_p}\right]' (T_{1i}-t_0)^{Q(s)} [m^{(s)}(t_0)]' K\left(\frac{T_{1i}-t_0}{h_{Nq}}\right) \hat{I}_{10} \hat{I}_{20}$$

We add and subtract

$$e_1[\hat{M}_{pN}(t_0)]^{-1} \sum_{s=p+1}^{\overline{p}} \frac{1}{h_{Nq}^2} E\left\{ \frac{D_j}{P(Z_j)} \left[\frac{T_{1i}-T_{0j}}{h_{Nq}} \quad Q_p \right]' (T_{1i}-T_{0j})^{Q(s)} K \quad \frac{T_{1i}-T_{0j}}{h_{Nq}} \quad |T_{0j}=t_0, X_j=x_0, Z_j=z_0 \right\} [m^{(s)}(t_0)]' \hat{I}_{10} \hat{I}_{20}$$

This gives us three terms. But the difference of the two terms is handled in the same way as in lemma 2. In particular, we take γ_n and Γ_n as before and define

$$g_N(T_{1i},T_{0j},D_j,X_j,Z_j) = N^{-3/2}\gamma_N(T_{0j})h_{Nq}^{-2}$$
$$\times \Big(\tfrac{D_j}{P(Z_j)}\Big[\Big(\tfrac{T_{1i}-T_{0j}}{h_{Nq}}\Big)^{Q_p}\Big]'(T_{1i}-T_{0j})^{Q(s)}K\Big(\tfrac{T_{1i}-T_{0j}}{h_{Nq}}\Big) - E\Big[\tfrac{D_j}{P(Z_j)}\Big[\Big(\tfrac{T_{1i}-T_{0j}}{h_{Nq}}\Big)^{Q_p}\Big]'(T_{1i}-T_{0j})^{Q(s)}K\Big(\tfrac{T_{1i}-T_{0j}}{h_{Nq}}\Big)\Big|T_{0j},X_j,Z_j\Big]\Big)$$
$$\times[m^{(s)}(T_{0j})]'\tilde{I}_{1j}\tilde{I}_{2j}$$

Let$\hat{g}_N(T_{1i},T_{0j},X_j,Z_j)$ and $g_{N0}(T_{1i},T_{0j},X_j,Z_j)$ be defined in the same way as before. Moreover, let $\mathcal{G}_{2N} := \{g_n(T_{1i},T_{0j},X_j,Z_j)|\gamma_N(T_{0j}) \in \Gamma_N\}$. Then going through the same steps as in lemma 2 we can show that $1/\sqrt{N}\sum_{j=1}^N \hat{R}_{21}(T_{0j},D_j) = \sum_{i=1}^N\sum_{j=1}^N[\hat{g}_N(T_{1i},T_{0j},D_j,X_j,Z_j) - g_{N0}(T_{1i},T_{0j},D_j,X_j,Z_j)] = o_p(1)$.
Then we deal with the term

$$e_1[\hat{M}_{pN}(t_0)]^{-1}\sum_{s=p+1}^{\overline{p}}\tfrac{1}{h_{Nq}^2}E\Big\{\tfrac{D_j}{P(Z_j)}\Big[\Big(\tfrac{T_{1i}-T_{0j}}{h_{Nq}}\Big)^{Q_p}\Big]'(T_{1i}-T_{0j})^{Q(s)}K\Big(\tfrac{T_{1i}-T_{0j}}{h_{Nq}}\Big)|T_{0j}=t_0,X_j=x_0,Z_j=z_0\Big\}[m^{(s)}(t_0)]'\hat{I}_{10}\hat{I}_{20}$$

which in turn equals

$$e_1\big([\hat{M}_{pN}(t_0)]^{-1}-[M_p(t_0)]^{-1}\big)$$
$$\cdot\sum_{s=p+1}^{\overline{p}}\tfrac{1}{h_{Nq}^2}E\Big\{\tfrac{D_j}{P(Z_j)}\Big[\Big(\tfrac{T_{1i}-T_{0j}}{h_{Nq}}\Big)^{Q_p}\Big]'(T_{1i}-T_{0j})^{Q(s)}K\Big(\tfrac{T_{1i}-T_{0j}}{h_{Nq}}\Big)|T_{0j}=t_0,X_j=x_0,Z_j=z_0\Big\}[m^{(s)}(t_0)]'\hat{I}_{10}\hat{I}_{20}$$
$$+e_1[M_p(t_0)]^{-1}\sum_{s=p+1}^{\overline{p}}\tfrac{1}{h_{Nq}^2}E\Big\{\tfrac{D_j}{P(Z_j)}\Big[\Big(\tfrac{T_{1i}-T_{0j}}{h_{Nq}}\Big)^{Q_p}\Big]'(T_{1i}-T_{0j})^{Q(s)}K\Big(\tfrac{T_{1i}-T_{0j}}{h_{Nq}}\Big)|T_{0j}=t_0,X_j=x_0,Z_j=z_0\Big\}[m^{(s)}(t_0)]'\hat{I}_{10}\hat{I}_{20}$$

The first expression can be treated in the same way as in lemma 2. If $t_0 = (h_0(x_0,P(z_0)),P(z_0))$, the last expression equals

$$e_1[M_p(t_0)]^{-1}\sum_{s=p+1}^{\overline{p}}\tfrac{1}{h_{Nq}^2P(z_0)}E\Big\{D_j\Big[\Big(\tfrac{T_{1i}-t_0}{h_{Nq}}\Big)^{Q_p}\Big]'(T_{1i}-t_0)^{Q(s)}K\Big(\tfrac{T_{1i}-t_0}{h_{Nq}}\Big)|T_{0j}=t_0,X_j=x_0,Z_j=z_0\Big\}[m^{(s)}(t_0)]'\hat{I}_{10}\hat{I}_{20}$$
$$=e_1[M_p(t_0)]^{-1}\sum_{s=p+1}^{\overline{p}}\tfrac{1}{h_{Nq}^2P(z_0)}E(D_j|Z_j=z_0)E\Big\{\Big[\Big(\tfrac{T_{1i}-t_0}{h_{Nq}}\Big)^{Q_p}\Big]'(T_{1i}-t_0)^{Q(s)}K\Big(\tfrac{T_{1i}-t_0}{h_{Nq}}\Big)\Big\}[m^{(s)}(t_0)]'$$
$$=h_{Nq}^{\overline{p}}e_1[M_p(t_0)]^{-1}\sum_{s=p+1}^{\overline{p}}\Big[\int u^{Q(0)}\cdot u^{Q(s)}m^{(s)}(t_0)'\cdot u^{Q(\overline{p}-1)}K(u)\mathrm{d}u,\ldots,\int u^{Q(p)}\cdot u^{Q(s)}m^{(s)}(t_0)'\cdot u^{Q(\overline{p}-s)}K(u)\mathrm{d}u\Big]f^{(\overline{p}-s)}(t_0)'$$

We need

$$\mathrm{plim}_{N\to\infty}1/\sqrt{N}\sum_{i=1}^N h_{Nq}^{\overline{p}}e_1[M_p(T_{0i})]^{-1}\times$$
$$\sum_{s=p+1}^{\overline{p}}\left[\int u^{Q(0)}\cdot u^{Q(s)}m^{(s)}(T_{0i})'\cdot u^{Q(\overline{p}-1)}K(u)\mathrm{d}u,\ldots,\int u^{Q(p)}\cdot u^{Q(s)}m^{(s)}(T_{0i})'\cdot u^{Q(\overline{p}-1)}K(u)\mathrm{d}u\right]$$
$$\times f^{(\overline{p}-s)}(T_{0i})' = b_q < \infty$$

All the terms involving $T_{0i} = (h_0(X_i,P(Z_i)),P(Z_i))$ are bounded with probability 1. Thus, if $Nh_{Nq}^{2\overline{p}} \to c < \infty$ then we are OK.

C.3.3 Third Term:

We claim that under our assumptions, for each evaluation point (d_0, x_0, z_0) such that $(x_0, z_0) \in A_1 \cap A_2$,

$$e_1[\hat{M}_{pN}(t_0)]^{-1}N^{-1}\frac{d_0}{P(z_0)}H'T_p'(t_0)W(t_0)r_{\overline{p}+1}(t_0)\hat{I}_{10}\hat{I}_{20} = o_p(h_{Nq}^{\overline{p}+1})$$

But, as in lemma 8 of HIT,

$$\begin{aligned} & N^{-1}||\frac{d_0}{P(z_0)}H'T_p'(t_0)W(t_0)r_{\overline{p}+1}(t_0)\hat{I}_{10}\hat{I}_{20}|| \leq N^{-1}h_{Nq}^{(\overline{p}+1)} \\ \times & \left|\left|\sum_{i\in I_1}\frac{d_0}{P(z_0)}\left[\left(\frac{T_{1i}-t_0}{h_{Nq}}\right)^{Q_p}\right]'\left(\frac{T_{1i}-t_0}{h_{Nq}}\right)^{Q(\overline{p}+1)}[m^{(\overline{p}+1)}(t_i)-m^{(\overline{p}+1)}(t_0)]\frac{1}{h_{Nq}^2}K\left(\frac{T_{1i}-t_0}{h_{Nq}}\right)\right|\right| \\ \leq & N^{-1}o(h_{Nq}^{\overline{p}+1})\sum_{i=1}^{N}\left|\left|\left[\left(\frac{T_{1i}-t_0}{h_{Nq}}\right)^{Q_p}\right]'\left(\frac{T_{1i}-t_0}{h_{Nq}}\right)^{Q(\overline{p}+1)}\frac{1}{h_{Nq}^2}K\left(\frac{T_{1i}-t_0}{h_{Nq}}\right)\right|\right| = o_p(h_{Nq}^{\overline{p}+1}) \end{aligned}$$

By lemma 5 of HIT, for any t_0 such that $f_{h_1(X,P(Z)),P(Z)}(t_0) > 0$, for sufficiently large N, $\hat{M}_{pN}(t_0)$ will be nonsingular. Therefore, every element of the matrix $[\hat{M}_{pN}(t_0)]^{-1}$ has finite norm.

C.3.4 Conclusion:

$$\tfrac{1}{\sqrt{N}}\textstyle\sum_{j=1}^{N}D_j\big(\hat{q}(h_{0j},P_j)-q(h_{0j},P_j)\big)\hat{I}_{1j}\hat{I}_{2j} =^{AE}$$
$$\tfrac{1}{\sqrt{N}N}\textstyle\sum_{j=1}^{N}\sum_{i=1}^{N}\frac{D_j}{P(Z_j)}[M_{pN}(h_{0j},P_j)]^{-1}\left[\left(\frac{(h_{1i},P_i)-(h_{0j},P_j)}{h_{Nq}}\right)^{Q_p}\right]'\frac{1}{h_{Nq}^2}K\left(\frac{(h_{1i},P_i)-(h_{0j},P_j)}{h_{Nq}}\right)\varepsilon_i^q I_{1j}I_{2j}+b_q$$

where $\varepsilon_i^q = D_iY_i - E\big[D_iY_i|h_1\big(X_i,P(Z_i)\big),P(Z_i)\big]$

C.4 Issues in Trimming

The estimation method in this paper uses two trimming functions. One of the trimming functions is based on the estimated density $\hat{f}_{X,Z}(x,z)$. The other one is based on $\hat{f}_{\hat{h}_1,\hat{P}}$. The first part of this appendix shows that the family of functions that contains the first trimming function has an envelope and satisfies the conditions of the equicontinuity lemma. The second part of the appendix verifies the same condition for the trimming function which is based on the kernel density estimator of $f_{h_1,P}$. We observe the values of (X,Z). Suppose the support of (X,Z) is a connected[29] subset of $\mathbb{R}^d$, with $d < \infty$. Also let

$$\begin{aligned}
\mathcal{H}_1 &:= \{f : \sup_{x,z} |f(x,z) - f_{X,Z}(x,z)| \le \epsilon_{f1},\ f \ \ has \ \ smoothness \ \ q > d,\ \inf_{(x,z)\in\overline{A}_q} ||Df(x,z)|| \ge \underline{\theta}_1\} \\
&\cap \ \{f : \sup_{x,z} |f_j(x,z) - (f_{X,Z})_j(x,z)| \le \epsilon^j_{f1},\} \\
\mathcal{I}_1 &:= \{I((x,z) \in \tilde{A}_1) : \tilde{A} = \{(x,z) : f(x,z) \ge q_{01}\} \ for \ some \ f \in \mathcal{H}_1\}
\end{aligned}$$

$$\begin{aligned}
\overline{A}_1 &:= \{(x,z) : f_{X,Z}(x,z) \ge q_{01} - \epsilon_{f1}\} \\
\overline{A}_q &:= \{(x,z) : q_{01} + \epsilon_{f1} \ge f_{X,Z}(x,z) \ge q_{01} - \epsilon_{f1}\} \\
A_1 &:= \{(x,z) : f_{X,Z}(x,z) \ge q_{01}\}
\end{aligned}$$

where $\underline{\theta}_1 > 0$, and the subscripts denote the j^{th} partial derivative of the associated function. First we observe that under the assumptions of Silverman's Theorem A on $f_{X,Z}$, the kernel function and the bandwidth sequence used to estimate this density function

$$\sup_{(x,z)\in \text{supp}(X,Z)} \left|\hat{f}(x,z) - f_{X,Z}(x,z)\right| \to 0 \ a.s. \qquad \sup_{(x,z)\in \text{supp}(X,Z)} \left|\hat{f}_j(x,z) - f_{X,Z,j}(x,z)\right| \to 0 \ a.s.$$

for $j \in \{1, ..., d\}$, where the subscript j denotes the j^{th}partial derivative. Using this result, we can claim that $1_{\overline{A}}(s)$ is an envelope for $\mathcal{I}_1$.

Next, we verify that $\mathcal{I}_1$ satisfies the third condition of the equicontinuity lemma. Our arguments will rely on a lemma due to Kolmogorov and Tihomirov.

Definition C.1 *Let E be a connected compact subset of a finite dimensional Banach space. Suppose the metric dimension of E equals d. Consider bounded real valued functions on E with norm equal $||f|| = \sup_{s\in E} |f(s)|$. A function in this space has smoothness $q > 0$ ($q = p + a$, p an integer, $0 < a \le 1$) if for arbitrary vectors $s \in E$, $s + h \in E$, we have*

$$f(s+h) = \sum_{k=0}^{p} \frac{1}{k!} B_k(h,s) + R(h,s) \tag{13}$$

where $B_k(h,s)$ is a homogeneous form of degree k in h and

$$|R(h,s)| \le C||h||^q$$

where C is a certain constant. All of the functions f that satisfy (C.1) and (13) with a fixed constant C form the class $F^E_q(C)$.

[29] Connectedness of the support is needed for the application of the Mean Value Theorem. If the support is not connected but could be partitioned into finitely many connected subsets, then the argument can be made for each connected piece separately.

Lemma C.2 ***(Kolmogorov-Tihomirov Lemma):*** *For every set $B \subset F_q^E(C)$ bounded in the sense of the metric as in Definition (C.1), we have $\left(\frac{1}{\varepsilon}\right)^{d/q} = O(\mathcal{C}_\varepsilon(B))$, where $\mathcal{C}_\varepsilon(B) = \log_2 D(\varepsilon, B)$, and $D(\varepsilon, B)$ is the ε packing number (in the norm as in Definition (C.1)) of the set B.*

By this lemma, we know that the class of functions which consists of restrictions of elements of $\mathcal{H}_1$ to $\overline{A}_1$ satisfies the third condition of the equicontinuity lemma in the sup norm. Using this information, we need to verify that $\mathcal{I}_1$ also satisfies that condition. Before we start with that proof, let us show thew following preliminary result:

Claim C.1 *$\overline{A}_1$ and $\overline{A}_q$ are compact.*

Proof: The sets $[q_0 - \epsilon_f, \infty)$ and $[q_{01} - \epsilon_{f1}, q_{01} + \epsilon_{f1}]$ are closed in $\mathbb{R}$ when it is equipped with the Euclidean (absolute value) metric. Since $f_{X,Z}$ is continuous, this implies that $\overline{A}_1$ and $\overline{A}_q$ are closed in $(\mathbb{R}^d, ||\cdot||)$. If we show that $\overline{A}_1$ is also bounded, we will have shown that both $\overline{A}_1$ and $\overline{A}_q$ are compact. Suppose toward a contradiction $\overline{A}_1$ is unbounded. This means that for each $J > 0$, $\mathbb{R}^d \setminus B_J(0)$ must contain infinitely many elements of $\overline{A}_1$. We will pick a sequence of elements of $\overline{A}_1$ recursively. Pick $s_1 \in B_1(0) \cap \overline{A}_1$. Suppose for $j \geq 2$, we have already picked $s_1, ..., s_{j-1}$. Then pick $s_j \in (B_j(0) \setminus B_{j-1/2}(0)) \cap \overline{A}_1$. The $\{s_j\}_{j=1}^\infty$ is an unbounded sequence contained in $\overline{A}_1$, such that $||s_i - s_j|| > 1/2$ whenever $i \neq j$. Let $r := (q_{01} - \epsilon_{f1})/2$. Since $f_{X,Z}$ is uniformly continuous, there exists $\tilde{\nu} > 0$ such that $||t - s|| < \tilde{\nu} \Rightarrow |f_{X,Z}(t) - f_{X,Z}(s)| < r$. Let $\nu := \min\{\tilde{\nu}, 1/2\}$. Then for each j, $t \in B_\nu(s_j) \Rightarrow f_{X,Z}(t) > (q_{01} - \epsilon_{f1})/2$. Moreover, for $i \neq j$, $B_\nu(s_i) \cap B_\nu(s_j) = \emptyset$. Then

$$1 \geq P\left(\cup_{j=1}^\infty B_\nu(s_j)\right) = \sum_{j=1}^\infty P(B_\nu(s_j)) > (q_{01} - \epsilon_{f1})/2 \sum_{j=1}^\infty Leb(B_\nu(s_j))$$

which is a contradiction. ■

Remark C.1 *The above arguments also imply that*

1. *$\overline{A}_q$ has finitely many disjoint components*[30]*, so that we could write $\overline{A}_q = \cup_{l=1}^L Q_l$, with $d_H(Q_l, Q_k) > 0$ for $l \neq k$,*
2. *there exists $M_1 \in \mathbb{R}$ such that $\forall f \in \mathcal{H}_1$, $\sup_{(x,z)\in\overline{A}_1} |f(x,z)| \leq M_1$, and*
3. *there exists $M_2 \in \mathbb{R}$ such that $\forall f \in \mathcal{H}_1$, $\sup_{(x,z)\in\overline{A}_1} ||Df(x,z)|| \leq M_2$.*

Claim C.2 *For any $f, g \in \mathcal{H}_1$, such that $\sup_{(x,z)\in\overline{A}_1} |f(x,z) - g(x,z)| < \eta \leq \epsilon_{f1}$, and for $\delta = \eta/\underline{\theta}_1$,*

$$\{f \geq q_0 > g\} \subseteq \big(\{s : f(s) = q_0\} \oplus B_\delta(0)\big) \cap \overline{A}_q$$

where $A \oplus B := \{a + b : a \in A,\ b \in B\}$, and $B_\delta(0)$ denotes the ball around 0 with radius δ.

[30] A component of a set E is a connected subset $E_0 \subseteq E$ such that there is no connected set in E containing E_0 other than E_0 itself.

Proof: First, consider $s \in \{f \geq q_0 > g\}$. $f \in \mathcal{H}_1 \Rightarrow f(s) - \epsilon_{f1} \leq f_{X,Z}(s) \leq f(s) + \epsilon_{f1}$. Combining this with $f(s) \geq q_{01}$ we get $f_S(s) \geq q_{01} - \epsilon_{f1}$. Similarly, $g \in \mathcal{H}_1 \Rightarrow g(s) - \epsilon_{f1} \leq f_{X,Z}(s) \leq g(s) + \epsilon_{f1}$. And combining this with $g(s) < q_{01}$ we get $f_{X,Z}(s) \leq q_{01} + \epsilon_{f1}$. Thus, $f_{X,Z}(s) \in [q_{01} - \epsilon_{f1}, q_{01} + \epsilon_{f1}]$. This shows that $\{f \geq q_0 > g\} \subseteq \overline{A}_q$. If $\overline{A}_q \setminus (\{x : f(x) = q_{01}\} \oplus B_\delta(0)) = \emptyset$, we have nothing more to prove. Else, consider any $s \in \overline{A}_q \setminus (\{x : f(x) = q_{01}\} \oplus B_\delta(0))$. Our goal is to show that $s \in \mathbb{R}^d \setminus \{f \geq q_0 > g\}$. Toward this goal, pick $u \in \{x : f(x) = q_{01}\}$ that is closest to s. Note that, since $||s - u||$ is a continuous function of u and $\{x : f(x) = q_{01}\}$ is compact, such a point must exist. Moreover, u must be in the same component of $\overline{A}_q$ as s, and the line segment joining u to s must be contained in that component of $\overline{A}_q$. Thus, $||Df(\tilde{u})|| > \underline{\theta}_1$ for any $\tilde{u}$ lying on the line segment between s and u. Using the mean value theorem, we know that $|f(s) - f(u)| = |f(s) - q_{01}| = |Df(\tilde{u}) \cdot (s - u)| = ||Df(\tilde{u})|| \cdot ||s - u|| \cdot |\cos\alpha|$, where α denotes the angle between s and u. But $||Df(\tilde{u})|| > \underline{\theta}_1$ by our previous arguments, and $|\cos\alpha| = 1$ because u is the member of the q_{01}-level set of f that is closest to s[31]. Combining these arguments, we get that $|f(s)| > q_{01} + \eta$. Note that if $f(s) < q_{01}$, there is nothing to prove. On the other hand, if $f(s) > q_{01} + \eta$, then $|f(s) - g(s)| < \eta$ implies that $g(s)$ must also be strictly larger than q_{01}, which means that s cannot belong to $\{f \geq q_0 > g\}$. ■

Claim C.3 *$\mathcal{I}_1$ satisfies the third condition of the equicontinuity lemma.*

Proof: Let $\tau > 0$ and let $\{f^1, ..., f^J\}$ be the maximal η-separated subset of $\mathcal{H}_1$ restricted to $\overline{A}_1$ in the sup norm, with $\tau > \eta > 0$. The relationship between η and τ will be clear later. For $j \in \{1, ..., J\}$, let $I^j := \{s : f^j(s) \geq q_{01}\}$. We are going to try to demonstrate that $\{I^1, ..., I^J\}$ is maximal τ-separated subset of $\mathcal{I}_1$ in the $\mathcal{L}^2$ norm. For this claim to be true it must be that for each element I of $\mathcal{I}_1$, there exists $j \in \{1, .., J\}$ such that

$$\int \left|I - I^j\right|^2 \mathrm{d}P = P(\{I \neq I^j\}) < \tau.$$

Note that

$$P(\{I \neq I^j\}) = P\left(\{I \neq I^j\}\right) + P\left(\{I \neq I^j\}\right) = P(\{f \geq q_0 > f^j\}) + P(\{f^j \geq q_0 > f\}),$$

where f is an element of $\mathcal{H}_1$ that is associated with I. Using claim (C.2), with $\delta = \eta / \underline{\theta}_1$,

$$\begin{aligned} P(\{f \geq q_0 > f^j\}) + P(\{f^j \geq q_0 > f\}) &\leq P\big((\{s : f(s) = q_0\} \oplus B_\delta(0)) \cap \overline{A}_q\big) \\ &+ P\big((\{s : f^j(s) = q_0\} \oplus B_\delta(0)) \cap \overline{A}_q\big) \end{aligned}$$

[31] To see this, let U^{su} be an open set containing the component of $\overline{A}_q$ that s and u belong to. Let $f|U^{su}$ denote the restriction of f to U^{su}. Now u solves $\min \sum_{i=1}^d (s_i - u_i)^2$ such that $u \in \{x : f|U^{su}(x) = q_{01}\}$. The first order conditions of this problem tell us that if $s \neq u$, then the $f_i(u)$ must be proportional to $s_i - u_i$. Reversing s and u in this problem says $f_i(s)$ must also be proportional to $s_i - u_i$. Since $\tilde{u}$ is on the line segment joining s and u, and since derivative is a linear operator, $f_i(\tilde{u})$ must also be proportional to $s_i - u_i$. This last fact tells us that $|\cos\alpha| = 1$.

$$= \sum_{l=1}^{L} P\big((\{s : f(s) = q_0\} \oplus B_\delta(0)) \cap Q_l\big) + \sum_{l=1}^{L} P\big((\{s : f^j(s) = q_0\} \oplus B_\delta(0)) \cap Q_l\big)$$

$$\leq M_1 \sum_{l=1}^{L} \Big(Leb\big((\{s : f(s) = q_0\} \oplus B_\delta(0)) \cap Q_l\big) + Leb\big((\{s : f^j(s) = q_0\} \oplus B_\delta(0)) \cap Q_l\big)\Big),$$

where we used the second part of Remark (C.1) to write the last inequality. On the other hand, using the last part of Remark (C.1), the compactness of $\overline{A}_q$ and the formula for surface areas of smooth, parametrized manifolds we could show that for each $f \in \mathcal{H}_1$, the $d-1$ dimensional volume of the smooth surface $\{s : f(s) = q_{01}\} \cap Q_l \leq R$ for some $R < \infty$. Then the last expression above is less than or equal to $\frac{4M_1LR\eta}{\underline{\theta}_1}$. Letting $\tau = \frac{4M_1LR\eta}{\underline{\theta}_1}$, we can conclude that τ-packing number for $\mathcal{I}_1$ in the $\mathcal{L}^2$ norm is the same as η-packing number for $\mathcal{H}_1$ in the sup norm[32]. Since by Kolmogorov-Tihomirov lemma, the latter satisfies the desired condition of the equicontinuity lemma, so does the former. ■

Next we turn to our trimming problem. We have to employ two trimming functions. The first function is needed to guarantee that the estimator $\hat{P(z)}$ is uniformly consistent for $E(D|Z)$. The second trimming function is needed because we need to have a uniformly consistent estimate for $E(DY|h_1(X, P(Z)), P(Z))$ evaluated at the value $(h_0(X, P(Z)), P(Z))$ takes. Our previous arguments the first trimming function satisfies the conditions of the equicontinuity lemma. We need to define a family of functions that will contain our second trimming function. For this purpose we define

$$\overline{B}_z = \{z \in \text{supp}(Z) : (x, z) \in \overline{A}_1, \text{for some } x \in \text{supp}(X)\}$$

and[33]

$$\begin{aligned}
\Psi_P &= \{g : \sup_{z \in \overline{B}_z} |g(z) - P(z)| \leq \epsilon_P,\ g \textit{ has smoothness } q > d,\ \inf_{z \in \overline{B}_{zq}} ||Dg(z)|| \geq \underline{\theta}_P\} \\
\Psi'_P &= \Psi_P \cap \{g : \sup_{z \in \overline{B}_z} |g(z) - P(z)| = o_P(\tilde{h}_{N2}^3)\} \\
\Psi_h &= \{\varphi : \sup_{\tilde{P} \in \Psi_P} \sup_{(x,z) \in \overline{A}_1} |\varphi(x, \tilde{P}(z)) - h_0(x, P(z))| \leq \epsilon_h,\ \varphi \textit{ has smoothness } q > d\} \\
&\cap \{\varphi : \sup_{\tilde{P} \in \Psi_P} \sup_{(x,z) \in \overline{A}_1} |\varphi(x, \tilde{P}(z)) - h_0(x, P(z))| = o_P(\tilde{h}_{N2}^3)\} \\
&\cap \{\varphi : \inf\{||D_x\varphi(x, \tilde{P}(z))|| : (x, z) \in \overline{A}_q, \tilde{P} \in \Psi_P\} \geq \underline{\theta}_{hx}\} \\
&\cap \{\varphi : \inf\{||D_P\varphi(x, \tilde{P}(z))|| : (x, z) \in \overline{A}_q, \tilde{P} \in \Psi_P\} \geq \underline{\theta}_{hP}\}
\end{aligned}$$

[32] Since $\sup|f - f^j| < \eta \rightarrow \sqrt{E|f - f^j|^2} < \eta$, the η-packing number for $\mathcal{H}_1$ in the sup norm is at least as large as the η-packing number for $\mathcal{H}_1$ in the $\mathcal{L}^2$ norm.

[33] In these definitions all the θ's are strictly greater than 0, and $\tilde{h}_{N2}$ denotes the smoothing parameter that is used in the trimmed kernel density estimation of $f_{h_1,P}$.

$$\begin{aligned}\mathcal{H}_2 &= \{f : \exists \tilde{I}_1 \in \mathcal{I}_1,\ \tilde{d}\left(f(\varphi,\tilde{P}), f_{h_1,P}(h_0,P)\right) \le \epsilon_{f2}\} \\ &\cap \ \{f : f \ \ has \ \ smoothness \ \ q > d, \inf_{(x,z,\tilde{P},\varphi)\in \overline{A}_q\times\Psi_P\times\Psi_h} ||Df(\varphi(x,\tilde{P}(z)),\tilde{P}(z))|| \ge \underline{\theta}_2\}\end{aligned}$$

where

$$\tilde{d}\big(f(\varphi,\tilde{P}),f_{h_1,P}(h_0,P)\big) := \sup_{(\varphi,\tilde{P})\in\Psi_h\times\Psi_P}\sup_{(x,z)\in\overline{A}_1}|f(\varphi(x,\tilde{P}(z)),\tilde{P}(z)) - f_{h_1,P}(h_0(x,P(z)),P(z))|$$

$$\mathcal{I}_2 = \{I((x,z)\in\tilde{A}_2) : \tilde{A}_2 = \{(x,z)\in\overline{A}_1 : f(\varphi(x,\tilde{P}(z)),\tilde{P}(z)) \ge q_{02}\} \text{ for some } {}_{f\in\mathcal{H}_2,\varphi\in\Psi_h,\tilde{P}\in\Psi_P}\}$$

Going through arguments similar to those above we can verify the third condition of the equicontinuity lemma for $\mathcal{I}_2$. But we still need to verify that $\tilde{d}\left(\hat{f}_{\hat{h}_1,\hat{P}}(\varphi,\tilde{P}), f_{h_1,P}(h_0,P)\right) \le \epsilon_{f2}$ for sufficiently large N with probability approaching to 1 [34]. To guarantee this, we need $\hat{h}_1(X_i,\hat{P}(Z_i))$ and $\hat{P}(Z_i)$ to be uniformly consistent for $h_1(X_i,P(Z_i))$ and $P(Z_i)$. However, this occurs only when the density of (X,Z) is bounded away from 0. Therefore, in the kernel density estimation of $f_{h_1,P}$ we have to trim out those observations at which $f_{X,Z}$ is very small. Let $\tilde{K}_2$ be a Lipschitz function with Lipschitz constant equal to M_3 [35] and define

$$\hat{f}_{\hat{h}_1,\hat{P}}\big(\varphi(x,\tilde{P}(z)),\tilde{P}(z)\big) := \frac{1}{N\tilde{h}_{N2}^2}\sum_{i=1}^N \tilde{K}_2\left(\frac{(\hat{h}_1(X_i,\hat{P}(Z_i)),\hat{P}(Z_i)) - (\varphi(x,\tilde{P}(z)),\tilde{P}(z))}{\tilde{h}_{N2}}\right)\hat{I}_1(X_i,Z_i)$$

$$\hat{f}_{h_1,P}\big(\varphi(x,\tilde{P}(z)),\tilde{P}(z)\big) := \frac{1}{N\tilde{h}_{N2}^2}\sum_{i=1}^N \tilde{K}_2\left(\frac{(h_1(X_i,P(Z_i)),P(Z_i)) - (\varphi(x,\tilde{P}(z)),\tilde{P}(z))}{\tilde{h}_{N2}}\right)\hat{I}_1(X_i,Z_i)$$

Adding and subtracting some terms yields

$$\begin{aligned}&|\hat{f}_{\hat{h}_1,\hat{P}}(\varphi(x,\tilde{P}(z)),\tilde{P}(z)) - f_{h_1,P}(h_0(x,P(z)),P(z))| \le \\ &\left|\hat{f}_{\hat{h}_1,\hat{P}}(\varphi(x,\tilde{P}(z)),\tilde{P}(z)) - \hat{f}_{\hat{h}_1,\hat{P}}(h_0(x,P(z)),P(z))\right| \\ &+\left|\hat{f}_{\hat{h}_1,\hat{P}}(h_0(x,P(z)),P(z)) - \hat{f}_{h_1,P}(h_0(x,P(z)),P(z))\right| \\ &+\left|\hat{f}_{h_1,P}(h_0(x,P(z)),P(z)) - f_{h_1,P}(h_0(x,P(z)),P(z))\right|\end{aligned} \tag{14}$$

Our goal is to show that each of the above terms is less than or equal to $\epsilon_{f2}/3$ with probability approaching to 1 for sufficiently large N. Let us start with the first one:

$$\begin{aligned}&|\hat{f}_{\hat{h}_1,\hat{P}}(\varphi(x,\tilde{P}(z)),\tilde{P}(z)) - \hat{f}_{\hat{h}_1,\hat{P}}(h_0(x,P(z)),P(z))| \le \\ &\tfrac{1}{N\tilde{h}_{N2}^2}\textstyle\sum_{i=1}^N\left|\tilde{K}_2\left(\tfrac{(\hat{h}_1(X_i,\hat{P}(Z_i)),\hat{P}(Z_i)) - (\varphi(x,\tilde{P}(z)),\tilde{P}(z))}{\tilde{h}_{N2}}\right) - \tilde{K}_2\left(\tfrac{(\hat{h}_1(X_i,\hat{P}(Z_i)),\hat{P}(Z_i)) - (h_0(x,P(z)),P(z))}{\tilde{h}_{N2}}\right)\right|\hat{I}_1(X_i,Z_i) \\ &\le \tfrac{M_3}{\tilde{h}_{N2}^3}\left[\left|\varphi(x,\tilde{P}(z)) - h_0(x,P(z))\right| + \left|\tilde{P}(z) - P(z)\right|\right]\end{aligned}$$

[34] Note that the first trimming function $\hat{I}_1$ would eventually eliminate observations which lie outside of $\overline{A}_1$ with probability approaching to 1. So in terms of the second trimming function, we only need to worry about (x,z) values in $\overline{A}_1$.

[35] Later, we may impose other conditions on this kernel function.

We know that on $\overline{A}_1$, both $\varphi(x, \tilde{P}(z))$, and $\tilde{P}(z)$ are uniformly consistent. Moreover, by assumption $|\varphi(x, \tilde{P}(z)) - h_0(x, P(z))|$ and $|\tilde{P}(z) - P(z)|$ are both $o_p(\tilde{h}_{N2}^3)$ on $\overline{A}_1$.

Next focus on the middle term of (14):

$$\left|\hat{f}_{\hat{h}_1,\hat{P}}(h_0(x, P(z)), P(z)) - \hat{f}_{h_1,P}(h_0(x, P(z)), P(z))\right| =$$

$$\frac{1}{N\tilde{h}_{N2}^2} \sum_{i=1}^N \; \tilde{K}_2 \; \frac{(\hat{h}_1(X_i,\hat{P}(Z_i)),\hat{P}(Z_i))-(h_0(x,P(z)),P(z))}{\tilde{h}_{N2}} \; -\tilde{K}_2 \; \frac{(h_1(X_i,P(Z_i)),P(Z_i))-(h_0(x,P(z)),P(z))}{\tilde{h}_{N2}} \quad \hat{I}_1(X_i,Z_i)$$

$$\leq \left|\frac{M_3}{N\tilde{h}_{N2}^3} \sum_{i=1}^N \left[\hat{h}_1(X_i, \hat{P}(Z_i)) - h_1(X_i, P(Z_i))\right]\hat{I}_1(X_i, Z_i)\right|$$

$$+\left|\frac{M_3}{N\tilde{h}_{N2}^3} \sum_{i=1}^N \left[\hat{P}(Z_i) - P(Z_i)\right]\hat{I}_1(X_i, Z_i)\right|$$

Using the results of Appendix C.1:

$$\left|\frac{M_3}{N\tilde{h}_{N2}^3} \sum_{i=1}^N \left[\hat{P}(Z_i) - P(Z_i)\right]\hat{I}_1(X_i, Z_i)\right|$$

$$\leq \left|\frac{M_3}{N^2\tilde{h}_{N2}^3} \sum_{i=1}^N \sum_{j=1}^N \psi_{NP}(D_j, X_j, Z_j; X_i, Z_i)\right| + \left|\frac{M_3}{N\tilde{h}_{N2}^3} \sum_{i=1}^N \hat{b}_P(X_i, Z_i)\right|$$

$$+\left|\frac{M_3}{N\tilde{h}_{N2}^3} \sum_{i=1}^N \hat{R}_P(X_i, Z_i)\right|$$

We will split the first term into two sums: one containing the terms where i and j are the same, and the other, where they are different. To deal with the sum containing terms with the same indices we will use a law of large numbers:

Theorem C.1 ***(Chebyshev)*** *Let $S_1, S_2, ...$ be uncorrelated with means $\mu_1, \mu_2, ...$ and variances $\sigma_1^2, \sigma_2^2, ...$ If $\sum_{i=1}^N \sigma_i^2 = o(N^2)$ as $N \to \infty$ then*

$$\frac{1}{N}\sum_{i=1}^N S_i - \frac{1}{N}\sum_{i=1}^N \mu_i \to^P 0$$

Now

$$\frac{M_3}{N^2\tilde{h}_{N2}^3} \sum_{i=1}^N \psi_{NP}(X_i, Z_i, D_i; X_i, Z_i) = \frac{1}{N} \sum_{i=1}^N \frac{M_3}{N\tilde{h}_{N2}^3 h_{NP}^{d_z}} e_1[M_{pN}^P(Z_i)]^{-1} e_1' K^P(0)\varepsilon_i^P I_{1i}$$

Using the law of iterated expectations we could show that each term of the above summation has 0 expectation. On the other hand,

$$\lim_{N\to\infty} \frac{1}{N^2} \sum_{i=1}^N var\left[\frac{M_3 e_1[M_{pN}^P(Z_i)]^{-1} e_1' K^P(0) I_{1i}\varepsilon_i^P}{N\tilde{h}_{N2}^3 h_{NP}^{d_z}}\right] = \lim_{N\to\infty} \frac{M_3^2 E\left(e_1[M_{pN}^P(Z_i)]^{-1}e_1'\right)^2\left(K^P(0)\right)^2 I_{1i}\left(\varepsilon_i^P\right)^2}{N^3\tilde{h}_{N2}^6 h_{NP}^{2d_z}}$$

Recall that $Nh_{NP}^{2d_z} \to \infty$ and M_{pN}^P to a nonsingular matrix. Thus, as long as $N\tilde{h}_{N2}^3$ does not converge to 0, or does not converge to 0 too fast, the variance condition needed to apply the theorem holds and we have[36]

$$plim_{N\to\infty}\left|\frac{1}{N}\sum_{i=1}^{N}\frac{M_3}{N\tilde{h}_{N2}^{3}h_{NP}^{d_z}}e_1[M_{pN}^{P}(Z_i)]^{-1}e_1'K^P(0)I_1(X_i,Z_i)\varepsilon_i^P\right|=0$$

Next, we focus on $\left|\frac{M_3}{N^2\tilde{h}_{N2}^3}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\psi_{NP}(X_j,Z_j,D_j;X_i,Z_i)\right|$. Our first step will be appealing to the Hoeffding, Powell, Stock and Stoker lemma to express the term sum inside the absolute value as a sum over one index only. Define

$$\begin{aligned}\zeta^P(D_i,Y_i,X_i,Z_i,D_j,Y_j,X_j,Z_j) &= \frac{1}{2h_{NP}^{d_z}}e_1[M_{pN}^{P}(Z_i)]^{-1}\left[\left(\frac{Z_j-Z_i}{h_{NP}}\right)\right]'K^P\left(\frac{Z_j-Z_i}{h_{NP}}\right)I_1(X_i,Z_i)\varepsilon_j^P\\ &+ \frac{1}{2h_{NP}^{d_z}}e_1[M_{pN}^{P}(Z_j)]^{-1}\left[\left(\frac{Z_i-Z_j}{h_{NP}}\right)\right]'K^P\left(\frac{Z_i-Z_j}{h_{NP}}\right)I_1(X_j,Z_j)\varepsilon_i^P\end{aligned}$$

By the law of iterated expectations $E\big[\zeta^P(D_i,Y_i,X_i,Z_i,D_j,Y_j,X_j,Z_j)\big]=0$. Moreover, using Cauchy-Schwarz inequality, and the facts that $Nh_{NP}^{2d_z}\to\infty$, M^PpN converges to a nonsingular matrix and K^P is 0 outside a compact set we could also show that $E\big[(\zeta^P(D_i,Y_i,X_i,Z_i,D_j,Y_j,X_j,Z_j))^2\big]=o(N)$. Therefore, using the Hoeffding, Powell, Stock and Stoker lemma, we can argue that

$$\begin{aligned}&plim_{N\to\infty}\tfrac{M_3}{N^2\tilde{h}_{N2}^3}\textstyle\sum_{i=1}^{N}\sum_{j\neq i}^{N}\psi_{NP}(X_j,Z_j,D_j;X_i,Z_i)\\ &= plim_{N\to\infty}\tfrac{M_3}{\sqrt{N\tilde{h}_{N2}^6}}plim_{N\to\infty}\textstyle\sum_{i=1}^{N}\sum_{j\neq i}^{N}\tfrac{\psi_{NP}(X_j,Z_j,D_j;X_i,Z_i)}{N^{3/2}}\\ &= plim_{N\to\infty}\tfrac{1}{N}\textstyle\sum_{i=1}^{N}\tfrac{M_3}{h_{NP}^{d_z}\tilde{h}_{N2}^3}E\left[e_1[M_{pN}^{P}(Z_j)]^{-1}I_1(X_j,Z_j)\left[\left(\tfrac{Z_i-Z_j}{h_{NP}}\right)\right]'K^P\left(\tfrac{Z_i-Z_j}{h_{NP}}\right)\varepsilon_i^P|D_i,X_i,Z_i\right]\end{aligned}$$

We can now apply Chebychev's law of large numbers to this expression. Expectation of each term is again 0. But we still have to verify that

$$\lim_{N\to\infty}\tfrac{1}{N^2}\textstyle\sum_{i=1}^{N}\tfrac{M_3^2}{\tilde{h}_{N2}^6}E\left\{\left(E\left[e_1[M_{pN}^{P}(Z_j)]^{-1}I_1(X_j,Z_j)\left[\tfrac{Z_i-Z_j}{h_{NP}}\right]'\tfrac{1}{h_{NP}^{d_z}}K^P\tfrac{Z_i-Z_j}{h_{NP}}\varepsilon_i^P|D_i,X_i,Z_i\right]\right)^2\right\}=0$$

By Jensen's inequality

$$\begin{aligned}&\left(E\left[e_1[M_{pN}^{P}(Z_j)]^{-1}I_1(X_j,Z_j)\left[\left(\tfrac{Z_i-Z_j}{h_{NP}}\right)\right]'\tfrac{1}{h_{NP}^{d_z}}K^P\left(\tfrac{Z_i-Z_j}{h_{NP}}\right)\varepsilon_i^P|D_i,X_i,Z_i\right]\right)^2\\ &\le E\left\{I_1(X_j,Z_j)\left(e_1[M_{pN}^{P}(Z_j)]^{-1}\left[\left(\tfrac{Z_i-Z_j}{h_{NP}}\right)\right]'\right)^2\tfrac{1}{h_{NP}^{2d_z}}\left(K^P\left(\tfrac{Z_i-Z_j}{h_{NP}}\right)\right)^2\left(\varepsilon_i^P\right)^2|D_i,X_i,Z_i\right\}\\ &= \tfrac{\left(\varepsilon_i^P\right)^2}{h_{NP}^{d_z}}E\left\{I_1(X_j,Z_j)\left(e_1[M_{pN}^{P}(Z_j)]^{-1}\left[\left(\tfrac{Z_i-Z_j}{h_{NP}}\right)\right]'\right)^2\tfrac{1}{h_{NP}^{d_z}}\left(K^P\left(\tfrac{Z_i-Z_j}{h_{NP}}\right)\right)^2|D_i,X_i,Z_i\right\}\end{aligned}$$

[36]Note that $T_N\xrightarrow{P}0\Leftrightarrow|T_N|\xrightarrow{P}0$

On A_1, $e_1[M^P_{pN}(Z_j)]^{-1}$ and the density $f_{X,Z}$ are bounded. Moreover, the kernel function K^P is assumed to have compact support, and hence for some positive $\tilde{C}$

$$E\left\{I_1(X_j,Z_j)\left(e_1[M^P_{pN}(Z_j)]^{-1}\left[\left(\frac{Z_i-Z_j}{h_{NP}}\right)\right]'\right)^2\frac{1}{h^{d_z}_{NP}}\left(K^P\left(\frac{Z_i-Z_j}{h_{NP}}\right)\right)^2|D_i,X_i,Z_i\right\}\leq\tilde{C}$$

Since, we also have $\sigma^2_P := E(\varepsilon^P_i)^2 < \infty$,

$$\frac{1}{N^2}\sum_{i=1}^N\frac{M_3^2}{\tilde{h}^6_{N2}}E\left\{\left(E\left[e_1[M^P_{pN}(Z_j)]^{-1}I_1(X_j,Z_j)\left[\left(\frac{Z_i-Z_j}{h_{NP}}\right)\right]'\frac{1}{h^{d_z}_{NP}}K^P\left(\frac{Z_i-Z_j}{h_{NP}}\right)\varepsilon^P_i|D_i,Y_i,X_i,Z_i\right]\right)^2\right\}$$
$$\leq\frac{1}{N^2}\sum_{i=1}^N\frac{M_3^2}{h^{d_z}_{NP}\tilde{h}^6_{N2}}\sigma^2_P\tilde{C}=\frac{M_3^2\sigma^2_P\tilde{C}}{\sqrt{Nh^{2d_z}_{NP}}\sqrt{N}\tilde{h}^6_{N2}}$$

We assumed that $Nh^{2d_z}_{NP}\to\infty$. Then if $\sqrt{N}\,\tilde{h}^6_{N2}$ does not go to 0, or if it does not go to 0 too fast, then the product of $\sqrt{N}h^{d_z}_{NP}$ and $\sqrt{N}\,\tilde{h}^6_{N2}$ will still go to ∞[37]. Next, we deal with

$$\left|\frac{M_3}{N\tilde{h}^3_{N2}}\sum_{i=1}^N\hat{b}_P(X_i,Z_i)\right|$$

From Appendix C.1, we know that $plim_{N\to\infty}\frac{1}{\sqrt{N}}\sum_{i=1}^N\hat{b}_P(X_i,Z_i)=b_P<\infty$. Then if $\lim_{N\to\infty}\sqrt{N}\tilde{h}^3_{N2}$ $=\infty$, this term too will be converging to 0 uniformly in probability. Finally, let us look at

$$\left|\frac{M_3}{N\tilde{h}^3_{N2}}\sum_{i=1}^N\hat{R}_P(X_i,Z_i)\right|=\left|\frac{M_3}{\sqrt{N}\tilde{h}^3_{N2}}\frac{1}{\sqrt{N}}\sum_{i=1}^N\hat{R}_P(X_i,Z_i)\right|$$

We know that $\frac{1}{\sqrt{N}}\sum_{i=1}^N\hat{R}_P(X_i,Z_i)=o_p(1)$. This and our previous assumption that $\lim_{N\to\infty}\sqrt{N}\tilde{h}^3_{N2}=$ ∞ jointly imply that this last term also goes to 0 uniformly in probability.

Next, we study the second part of the middle term in (14). Using Appendix C.2, we can write

$$\left|\frac{M_3}{N\tilde{h}^3_{N2}}\sum_{i=1}^N\left[\hat{h}_1(X_i,\hat{P}(Z_i))-h_1(X_i,P(Z_i))\right]\hat{I}_1(X_i,Z_i)\right|\leq$$
$$\left|\frac{M_3}{N^2\tilde{h}^3_{N2}}\sum_{i=1}^N\sum_{j=1}^N\psi_{N\hat{h}_1}(D_j,Y_j,X_j,Z_j;X_i,Z_i)\right|+\left|\frac{M_3}{N\tilde{h}^3_{N2}}\sum_{i=1}^N\hat{b}_{\hat{h}_1}(X_i,Z_i)\right|$$
$$+\left|\frac{M_3}{N\tilde{h}^3_{N2}}\sum_{i=1}^N\hat{R}_{\hat{h}_1}(X_i,Z_i)\right|$$

Again, we know that $plim_{N\to\infty}\frac{1}{\sqrt{N}}\sum_{i=1}^N\hat{b}_{\hat{h}_1}(X_i,Z_i)=b_{h_1}+b_{h_1P}<\infty$. So if $\sqrt{N}\tilde{h}^3_{N2}\to\infty$, the middle term goes to 0 in probability by the continuous mapping theorem. Similarly, we know $plim_{N\to\infty}\frac{1}{\sqrt{N}}\sum_{i=1}^N\hat{R}_{\hat{h}_1}(X_i,Z_i)=0$. Thus, the same condition guarantees that the last sum converges to 0 in probability. As for the first sum, again we can split it into two pieces. One

[37]We could for example, choose $\tilde{h}_{N2}=h^{d_z/6}_{NP}$.

piece contains the terms where the two indices equal, the other piece contains the terms where the indices are different:

$$\left|\frac{M_3}{N^2\tilde{h}_{N2}^3}\sum_{i=1}^N \psi_{N\hat{h}_1}(D_i, Y_i, X_i, Z_i; X_i, Z_i)\right| =$$
$$\left|\frac{1}{N}\sum_{i=1}^N \frac{M_3}{N\tilde{h}_{N2}^3}\left(\psi_{Nh_1}(D_i, Y_i, X_i, Z_i; X_i, Z_i) + \frac{\partial h_1}{\partial P}(X_i, P(Z_i))\psi_{NP}(D_i, X_i, Z_i; X_i, Z_i)\right)\right|$$

Each term has 0 expectation. To apply Chebychev's law of large numbers, we need to verify that the sum of the variances is $o(N^2)$. By Cauchy-Schwarz inequality it suffices to verify that

$$\lim_{N\to\infty}\frac{1}{N^2}\sum_{i=1}^N E\big[\frac{M_3^2}{N^2\tilde{h}_{N2}^6}\psi_{Nh_1}(D_i, Y_i, X_i, Z_i; X_i, Z_i)^2\big] =$$
$$\lim_{N\to\infty}\frac{M_3^2}{N^3\tilde{h}_{N2}^6 h_{Nh_1}^{2(d_x+1)}}E\left[\left(e_1[M_{pN}^{h_1}(X_i, P(Z_i))]^{-1}e_1'\right)^2\left(K^{h_1}(0)\right)^2 I_1(X_i, Z_i)\left(\varepsilon_i^{h_1}\right)^2\right] = 0$$

and

$$\lim_{N\to\infty}\frac{1}{N^2}\sum_{i=1}^N E\big[\frac{M_3^2}{N^2\tilde{h}_{N2}^6}\left(\frac{\partial h_1}{\partial P}(X_i, P(Z_i))\right)^2\psi_{NP}(D_i, X_i, Z_i; X_i, Z_i)^2\big] =$$
$$\lim_{N\to\infty}\frac{M_3^2}{N^3\tilde{h}_{N2}^6 h_{NP}^{2d_z}}E\left[\left(\frac{\partial h_1}{\partial P}(X_i, P(Z_i))\right)^2\left(e_1[M_{pN}^P(Z_i)]^{-1}e_1'\right)^2\left(K^P(0)\right)^2 I_1(X_i, Z_i)\left(\varepsilon_i^P\right)^2\right] = 0$$

The first one is true because the term inside the parentheses is bounded, $N\tilde{h}_{N2}^6 \to \infty$ and $Nh_{Nh_1}^{2(d_x+1)} \to \infty$. The second one is true because $Nh_{NP}^{2d_z} \to \infty$, the term inside the parentheses is bounded, and $N\tilde{h}_{N2}^6 \to \infty$. So the sum of terms with $i = j$ converges to 0 in probability. For the other sum, we again use Hoeffding, Powell, Stock and Stoker lemma. By arguments similar to those in Appendix B.2, we can show that

$$plim_{N\to\infty}\frac{M_3}{N^2\tilde{h}_{N2}^3}\sum_{i=1}^N\sum_{j\neq i}^N \psi_{N\hat{h}_1}(D_j, Y_j, X_j, Z_j; X_i, Z_i) =$$
$$= plim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N\frac{M_3}{\tilde{h}_{N2}^3}E\left[\psi_{N\hat{h}_1}(D_i, Y_i, X_i, Z_i; D_j, Y_j, X_j, Z_j)|D_i, Y_i, X_i, Z_i\right]$$

Then we apply the Chebyshev's theorem one last time. Again, the expectation of i^{th} term is 0. And given that we have already assumed $Nh_{NP}^{2d_z} \to \infty$, $Nh_{Nh_1}^{2(d_x+1)} \to \infty$ and $N\tilde{h}_{N2}^{12}$ does not go to 0, the variance condition is satisfied. Therefore, this sum converges to 0 in probability as well.

This leaves us with the last piece of (14):

$$\left|\hat{f}_{h_1,P}(h_0(x, P(z)), P(z)) - f_{h_1,P}(h_0(x, P(z)), P(z))\right| =$$
$$\left|\frac{1}{N\tilde{h}_{N2}^2}\sum_{i=1}^N \tilde{K}_2\left(\frac{\left(h_1(X_i, P(Z_i)), P(Z_i)\right) - (h_0(x, P(z)), P(z))}{\tilde{h}_{N2}}\right)\hat{I}_1(X_i, Z_i) - f_{h_1,P}(h_0(x, P(z)), P(z))\right|$$

$$\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\tilde{h}_{N2}^2}\tilde{K}_2\left(\frac{\Big(h_1(X_i,P(Z_i)),P(Z_i)\Big)-(h_0(x,P(z)),P(z))}{\tilde{h}_{N2}}\right)\hat{I}_1(X_i,Z_i)=$$

$$\frac{1}{N\tilde{h}_{N2}^2}\sum_{i=1}^{N}\tilde{K}_2\left(\frac{\Big(h_1(X_i,P(Z_i)),P(Z_i)\Big)-(h_0(x,P(z)),P(z))}{\tilde{h}_{N2}}\right)I_1(X_i,Z_i)+\frac{1}{N\tilde{h}_{N2}^2}\sum_{i=1}^{N}\tilde{K}_2\left(\frac{\Big(h_1(X_i,P(Z_i)),P(Z_i)\Big)-(h_0(x,P(z)),P(z))}{\tilde{h}_{N2}}\right)$$

$$\times[\hat{\sigma}(X_i,Z_i)]^{-1}\tilde{J}_2^{-}\quad\frac{f_{X,Z}(X_i,Z_i)-q_{01}}{\hat{\sigma}(X_i,Z_i)}\quad\Big[\hat{f}(X_i,Z_i)-f_{X,Z}(X_i,Z_i)\Big]1\{\hat{f}(X_i,Z_i)>f(X_i,Z_i)\}$$

$$+\frac{1}{N\tilde{h}_{N2}^2}\sum_{i=1}^{N}\tilde{K}_2\left(\frac{\Big(h_1(X_i,P(Z_i)),P(Z_i)\Big)-(h_0(x,P(z)),P(z))}{\tilde{h}_{N2}}\right)[\hat{\sigma}(X_i,Z_i)]^{-1}\tilde{J}_2^{+}\quad\frac{f_{X,Z}(X_i,Z_i)-q_{01}}{\hat{\sigma}(X_i,Z_i)}\quad\Big[\hat{f}(X_i,Z_i)-f_{X,Z}(X_i,Z_i)\Big]$$

$$\times 1\{\hat{f}(X_i,Z_i)\le f(X_i,Z_i)\}$$

My goal is to show that each of the last two terms is uniformly $o_p(1)$. Let's focus on the first of those two. That term equals

$$\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{N^2\tilde{h}_{N2}^2\tilde{h}_{N1}^d}\tilde{K}_2\left(\frac{\Big(h_1(X_i,P(Z_i)),P(Z_i)\Big)-(h_0(x,P(z)),P(z))}{\tilde{h}_{N2}}\right)[\hat{\sigma}(X_i,Z_i)]^{-1}\tilde{J}_2^{-}\quad\frac{f_{X,Z}(X_i,Z_i)-q_{01}}{\hat{\sigma}(X_i,Z_i)}\quad 1\{\hat{f}(X_i,Z_i)>f(X_i,Z_i)\}$$

$$\times\ \tilde{K}_1\quad\frac{(X_j,Z_j)-(X_i,Z_i)}{\tilde{h}_{N1}}\quad -E\quad\tilde{K}_1\quad\frac{(X_j,Z_j)-(X_i,Z_i)}{\tilde{h}_{N1}}\quad|X_i,Z_i \tag{15}$$

$$+\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{N^2\tilde{h}_{N2}^2\tilde{h}_{N1}^d}\tilde{K}_2\left(\frac{\Big(h_1(X_i,P(Z_i)),P(Z_i)\Big)-(h_0(x,P(z)),P(z))}{\tilde{h}_{N2}}\right)[\hat{\sigma}(X_i,Z_i)]^{-1}\tilde{J}_2^{-}\quad\frac{f_{X,Z}(X_i,Z_i)-q_{01}}{\hat{\sigma}(X_i,Z_i)}\quad 1\{\hat{f}(X_i,Z_i)>f(X_i,Z_i)\}$$

$$\times\ E\quad\tilde{K}_1\quad\frac{(X_j,Z_j)-(X_i,Z_i)}{\tilde{h}_{N1}}\quad|X_i,Z_i\quad -f_{X,Z}(X_i,Z_i) \tag{16}$$

Using the equicontinuity lemma we will show that (15) is $o_p(1)$. For this purpose, for $g\in\mathcal{H}_1$, define $\tilde{\sigma}(X_i,Z_i)=|g(X_i,Z_i)-f_{X,Z}(X_i,Z_i)|$, $\tilde{L}_i=1\{g(X_i,Z_i)>f_{X,Z}(X_i,Z_i)\}$. Then

$$\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{N^2\tilde{h}_{N2}^2\tilde{h}_{N1}^d}\tilde{K}_2\left(\frac{\Big(h_1(X_i,P(Z_i)),P(Z_i)\Big)-(h_0(x,P(z)),P(z))}{\tilde{h}_{N2}}\right)[\tilde{\sigma}(X_i,Z_i)]^{-1}\tilde{J}_2^{-}\quad\frac{f_{X,Z}(X_i,Z_i)-q_{01}}{\tilde{\sigma}(X_i,Z_i)}\quad\tilde{L}_i$$

$$\times\ \tilde{K}_1\quad\frac{(X_j,Z_j)-(X_i,Z_i)}{\tilde{h}_{N1}}\quad -E\quad\tilde{K}_1\quad\frac{(X_j,Z_j)-(X_i,Z_i)}{\tilde{h}_{N1}}\quad|X_i,Z_i\quad -f_{X,Z}(X_i,Z_i)$$

is a degenerate U-process of order one which satisfies the conditions of the equicontinuity lemma. Finally, (16) is $o_p(\tilde{h}_{N1})$ by the smoothness of $f_{X,Z}$. On the other hand, by using the same tricks, we can also show that the symmetric term (i.e. the term involving $\tilde{J}_+$) is also uniformly $o_p(1)$. As a result, $\hat{f}_{h_1,P}(h_0(x,P(z)),P(z))$ converges in probability uniformly to

$$\tilde{f}_{h_1,P}(h_0(x,P(z)),P(z))=\frac{1}{N\tilde{h}_{N2}^2}\sum_{i=1}^{N}\tilde{K}_2\left(\frac{\big(h_1(X_i,P(Z_i)),P(Z_i)\big)-(h_0(x,P(z)),P(z))}{\tilde{h}_{N2}}\right)I_1(X_i,Z_i)$$

Given our assumptions on $\tilde{K}_2$ we can use a strong law of large numbers to show that this converges to

$$E\left[\frac{1}{\tilde{h}_{N2}^2}\tilde{K}_2\left(\frac{\big(h_1(X_i,P(Z_i)),P(Z_i)\big)-(h_0(x,P(z)),P(z))}{\tilde{h}_{N2}}\right)1(A_1)\right]$$

Now the set A_1 is closed, but we can find a sequence of open sets that are all contained in A_1. Moreover the limit of this sequence of open sets will be A_1. Using change of variables theorem by breaking the set A_1 into disjoint regions where P and h_1 have non-zero derivatives, if necessary, and then using Silverman's theorem we have the desired result.