Addendum to:

"Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score,"

K. Hirano, G. Imbens & G. Ridder

November, 2002

Expanded Proof of Theorem 1:

The estimated weight estimator $\hat{\beta}_{ew}$ is

$$\hat{\beta}_{ew} = \frac{1}{N}\sum_{i=1}^{N}\frac{T_i.Y_i}{\hat{p}_K(X_i)} \tag{1}$$

with $\hat{p}_K(X_i) = l(R^K(X_i)'\hat{\pi}_K)$. In the proof we show that

$$\left|\sqrt{N}(\hat{\beta}_{ew} - \beta_0) - \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{\left(\frac{T_i.Y_i}{p_0(X_i)} - \beta_0\right) - \frac{\mathrm{E}(Y \mid X_i)}{p_0(X_i)}(T_i - p_0(X_i))\right\}\right| = o_P(1) \tag{2}$$

Hence, $\hat{\beta}_{ew}$ is asymptotically linear, i.e. behaves asymptotically as a sample average, with score function

$$\psi(Y,T,X) = \left(\frac{T.Y}{p_0(X)} - \beta_0\right) - \frac{\mathrm{E}(Y \mid X)}{p_0(X)}(T - p_0(X)) \tag{3}$$

The first term of the score function is equal to the score that would obtain if we substitute the population p_0 for the estimated $\hat{p}_K$. The second term gives the contribution of the nonparametric estimator $\hat{p}_K$ to the asymptotic distribution of $\hat{\beta}_{ew}$. The contribution is linear in $T - p_0(X)$. Hence, the score linearizes the estimator with respect to β (the estimator is already linear in β) and p. The asymptotic variance of $\hat{\beta}_{ew}$ is equal to the variance of $\psi(Y,T,x)$ (note that its mean is 0).

In the proof of (2) we rewrite the difference by adding and subtracting a number of terms, so that we can bound the differences. We give the asymptotic order of all differences, which makes it easier to understand the role of the assumptions. We have

$$\sqrt{N}(\hat{\beta}_{ew} - \beta_0) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{T_iY_i}{\hat{p}_K(X_i)} - \frac{T_iY_i}{p_0(X_i)} + \frac{T_iY_i}{p_0(X_i)^2}(\hat{p}_K(X_i) - p_0(X_i))\right) + \tag{4}$$

$$+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(-\frac{T_iY_i}{p_0(X_i)^2}(\hat{p}_K(X_i)-p_0(X_i))+\int_{\mathcal{X}}\frac{\mathrm{E}(Y\mid x)}{p_0(x)}(\hat{p}_K(x)-p_0(x))\mathrm{d}F_0(x)\right)+ \quad (5)$$

$$-\sqrt{N}\int_{\mathcal{X}}\frac{\mathrm{E}(Y\mid x)}{p_0(x)}(\hat{p}_K(x)-p_0(x))\mathrm{d}F_0(x)-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\tilde{\delta}_K(X_i)\frac{T_i-p_K(X_i)}{\sqrt{p_K(X_i)(1-p_K(X_i))}}+ \quad (6)$$

$$+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\tilde{\delta}_K(X_i)-\delta_K(X_i))\frac{T_i-p_K(X_i)}{\sqrt{p_K(X_i)(1-p_K(X_i))}}+ \quad (7)$$

$$+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\delta_K(X_i)\frac{T_i-p_K(X_i)}{\sqrt{p_K(X_i)(1-p_K(X_i))}}-\delta_0(X_i)\frac{T_i-p_0(X_i)}{\sqrt{p_0(X_i)(1-p_0(X_i))}}\right)+ \quad (8)$$

$$+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{\left(\frac{T_i.Y_i}{p_0(X_i)}-\beta_0\right)+\delta_0(X_i)\frac{T_i-p_0(X_i)}{\sqrt{p_0(X_i)(1-p_0(X_i))}}\right\} \quad (9)$$

In this expression F_0 is the population cdf of X and

$$\tilde{\delta}_K(x)=-\int_{\mathcal{X}}\frac{\mathrm{E}(Y\mid z)}{p_0(z)}L'(R^K(z)'\tilde{\pi}_K)R^K(z)'\mathrm{d}F_0(z)\tilde{\Sigma}_K^{-1}\sqrt{L'(R^K(x)'\pi_K)}R^K(x) \quad (10)$$

$$\delta_K(x)=-\int_{\mathcal{X}}\frac{\mathrm{E}(Y\mid z)}{p_0(z)}L'(R^K(z)'\pi_K)R^K(z)'\mathrm{d}F_0(z)\Sigma_K^{-1}\sqrt{L'(R^K(x)'\pi_K)}R^K(x) \quad (11)$$

$$\delta_0(x)=-\frac{\mathrm{E}(Y\mid x)}{p_0(x)}\sqrt{p_0(X_i)(1-p_0(X_i))} \quad (12)$$

Note that (9) is equal to the linearized expression for $\sqrt{N}(\hat{\beta}_{ew}-\beta_0)$. To show that the estimator is indeed asymptotically linear, we must derive bounds on the terms (4)-(8). If a bound depends on both K and N, we derive the bound for sequences $K(N)$ that go to ∞ with N. Because during the derivation some restrictions on these sequences are imposed, the resulting bounds are not uniform in K. We have seen this type of argument in the derivation of the order of $\|\hat{\pi}_{K(N)}-\pi_{K(N)}\|$ where we imposed the large sample identification condition $\zeta(K(N))^4/N\to 0$.

0.1 Bound on (4)

Using the fact that

$$\frac{TY}{\hat{p}_K(X)} - \frac{TY}{p_0(X)} + \frac{TY}{p_0(X)^2}(\hat{p}_K(X) - p_0(X)) = \frac{TY}{p_0(X)^2\hat{p}_K(X)}(\hat{p}_K(X) - p_0(X))^2 \tag{13}$$

we rewrite (4) as

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{T_iY_i}{p_0(X_i)^2\hat{p}_{K(N)}(X_i)}(\hat{p}_{K(N)}(X_i) - p_{K(N)}(X_i))^2 + \tag{14}$$

$$+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{T_iY_i}{p_0(X_i)^2\hat{p}_{K(N)}(X_i)}(p_{K(N)}(X_i) - p_0(X_i))^2 + \tag{15}$$

$$+\frac{2}{\sqrt{N}}\sum_{i=1}^{N}\frac{T_iY_i}{p_0(X_i)^2\hat{p}_{K(N)}(X_i)}(\hat{p}_{K(N)}(X_i) - p_{K(N)}(X_i))(p_{K(N)}(X_i) - p_0(X_i)) \tag{16}$$

By the mean value theorem

$$\hat{p}_K(x) - p_K(x) = L'(R^K(x)'\tilde{\pi}_K)R^K(x)'(\hat{\pi}_K - \pi_K) \tag{17}$$

and hence we find for (14)

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{T_iY_i}{p_0(X_i)^2\hat{p}_{K(N)}(X_i)}L'(R^{K(N)}(X_i)'\tilde{\pi}_{K(N)})^2(R^{K(N)}(X_i)'(\hat{\pi}_{K(N)} - \pi_{K(N)}))^2 \tag{18}$$

Because $\hat{p}_{K(N)}$ is bounded from 0 on $\mathcal{X}$ with probability 1 for sequences $K(N)$ that satisfy the large sample identification condition, we have

$$\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{T_iY_i}{p_0(X_i)^2\hat{p}_{K(N)}(X_i)}L'(R^{K(N)}(X_i)'\tilde{\pi}_{K(N)})^2(R^{K(N)}(X_i)'(\hat{\pi}_{K(N)} - \pi_{K(N)}))^2\right| \leq \tag{19}$$

$$\leq \frac{C}{\sqrt{N}}\sum_{i=1}^{N}|Y_i|(R^K(X_i)'(\hat{\pi}_K - \pi_K))^2 + o_P(1) \leq$$

$$\leq C\sqrt{N}\zeta(K(N))^2\|\hat{\pi}_{K(N)} - \pi_{K(N)}\|^2 \frac{1}{N}\sum_{i=1}^{N}|Y_i| + o_P(1) \leq$$

$$\leq C\sqrt{N}\zeta(K(N))^2\|\hat{\pi}_{K(N)} - \pi_{K(N)}\|^2 + o_P(1)$$

because $\mathrm{E}(Y^2) < \infty$. The $o_P(1)$ terms converge to 0 in probability for all sequences $K(N)$ that satisfy the large sample identification condition. Next, using (39), (15) is bounded by

$$\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{T_iY_i}{p_0(X_i)^2\hat{p}_{K(N)}(X_i)}(p_{K(N)}(X_i) - p_0(X_i))^2\right| \leq C\sqrt{N}\zeta(K)^2K^{-\frac{s}{r}} + o_P(1) \tag{20}$$

Finally, (16) is bounded by

$$\frac{C}{\sqrt{N}}\sum_{i=1}^{N}|\hat{p}_{K(N)}(X_i) - p_{K(N)}(X_i)||p_{K(N)}(X_i) - p_0(X_i)| + o_P(1) \leq \tag{21}$$

$$\leq C\frac{\zeta(K(N))K^{-\frac{s}{2r}}}{\sqrt{N}}\sum_{i=1}^{N}|R^{K(N)}(X_i)'(\hat{\pi}_{K(N)} - \pi_{K(N)})| + o_P(1) \leq$$

$$\leq C\zeta(K(N))^2K^{-\frac{s}{2r}}\sqrt{N}\|\hat{\pi}_{K(N)} - \pi_{K(N)})\| + o_P(1)$$

Hence, combining (19), (20) and (21), we find by the Markov inequality

$$\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{T_iY_i}{\hat{p}_K(X_i)} - \frac{T_iY_i}{p_0(X_i)} + \frac{T_iY_i}{p_0(X_i)^2}(\hat{p}_K(X_i) - p_0(X_i))\right)\right| = \tag{22}$$

$$= O_P\left(\frac{\zeta(K(N))^3}{\sqrt{N}}\right) + O_P\left(\sqrt{N}\zeta(K(N))^2K(N)^{-\frac{s}{r}}\right) + O_P\left(\zeta(K(N))^{5/2}K(N)^{-\frac{s}{2r}}\right)$$

0.2 Bound on (5)

First we rewrite (5) as

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(-\frac{T_iY_i}{p_0(X_i)^2}(\hat{p}_K(X_i) - p_K(X_i)) + \int_{\mathcal{X}}\frac{\mathrm{E}(Y\mid x)}{p_0(x)}(\hat{p}_K(x) - p_K(x))\mathrm{d}F_0(x)\right) + \tag{23}$$

$$+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(-\frac{T_iY_i}{p_0(X_i)^2}(p_K(X_i)-p_0(X_i))+\int_{\mathcal{X}}\frac{\mathrm{E}(Y\mid x)}{p_0(x)}(p_K(x)-p_0(x))\mathrm{d}F_0(x)\right) \tag{24}$$

Denote (24) by V_K with $\mathrm{E}(V_K)=0$. The variance is, using

$$\mathrm{Var}(TY\mid X)=\mathrm{E}(Y^2\mid X,T=1)p_0(X)-\mathrm{E}(Y\mid X)^2p_0(X)^2 \tag{25}$$

and $\mathrm{Var}(V_K)=\mathrm{E}(\mathrm{Var}(V_K\mid X_1,\ldots,X_N))+\mathrm{Var}(\mathrm{E}(V_K\mid X_1,\ldots,X_N))$, equal to

$$\mathrm{Var}(V_K)=\mathrm{E}_X\left[\frac{\mathrm{E}(Y^2\mid X,T=1)-\mathrm{E}(Y\mid X)^2p_0(X)}{p_0(X)^3}(p_K(X)-p_0(X))^2\right]+ \tag{26}$$

$$+\mathrm{E}_X\left[\frac{\mathrm{E}(Y\mid X)^2}{p_0(X)^2}(p_K(X)-p_0(X))^2\right]-$$

$$-\left(\mathrm{E}_X\left[\frac{\mathrm{E}(Y\mid X)}{p_0(X)}(p_K(X)-p_0(X))\right]\right)^2$$

which is bounded by

$$\mathrm{E}_X\left[\frac{\mathrm{E}(Y^2\mid X,T=1)}{p_0(X)^3}(p_K(X)-p_0(X))^2\right] \tag{27}$$

By (39), the assumption that p_0 is greater than $\varepsilon>0$ on $\mathcal{X}$, and the assumption $\mathrm{E}(Y^2)<\infty$

$$\mathrm{E}_X\left[\frac{\mathrm{E}(Y^2\mid X,T=1)}{p_0(X)^3}(p_K(X)-p_0(X))^2\right]<C\zeta(K)^2K^{-\frac{s}{r}} \tag{28}$$

and we conclude that

$$\mathrm{E}(|V_K|)\leq\sqrt{\mathrm{Var}(V_K)}<C\zeta(K)K^{-\frac{s}{2r}} \tag{29}$$

and hence

$$|V_K|=O_P(\zeta(K)K^{-\frac{s}{2r}}) \tag{30}$$

This bounds the terms in (24). Note that the bound is valid for each sequence $K(N)$ for which $\mathrm{E}(|V_{K(N)}|)$ is bounded.

Now consider (23). By the mean value theorem this is equal to

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[-\frac{T_iY_i}{p_0(X_i)^2}L'(R^{K(N)}(X_i)'\tilde{\pi}_{K(N)})R^{K(N)}(X_i)'+ \tag{31}\right.$$

$$\left.+\int_{\mathcal{X}}\frac{\mathrm{E}(Y\mid x)}{p_0(x)}L'(R^{K(N)}(x)'\tilde{\pi}_K(N))R^{K(N)}(x)'\mathrm{d}F_0(x)\right].(\hat{\pi}_{K(N)}-\pi_{K(N)})$$

We consider the two factors in (31) separately. By a second application of the mean value theorem we write the first factor of (31) as

$$W_{1K(N)}-W_{2K(N)}+W_{3K(N)} \tag{32}$$

with

$$W_{1K}=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[-\frac{T_iY_i}{p_0(X_i)^2}L'(R^K(X_i)'\pi_K)R^K(X_i)+\int_{\mathcal{X}}\frac{\mathrm{E}(Y\mid x)}{p_0(x)}L'(R^K(x)'\pi_K)R^K(x)\mathrm{d}F_0(x)\right] \tag{33}$$

$$W_{2K}=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{T_iY_i}{p_0(X_i)^2}L''(R^K(X_i)'\overline{\pi}_K)R^K(X_i)R^K(X_i)'(\tilde{\pi}_K-\pi_K) \tag{34}$$

$$W_{3K}=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\int_{\mathcal{X}}\frac{\mathrm{E}(Y\mid x)}{p_0(x)}L''(R^K(x)'\overline{\pi}_K)R^{K(N)}(x)R^{K(N)}(x)'\mathrm{d}F_0(x)(\tilde{\pi}_K-\pi_K) \tag{35}$$

We first compute the variance of W_{1K}, which is, using $\mathrm{Var}(W_{1K}) = \mathrm{E}(\mathrm{Var}(W_{1K}\mid X_1,\ldots,X_N))+\mathrm{Var}(\mathrm{E}(W_{1K}\mid X_1,\ldots,X_N))$ equal to

$$\mathrm{Var}(W_{1K})=\mathrm{E}_X\left[\frac{\mathrm{E}(Y^2\mid X,T=1)-\mathrm{E}(Y\mid X)^2p_0(X)}{p_0(X)^3}L'(R^K(X)'\pi_K)^2R^K(X)R^K(X)'\right]+ \tag{36}$$

$$+\mathrm{E}_X\left[\frac{\mathrm{E}(Y\mid X)^2}{p_0(X)^2}L'(R^K(X)'\pi_K)^2R^K(X)R^K(X)'\right]-$$

$$-\mathrm{E}_X\left[\frac{\mathrm{E}(Y \mid X)}{p_0(X)}L'(R^K(X)'\pi_K)R^K(X)\right]\mathrm{E}_X\left[\frac{\mathrm{E}(Y \mid X)}{p_0(X)}L'(R^K(X)'\pi_K)R^K(X)\right]'$$

This is bounded by

$$\mathrm{E}_X\left[\frac{\mathrm{E}(Y^2 \mid X, T=1)}{p_0(X)^3}L'(R^K(X)'\pi_K)^2 R^K(X)R^K(X)'\right] \tag{37}$$

Because $\mathrm{E}(Y^2) < \infty$, p_0 is bounded from 0, and $L' \leq 1/4$, (37) is bounded by

$$C\mathrm{E}_X(R^K(X)R^K(X)') \tag{38}$$

Hence

$$\mathrm{E}(\|W_{1K}\|) \leq \sqrt{\mathrm{tr}(\mathrm{Var}(W_K))} \leq C\sqrt{\mathrm{tr}(\mathrm{E}_X(R^K(X)R^K(X)')} = \tag{39}$$

$$= C\sqrt{\mathrm{E}(\|R^K(X)\|^2)} \leq C\zeta(K)$$

Next we consider $W_{2K(N)}$. Note that the terms in the sum are not independent. For that reason we derive a bound on each term and use the triangle inequality to obtain an overall bound

$$\left\|\frac{T_iY_i}{p_0(X_i)^2}L''(R^K(X_i)'\overline{\pi}_K)R^K(X_i)R^K(X_i)'(\tilde{\pi}_K-\pi_K)\right\| \leq \tag{40}$$

$$\left|\frac{T_iY_i}{p_0(X_i)^2}L''(R^K(X_i)'\overline{\pi}_K)\right|\left\|R^K(X_i)R^K(X_i)'\right\|\|\tilde{\pi}_K-\pi_K\|$$

Because $L'' = L(1-L)(1-2L)$ is bounded and $\mathrm{E}(Y^2) < \infty$, we find, using the Cauchy-Schwartz inequality and (51)[1] , that

$$\mathrm{E}\left[\left|\frac{T_iY_i}{p_0(X_i)^2}L''(R^K(X_i)'\overline{\pi}_K)\right|\left\|R^K(X_i)R^K(X_i)'\right\|\|\tilde{\pi}_K-\pi_K\|\right] \leq C\frac{\zeta(K)^{3/2}}{N^{1/2}} \tag{41}$$

[1] As before (51) only holds for sequences $K(N)$ that satisfy the asymptotic identification condition.

Hence,

$$\mathrm{E}(\|W_{2K}\|) \leq C\zeta(K)^{3/2} \tag{42}$$

Using an analogous argument we show that

$$\mathrm{E}(\|W_{3K}\|) \leq C\zeta(K)^{3/2} \tag{43}$$

By the triangle and Cauchy-Schwartz inequalities, we find, using (39), (42),(43), and (51)

$$\mathrm{E}(|(W_{1K(N)} - W_{2K(N)} + W_{3K(N)})'(\hat{\pi}_{K(N)} - \pi_{K(N)})|) \leq C\frac{\zeta(K(N))^2}{\sqrt{N}} \tag{44}$$

Combining (44) with (30) we find that[2]

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(-\frac{T_iY_i}{p_0(X_i)^2}(\hat{p}_K(X_i) - p_0(X_i)) + \int_{\mathcal{X}}\frac{\mathrm{E}(Y \mid x)}{p_0(x)}(\hat{p}_K(x) - p_0(x))\mathrm{d}F_0(x)\right) = \tag{45}$$

$$= O_P(\zeta(K(N))K(N)^{-\frac{s}{2r}}) + O_P\left(\frac{\zeta(K(N))^2}{\sqrt{N}}\right)$$

0.3 Bound on (6)

Consider

$$-\sqrt{N}\int_{\mathcal{X}}\frac{\mathrm{E}(Y \mid x)}{p_0(x)}(\hat{p}_K(x) - p_0(x))\mathrm{d}F_0(x) = \tag{46}$$

$$-\sqrt{N}\int_{\mathcal{X}}\frac{\mathrm{E}(Y \mid x)}{p_0(x)}(\hat{p}_K(x) - p_K(x))\mathrm{d}F_0(x) - \sqrt{N}\int_{\mathcal{X}}\frac{\mathrm{E}(Y \mid x)}{p_0(x)}(p_K(x) - p_0(x))\mathrm{d}F_0(x) \tag{47}$$

We concentrate on the first term in (47). Upon substitution of (17) in the first term in (47) we find

$$-\sqrt{N}\int_{\mathcal{X}}\frac{\mathrm{E}(Y \mid x)}{p_0(x)}(\hat{p}_K(x) - p_K(x))\mathrm{d}F_0(x) = \tag{48}$$

[2]Note that the second term is $o_P(1)$ for $K(N)$ sequences that satisfy the asymptotic identification condition. As noted, we only impose this restriction when we combine the bounds.

$$-\int_{\mathcal{X}} \frac{\mathrm{E}(Y \mid x)}{p_0(x)} L'(R^K(x)'\tilde{\pi}_K) R^K(x)' \mathrm{d}F_0(x) \sqrt{N}(\hat{\pi}_K - \pi_K) \tag{49}$$

We denote

$$\tilde{\Psi}_K = -\int_{\mathcal{X}} \frac{\mathrm{E}(Y \mid x)}{p_0(x)} L'(R^K(x)'\tilde{\pi}_K) R^K(x) \mathrm{d}F_0(x) \tag{50}$$

and using this notation, the definition in (10) and (48), (49) is equal to

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\tilde{\Psi}_K' \tilde{\Sigma}_K^{-1} \sqrt{p_K(X_i)(1-p_K(X_i))} \right) \frac{T_i - p_K(X_i)}{\sqrt{p_K(X_i)(1-p_K(X_i))}} = \tag{51}$$

$$= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \tilde{\delta}_K(X_i) \frac{T_i - p_K(X_i)}{\sqrt{p_K(X_i)(1-p_K(X_i))}} \tag{52}$$

Hence

$$\left| -\sqrt{N} \int_{\mathcal{X}} \frac{\mathrm{E}(Y \mid x)}{p_0(x)} (\hat{p}_K(x) - p_0(x)) \mathrm{d}F_0(x) - \right. \tag{53}$$

$$\left. - \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \tilde{\delta}_K(X_i) \frac{T_i - p_K(X_i)}{\sqrt{p_K(X_i)(1-p_K(X_i))}} \right| =$$

$$= \left| \sqrt{N} \int_{\mathcal{X}} \frac{\mathrm{E}(Y \mid x)}{p_0(x)} (p_K(x) - p_0(x)) \mathrm{d}F_0(x) \right| \tag{54}$$

Because we assume that $\mathrm{E}(Y^2)$ and hence $\mathrm{E}(Y)$ are finite and that p_0 is bounded from 0 on $\mathcal{X}$, we have by (39)

$$\left| \sqrt{N} \int_{\mathcal{X}} \frac{\mathrm{E}(Y \mid x)}{p_0(x)} (p_{K(N)}(x) - p_0(x)) \mathrm{d}F_0(x) \right| \tag{55}$$

$$< C\sqrt{N}\zeta(K(N))K(N)^{-\frac{s}{2r}} = O(\sqrt{N}\zeta(K(N))K(N)^{-\frac{s}{2r}})$$

This gives the bound on (6) which holds for all sequences $K(N)$.

0.4 Bound on (7)

Define

$$\Psi_K = -\int_{\mathcal{X}} \frac{\mathrm{E}(Y \mid x)}{p_0(x)} L'(R^K(x)'\pi_K) R^K(x) \mathrm{d}F_0(x) \tag{56}$$

We rewrite (7) as

$$\left(\tilde{\Psi}'_{K(N)} \tilde{\Sigma}^{-1}_{K(N)} - \Psi'_{K(N)} \Sigma^{-1}_{K(N)}\right) V_{K(N)} \tag{57}$$

with

$$V_K = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} R^K(X_i)(T_i - p_K(X_i)) \tag{58}$$

Rewrite (57) as

$$\left(\tilde{\Psi}_{K(N)} - \Psi_{K(N)}\right)' \tilde{\Sigma}^{-1}_{K(N)} V_{K(N)} + \Psi'_{K(N)} \left(\tilde{\Sigma}^{-1}_{K(N)} - \Sigma^{-1}_{K(N)}\right) V_{K(N)} \tag{59}$$

By the properties of the matrix norm, the first term is bounded by

$$\left| \left(\tilde{\Psi}_{K(N)} - \Psi_{K(N)}\right)' \tilde{\Sigma}^{-1}_{K(N)} V_{K(N)} \right| \leq \frac{1}{\lambda_{min}(\tilde{\Sigma}_{K(N)})} \left\| V_{K(N)} \right\| \left\| \tilde{\Psi}_{K(N)} - \Psi_{K(N)} \right\| \tag{60}$$

By the mean value theorem (L'' is bounded)

$$\|\tilde{\Psi}_{K(N)} - \Psi_{K(N)}\| \leq \tag{61}$$

$$\leq C \int_{\mathcal{X}} \frac{|\mathrm{E}(Y|x)|}{p_0(x)} \left| L''(R^{K(N)}(x)'\overline{\pi}_{K(N)}) \right| \|R^{K(N)}(x)\|^2 \mathrm{d}F_0(x) \|\tilde{\pi}_{K(N)} - \pi_{K(N)}\| \leq$$

$$\leq C \int_{\mathcal{X}} \|R^{(K(N))}(x)\|^2 \mathrm{d}F_0(x) \|\tilde{\pi}_{K(N)} - \pi_{K(N)}\| \leq C\zeta(K(N))^2 \|\tilde{\pi}_{K(N)} - \pi_{K(N)}\|$$

because $E(Y^2) < \infty$ and p_0 is bounded from 0 on $\mathcal{X}$. Hence (60) is bounded by $C\zeta(K(N))^2\|\tilde{\pi}_{K(N)} - \pi_{K(N)}\|\|V_{K(N)}\|$

The second term is bounded by

$$
|\Psi'_{K(N)}\Sigma^{-1}_{K(N)}\left(\tilde{\Sigma}_{K(N)} - \Sigma_{K(N)}\right)\tilde{\Sigma}^{-1}_{K(N)}V_{K(N)}| \leq \tag{62}
$$

$$
\leq \frac{1}{\lambda_{min}(\tilde{\Sigma}_{K(N)})}\left\|V_{K(N)}\right\|\left\|\left(\tilde{\Sigma}_{K(N)} - \Sigma_{K(N)}\right)\Sigma^{-1}_{K(N)}\Psi_{K(N)}\right\|
$$

Now with $\hat{\Sigma}_K = \frac{1}{N}\sum_{i=1}^{N} L'(R^K(X_i)\pi_K)R^K(X_i)R^K(X_i)'$ and $W_K = \Sigma_K^{-1}\Psi_K$

$$
\left\|\left(\tilde{\Sigma}_{K(N)} - \Sigma_{K(N)}\right)W_{K(N)}\right\| \leq \left\|\left(\tilde{\Sigma}_{K(N)} - \hat{\Sigma}_{K(N)}\right)W_{K(N)}\right\| + \left\|\left(\hat{\Sigma}_{K(N)} - \Sigma_{K(N)}\right)W_{K(N)}\right\| \leq \tag{63}
$$

$$
\leq \frac{1}{N}\sum_{i=1}^{N}\left\|\left(R^{K(N)}(X_i)'(\tilde{\pi}_{K(N)} - \pi_{K(N)})L''(R^{K(N)}(X_i)'\overline{\pi}_{K(N)})R^{K(N)}(X_i)R^{K(N)}(X_i)'\right)W_{K(N)}\right\| +
$$

$$
+\frac{1}{\sqrt{N}}\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(L'(R^{K(N)}(X_i)'\pi_{K(N)})R^{K(N)}(X_i)R^{K(N)}(X_i)' - \right.\right.
$$

$$
\left.\left. -\mathrm{E}\left[L'(R^{K(N)}(X)'\pi_{K(N)})R^{K(N)}(X)R^{K(N)}(X)'\right]\right)W_{K(N)}\right\| \leq
$$

The first term on the right hand side is bounded by $C\zeta(K(N))^3\|W_{K(N)}\|\|\tilde{\pi}_{K(N)} - \pi_{K(N)}\| = C\zeta(K(N))^4\|\tilde{\pi}_{K(N)} - \pi_{K(N)}\|$, because $\|W_K\| \leq C\|\Psi_K\|$ and $|\Psi_K\| \leq C\zeta(K)$. For the second term note that

$$
\left\|\mathrm{Var}\left(L'(R^{K(N)}(X)'\pi_{K(N)})R^{K(N)}(X)R^{K(N)}(X)'W_{K(N)}\right)\right\| = \tag{64}
$$

$$
W'_{K(N)}\mathrm{Var}\left(L'(R^{K(N)}(X)'\pi_{K(N)})R^{K(N)}(X)R^{K(N)}(X)'\right)W_{K(N)} \leq
$$

$$
W'_{K(N)}\mathrm{E}\left(R^{K(N)}(X)R^{K(N)}(X)'\right)W_{K(N)} \leq C\zeta(K(N))^4
$$

Because

$$
V_K = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} R^K(X_i)(T_i - p_0(X_i)) + \frac{1}{\sqrt{N}}\sum_{i=1}^{N} R^K(X_i)(p_K(X_i) - p_0(X_i)) \tag{65}
$$

the variance matrix of V_K is

$$\mathrm{E}_X\left[p_0(X)(1-p_0(X))R^K(X)R^K(X)'\right]+\mathrm{E}_X\left[(p_K(X)-p_0(X))^2R^K(X)R^K(X)'\right]\leq \tag{66}$$

$$\leq C_1\zeta(K)^2+C_2K^{-\frac{s}{r}}\zeta(K)^2\leq C\zeta(K)^2 \tag{67}$$

and $\mathrm{E}(\|V_{K(N)}\|^2)\leq C\zeta(K(N))^2$.

Combining these results we find that by the Cauchy-Schwartz inequality for (60)

$$\mathrm{E}\left(\left|\left(\tilde{\Psi}_{K(N)}-\Psi_{K(N)}\right)'\tilde{\Sigma}^{-1}_{K(N)}V_{K(N)}\right|\right)\leq C\frac{\zeta(K(N))^{7/2}}{N^{1/2}} \tag{68}$$

and for (62)

$$\mathrm{E}\left(\left|\Psi'_{K(N)}\left(\tilde{\Sigma}^{-1}_{K(N)}-\Sigma^{-1}_{K(N)}\right)V_{K(N)}\right|\right)\leq C_1\frac{\zeta(K(N))^{9/2}}{N^{1/2}}+C_2\frac{\zeta(K(N))^2}{N^{1/2}} \tag{69}$$

Combining (68) and (69), we obtain the stochastic order of (7)

$$\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\hat{\delta}_{K(N)}(X_i)-\delta_{K(N)}(X_i))\frac{T_i-p_{K(N)}(X_i)}{\sqrt{p_{K(N)}(X_i)(1-p_{K(N)}(X_i))}}\right|=O_P\left(\frac{\zeta(K(N))^{9/2}}{N^{1/2}}\right) \tag{70}$$

0.5 Bound on (8)

Because (8) is a normalized sum of independently and identically distributed (but not zero mean) random variables, we find the stochastic order of (8) by deriving the order of its second moment. The key result is that this moment decreases to 0, if K increases without bounds. The reason is that if $\delta_0(x)$ is differentiable, $\delta_K(x)$ converges uniformly to $\delta_0(x)$ if K increases without bounds. The second moment is

$$\mathrm{E}\left[\left(\delta_K(X)\frac{T-p_K(X)}{\sqrt{p_K(X)(1-p_K(X))}}-\delta_0(X)\frac{T-p_0(X)}{\sqrt{p_0(X)(1-p_0(X))}}\right)^2\right] \tag{71}$$

By adding and subtracting

$$\delta_K(X)\frac{T-p_0(X)}{\sqrt{p_0(X)(1-p_0(X))}} \tag{72}$$

we obtain

$$\mathrm{E}\left[(\delta_K(X)-\delta_0(X))^2\left(\frac{T-p_0(X)}{\sqrt{p_0(X)(1-p_0(X))}}\right)^2\right]+ \tag{73}$$

$$+\mathrm{E}\left[\delta_K(X)^2\left(\frac{T-p_K(X)}{\sqrt{p_K(X)(1-p_K(X))}}-\frac{T-p_0(X)}{\sqrt{p_0(X)(1-p_0(X))}}\right)^2\right]+ \tag{74}$$

$$+\mathrm{E}\left[(\delta_K(X)-\delta_0(X))\delta_K(X)\left(\frac{T-p_0(X)}{\sqrt{p_0(X)(1-p_0(X))}}\right)\cdot\right. \tag{75}$$

$$\left.\cdot\left(\frac{T-p_K(X)}{\sqrt{p_K(X)(1-p_K(X))}}-\frac{T-p_0(X)}{\sqrt{p_0(X)(1-p_0(X))}}\right)\right]$$

First consider (73). From (11) and (12) we see that $\delta_K(x)$ is the least squares projection of $\delta_0(x)$ on $R^K(x)\sqrt{p_K(x)(1-p_K(x))}$. If we assume that $\delta_0(x)$ is t times continuously differentiable, then (compare with (27))

$$\sup_{x\in\mathcal{X}}|\delta_0(x)-\delta_K(x)|<CK^{-\frac{t}{r}} \tag{76}$$

Hence

$$\mathrm{E}\left[(\delta_K(X)-\delta_0(X))^2\left(\frac{T-p_0(X)}{\sqrt{p_0(X)(1-p_0(X))}}\right)^2\right]=\mathrm{E}_X\left[(\delta_K(X)-\delta_0(X))^2\right]\le CK^{-2\frac{t}{r}} \tag{77}$$

Next consider (74) that is equal to

$$\mathrm{E}_X\left[\delta_K(X)^2\left(\frac{\sqrt{p_0(X)(1-p_0(X))}-\sqrt{p_K(X)(1-p_K(X))}}{\sqrt{p_0(X)p_K(X)(1-p_0(X))(1-p_K(X))}}\right)^2\right]+ \tag{78}$$

$$+\mathrm{E}_X\left[\delta_K(X)^2\frac{(p_K(X)-p_0(X))^2}{p_K(X)(1-p_K(X))}\right] \tag{79}$$

Now from (76)

$$\delta_K(x)^2 \leq \delta_0(x)^2 + |\delta_0(x)|CK^{-\frac{t}{r}} \tag{80}$$

Hence

$$\mathrm{E}_X\left[\delta_K(X)^2\frac{(p_K(X)-p_0(X))^2}{p_K(X)(1-p_K(X))}\right] \leq \tag{81}$$

$$\leq \mathrm{E}_X\left[\delta_0(X)^2\frac{(p_K(X)-p_0(X))^2}{p_K(X)(1-p_K(X))}\right] + CK^{-\frac{t}{r}}\mathrm{E}_X\left[|\delta_0(X)|\frac{(p_K(X)-p_0(X))^2}{p_K(X)(1-p_K(X))}\right] \leq \tag{82}$$

$$\leq \mathrm{E}_X\left[\frac{\mathrm{E}(Y\mid X)^2}{p_0(X)^2}\frac{p_0(X)(1-p_0(X))}{p_K(X)(1-p_K(X))}(p_K(X)-p_0(X))^2\right] + \tag{83}$$

$$+CK^{-\frac{t}{r}}\mathrm{E}_X\left[\frac{|\mathrm{E}(Y\mid X)|}{p_0(X)}\frac{\sqrt{p_0(X)(1-p_0(X))}}{p_K(X)(1-p_K(X))}(p_K(X)-p_0(X))^2\right] \tag{84}$$

Because p_0 is bounded from 0 and 1 on $\mathcal{X}$ and hence also, if K is sufficiently large, p_K by virtue of (33) and $\mathrm{E}(Y) < \infty$, we have again by (33) that

$$\mathrm{E}_X\left[\delta_K(X)^2\frac{(p_K(X)-p_0(X))^2}{p_K(X)(1-p_K(X))}\right] \leq C_1\zeta(K)^2K^{-\frac{s}{r}} + C_2\zeta(K)^2K^{-2\frac{t}{r}-\frac{s}{r}} \leq C\zeta(K)^2K^{-\frac{s}{r}} \tag{85}$$

By an analogous argument we have

$$\mathrm{E}_X\left[\delta_K(X)^2\left(\frac{\sqrt{p_0(X)(1-p_0(X))}-\sqrt{p_K(X)(1-p_K(X))}}{\sqrt{p_0(X)p_K(X)(1-p_0(X))(1-p_K(X))}}\right)^2\right] \leq C\zeta(K)^2K^{-\frac{s}{r}} \tag{86}$$

Finally, consider (75). Because $\delta_K(x)$ is bounded and p_0, p_K are bounded from 0 and 1 on $\mathcal{X}$, we have that (75) is bounded by

$$C_1\mathrm{E}(|\delta_K(X)-\delta_0(X)|) \leq CK^{-\frac{t}{r}} \tag{87}$$

Hence from (77), (86), and (87)

$$\mathrm{E}\left[\left(\delta_K(X)\frac{T-p_K(X)}{\sqrt{p_K(X)(1-p_K(X))}}-\delta_0(X)\frac{T-p_0(X)}{\sqrt{p_0(X)(1-p_0(X))}}\right)^2\right]\leq \tag{88}$$

$$\leq C_1K^{-\frac{t}{r}}+C_2\zeta(K)^2K^{-\frac{s}{r}}$$

and this gives for (8)

$$\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\delta_{K(N)}(X_i)\frac{T_i-p_{K(N)}(X_i)}{\sqrt{p_{K(N)}(X_i)(1-p_{K(N)}(X_i))}}-\delta_0(X_i)\frac{T_i-p_0(X_i)}{\sqrt{p_0(X_i)(1-p_0(X_i))}}\right)\right| = \tag{89}$$

$$=O_P\left(\max\left(K(N)^{-\frac{1}{2}\frac{t}{r}},\zeta(K(N))K(N)^{-\frac{s}{2r}}\right)\right)$$

0.6 Combining the bounds

From (22),(45), (55), (70), and (89), we obtain

$$\left|\sqrt{N}(\hat{\beta}_{ew}-\beta_0)-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{\left(\frac{T_i.Y_i}{p_0(X_i)}-\beta_0\right)-\frac{\mathrm{E}(Y\mid X_i)}{p_0(X_i)}(T_i-p_0(X_i))\right\}\right|= \tag{90}$$

$$=O_P\left(\frac{\zeta(K(N))^3}{\sqrt{N}}\right)+O_P\left(\sqrt{N}\zeta(K(N))^2K(N)^{-\frac{s}{r}}\right)+O_P\left(\zeta(K(N))^{5/2}K(N)^{-\frac{s}{2r}}\right)+$$

$$+O_P(\zeta(K(N))K(N)^{-\frac{s}{2r}})+O_P\left(\frac{\zeta(K(N))^2}{\sqrt{N}}\right)+O(\sqrt{N}\zeta(K(N))K(N)^{-\frac{s}{2r}})+O_P\left(\frac{\zeta(K(N))^{9/2}}{\sqrt{N}}\right)$$

$$+O_P\left(\max\left(K(N)^{-\frac{1}{2}\frac{t}{r}},\zeta(K(N))K(N)^{-\frac{s}{2r}}\right)\right)=$$

$$=O_P\left(\sqrt{N}\zeta(K(N))^2K(N)^{-\frac{s}{r}}\right)+O_P\left(\zeta(K(N))^{5/2}K(N)^{-\frac{s}{2r}}\right)+O_P\left(\frac{\zeta(K(N))^{9/2}}{\sqrt{N}}\right)$$

Note that the second term of the final expression is a bias term, the third a variance term, and the first a combination of a variance and bias term.

As noted $\zeta(K)$ depends on the sequence of approximating functions. For power series we have $\zeta(K) = O(K)$. If we consider sequences $K(N) = N^c$ we can find the range of c for which (90) is $o_P(1)$. Substitution in the right-hand side of (90) gives that the first term on the right hand side requires that $c > \frac{1}{2(s/r-2)}$, the second that $s/r > 5$ and the third that $c < 1/9$. These inequalities can be simultaneously satisfied if $s/r \geq 7$. $\square$

Addendum to Proof of Theorem 2:

The last step in Theorem 2 follows from the fact that we can write the difference between (18) and (17) as

$$\frac{1}{N}\sum_{i=1}^{N}\Bigg(V(T_i, Y_i, X_i) - \frac{T_i Y_i}{\hat{p}_K(X_i)p_0(X_i)}(\hat{p}_K(X_i) - p_0(X_i)) \tag{91}$$

$$+(\beta_0 - \hat{\beta}_{ew}) + (\hat{\alpha}_K(T_i, X_i) - \alpha(T_i, X_i))\Bigg)^2$$

$$= \frac{1}{N}\sum_{i=1}^{N} V(T_i, Y_i, X_i)^2 \tag{92}$$

$$+\frac{1}{N}\sum_{i=1}^{N}\frac{T_i Y_i^2}{\hat{p}_K(X_i)^2 p_0(X_i)^2}(\hat{p}_K(X_i) - p_0(X_i))^2 + (\beta_0 - \hat{\beta}_{ew})^2 \tag{93}$$

$$\frac{1}{N}\sum_{i=1}^{N}(\hat{\alpha}_K(T_i, X_i) - \alpha(T_i, X_i))^2 \tag{94}$$

$$-2\frac{1}{N}\sum_{i=1}^{N} V(T_i, Y_i, X_i)\frac{T_i Y_i}{\hat{p}_K(X_i)p_0(X_i)}(\hat{p}_K(X_i) - p_0(X_i)) \tag{95}$$

$$+2(\beta_0 - \hat{\beta}_{ew})\frac{1}{N}\sum_{i=1}^{N} V(T_i, Y_i, X_i) \tag{96}$$

$$+2\frac{1}{N}\sum_{i=1}^{N} V(T_i, Y_i, X_i)(\hat{\alpha}_K(T_i, X_i) - \alpha(T_i, X_i)) \tag{97}$$

$$-2(\beta_0 - \hat{\beta}_{ew})\frac{1}{N}\sum_{i=1}^{N}\frac{T_i Y_i}{\hat{p}_K(X_i)p_0(X_i)}(\hat{p}_K(X_i) - p_0(X_i)) \tag{98}$$

$$+2\frac{1}{N}\sum_{i=1}^{N}\frac{T_i Y_i}{\hat{p}_K(X_i)p_0(X_i)}(\hat{p}_K(X_i) - p_0(X_i))(\hat{\alpha}_K(T_i, X_i) - \alpha(T_i, X_i)) \tag{99}$$

$$+2(\beta_0 - \hat{\beta}_{ew})\frac{1}{N}\sum_{i=1}^{N}(\hat{\alpha}_K(T_i, X_i) - \alpha(T_i, X_i)) \tag{100}$$

with

$$V(t, y, x) = \frac{ty}{p_0(x)} - \beta_0 + \alpha(t, x) \tag{101}$$

(92) is a sample average that converges to the variance V. If the assumptions of Theorem 1 hold, the bound on the sum of (93) to (100) is given by the bound on (97) which is (78). Under the rates specified in Theorem 1 this bound is $o_p(1)$. Hence, (18) is a consistent estimator of the variance (17).