Supplement Material for "Inverse Probability Weighted Estimation of Local Average Treatment Effects: A Higher Order MSE Expansion"

Stephen G. Donald

Department of Economics

University of Texas at Austin

Yu-Chin Hsu

Institute of Economics

Academia Sinica

Robert P. Lieli

Department of Economics

Central European University, Budapest

and Magyar Nemzeti Bank

A The proof of Theorem 3.1

We give a detailed proof for the estimator $\hat{\tau}$. We analyze the numerator and denominator of $\hat{\tau}$ separately. Let

$$\widehat{\Delta} = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{Z_iY_i}{\hat{q}(X_i)} - \frac{(1-Z_i)Y_i}{1-\hat{q}(X_i)}\right\}, \qquad \widehat{\Gamma} = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{Z_iD_i}{\hat{q}(X_i)} - \frac{(1-Z_i)D_i}{1-\hat{q}(X_i)}\right\}.$$

so that $\hat{\tau} = \widehat{\Delta}/\widehat{\Gamma}$. The asymptotic properties of $\hat{\Delta}$ and $\hat{\Gamma}$ are established in the following lemma.

Lemma A.1 *Under the conditions of Theorem 3.1,* $\sqrt{n}(\widehat{\Delta}-\Delta) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\delta(Y_i,D_i,Z_i,X_i)+o_p(1)$ *and* $\sqrt{n}(\widehat{\Gamma}-\Gamma) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\gamma(Y_i,D_i,Z_i,X_i)+o_p(1)$, *where*

$$\delta(Y_i,D_i,Z_i,X_i) = \frac{Z_iY_i}{q(X_i)} - \frac{(1-Z_i)Y_i}{1-q(X_i)} - \Delta - \left(\frac{m_1(X_i)}{q(X_i)} + \frac{m_0(X_i)}{1-q(X_i)}\right)(Z_i - q(X_i))$$
$$\gamma(Y_i,D_i,Z_i,X_i) = \frac{Z_iD_i}{q(X_i)} - \frac{(1-Z_i)D_i}{1-q(X_i)} - \Gamma - \left(\frac{\mu_1(X_i)}{q(X_i)} + \frac{\mu_0(X_i)}{1-q(X_i)}\right)(Z_i - q(X_i)).$$

Taking Lemma A.1 as given for now, we can use the first order Taylor expansion of the bivariate function $f(\hat{\Delta},\hat{\Gamma}) = \hat{\Delta}/\hat{\Gamma}$ around the point (Δ,Γ) to write

$$\sqrt{n}(\hat{\tau}-\tau) = \sqrt{n}\left(\frac{\widehat{\Delta}}{\widehat{\Gamma}} - \frac{\Delta}{\Gamma}\right) = \frac{1}{\Gamma}\sqrt{n}(\widehat{\Delta}-\Delta) - \frac{\tau}{\Gamma}\sqrt{n}(\widehat{\Gamma}-\Gamma) + o_p(1). \tag{A.1}$$

Substituting the influence function representations given in Lemma A.1 into (A.1) establishes the representation we need. It is easy to check that under Assumption 2.1 (i), $E[\psi(Y,D,Z,X)] = 0$ and $E[\psi^2(Y,D,Z,X)] < \infty$. Applying the Lindeberg-Levy CLT and Slutsky's theorem shows $\sqrt{n}(\hat{\tau}-\tau) \xrightarrow{d} \mathcal{N}(0,\mathcal{V})$.

We complete the proof of Theorem 3.1 by verifying Lemma A.1. ∎

The proof of Lemma A.1 Our argument is based on Ichimura and Linton (2005) with the generalization that X is allowed to be an r-dimensional vector rather than a scalar. For a matrix $A = (a_{ij})$ we write $\|A\|_\infty = \sup|a_{ij}|$ and $\|A\|_1 = \sum|a_{ij}|$.

STEP 1 (Some properties of $\hat{q}(X_i)$). For $x \in \mathbb{R}^r$ and $\lambda \in \mathbb{N}^r$ define $x^\lambda = x^{\lambda_1}\cdot\ldots\cdot x_r^{\lambda_r} \in \mathbb{R}$. For a non-negative integer ℓ, let $x^{\Lambda(\ell)}$ denote the vector $(x^\lambda)_{\lambda_1+\ldots+\lambda_r=\ell}$ (along with some rule to order these elements). Thus, $x^{\Lambda(\ell)}$ contains all polynomial terms of order exactly ℓ that can be constructed from the components of x, and is interpreted as a row vector if x is a row vector and as a column vector if x is a column vector. E.g., $x^{\Lambda(0)} = 1$, $(x_1, \ldots, x_r)^{\Lambda(1)} = (x_1, \ldots, x_r)$, $(x_1, x_2)^{\Lambda(2)} = (x_1^2, x_2^2, x_1 x_2)$, etc. For each observation X_t on the vector of covariates, we define

$$\tilde{X}_t^{(i)} = [(X_t' - X_i')^{\Lambda(0)}, (X_t' - X_i')^{\Lambda(1)}, \ldots, (X_t' - X_i')^{\Lambda(r)}]'.$$

Then the leave-one-out local polynomial regression estimator of $q(X_i)$ is the first component of the vector $\hat{\beta}$ that solves $\min_\beta \sum_{t:t\neq i} K\left(\frac{X_t - X_i}{h}\right)(Z_t - \tilde{X}_t^{(i)\prime}\beta)^2$. Letting e_1 denote the first unit vector having the same dimension as $X_t^{(i)}$, we can write this estimator as

$$\hat{q}(X_i) = e_1' \left(\sum_{t:t\neq i} K\left(\frac{X_t - X_i}{h}\right) \tilde{X}_t^{(i)} \tilde{X}_t^{(i)\prime} \right)^{-1} \sum_{t:t\neq i} K\left(\frac{X_t - X_i}{h}\right) \tilde{X}_t^{(i)} Z_t = \sum_{t:t\neq i} \omega_{it} Z_t,$$

where ω_{it} depends only on $X, ..., X_n$ and is given by

$$\omega_{it} = e_1' \left(\sum_{j:j\neq i} K\left(\frac{X_j - X_i}{h}\right) \tilde{X}_j^{(i)} \tilde{X}_j^{(i)\prime} \right)^{-1} \tilde{X}_t^{(i)} K\left(\frac{X_t - X_i}{h}\right). \tag{A.2}$$

The first property we will need in later arguments is a bound on $|\omega_{it} - \omega_{ti}|$. By Assumption 3.4, $\omega_{it} = \omega_{ti} = 0$ for $\|X_t - X_i\|_\infty > h$. Now assume $\|X_t - X_i\|_\infty \leq h$. Let $H = diag(1, hu_1, ..., h^r u_r)$, where u_j is a vector of ones with the same dimensionality as $(X_t' - X_i')^{\Lambda(j)}$. Then, noting that $e_1' H = e_1'$, we can write

$$\omega_{it} = e_1' \left(H^{-1} \frac{1}{nh^r} \sum_{j:j\neq i} K\left(\frac{X_j - X_i}{h}\right) \tilde{X}_j^{(i)} \tilde{X}_j^{(i)\prime} H^{-1} \right)^{-1} H^{-1} \frac{1}{nh^r} \tilde{X}_t^{(i)} K\left(\frac{X_t - X_i}{h}\right).$$

Let the matrix inside the inverse operator be denoted as $\widehat{\mathcal{K}}(X_i)$. Then

$$\begin{aligned}
|\omega_{it} - \omega_{ti}| &= \frac{1}{nh^r} \left| K\left(\frac{X_i - X_t}{h}\right) \right| \times \left| e_1' \widehat{\mathcal{K}}^{-1}(X_i) H^{-1} \tilde{X}_t^{(i)} - e_1' \widehat{\mathcal{K}}^{-1}(X_t) H^{-1} \tilde{X}_i^{(t)} \right| \\
&\leq \frac{1}{nh^r} \left| K\left(\frac{X_i - X_t}{h}\right) \right| \\
&\times \left\{ \left| e_1' [\widehat{\mathcal{K}}^{-1}(X_i) - \widehat{\mathcal{K}}^{-1}(X_t)] H^{-1} \tilde{X}_t^{(i)} \right| + \left| e_1' \widehat{\mathcal{K}}^{-1}(X_t) H^{-1} [\tilde{X}_t^{(i)} - \tilde{X}_i^{(t)}] \right| \right\}. \qquad (A.3)
\end{aligned}$$

The first term in the braces in (A.3) is bounded as follows. The elements of $\widehat{\mathcal{K}}(X_i)$ are of the form

$$\frac{1}{nh^{r+\sum \lambda_k}} \sum_{j:j\neq i} (X_j - X_i)^\lambda K\left(\frac{X_j - X_i}{h}\right) \tag{A.4}$$

for some r-vector of nonnegative integers λ with $0 \leq \sum \lambda_k \leq 2r$. Using arguments similar to those in, e.g., Section 3.7 of Fan and Gijbels (1996), one can show that (A.4) converges in probability to $f(X_i)\int u^\lambda K(u)du$ uniformly in X_i. Hence $\sup_i \|\widehat{\mathcal{K}}(X_i) - f(X_i)\mathcal{K}\|_\infty = o_p(1)$ for a constant, symmetric matrix $\mathcal{K}$ that only depends on the kernel K. By continuity and $\mathcal{X}$ compact, it follows that $\sup_i \|\widehat{\mathcal{K}}^{-1}(X_i) - \frac{1}{f(X_i)}\mathcal{K}^{-1}\|_\infty = o_p(1)$. Since f is continuously differentiable and is bounded away from zero, $|f^{-1}(x_1) - f^{-1}(x_2)| \le C\|x_1 - x_2\|_\infty$ for all $x_1, x_2 \in \mathcal{X}$. Combining these observations yields

$$
\begin{aligned}
&\|\widehat{\mathcal{K}}^{-1}(X_i) - \widehat{\mathcal{K}}^{-1}(X_t)\|_\infty \\
\le& \sup_i \|\widehat{\mathcal{K}}^{-1}(X_i) - f^{-1}(X_i)\mathcal{K}^{-1}\|_\infty + \|f^{-1}(X_i)\mathcal{K}^{-1} - f^{-1}(X_t)\mathcal{K}^{-1}\|_\infty + \sup_t \|\widehat{\mathcal{K}}^{-1}(X_t) - f^{-1}(X_t)\mathcal{K}^{-1}\|_\infty \\
\le& C\|X_i - X_t\|_\infty + o_p(1) = Ch + o_p(1) \equiv M_n,
\end{aligned}
$$

where $M_n = o_p(1)$ and is independent of X_i and X_t. Hence the first term in the braces in (A.3) is bounded by $M_n\|H^{-1}X_t^{(i)}\|_1 \le \tilde{M}_n$, where $\tilde{M}_n$ incorporates a multiplicative constant that only depends on r and bounds $\|H^{-1}X_t^{(i)}\|_1$ (each component of $H^{-1}X_t^{(i)}$ is bounded by one since $\|X_t - X_i\|_\infty \le h$).

The second term within the braces in (A.3) is bounded as follows. Suppose that r is even. Then

$$
H^{-1}[\tilde{X}_t^{(i)} - \tilde{X}_i^{(t)}] = \left(0, \quad 2\left(\frac{X_t' - X_i'}{h}\right)^{\Lambda(1)}, \quad z_2 \quad , \quad 2\left(\frac{X_t' - X_i'}{h}\right)^{\Lambda(3)}, \quad \ldots \quad , \quad z_r\right)', \tag{A.5}
$$

where the z_j, j even, are zero vectors with the same dimensionality as $(X_t' - X_i')^{\Lambda(j)}$. (The only difference when r is odd is that the last term in this alternating partition is not a zero vector.) Note that each component of (A.5) is bounded by 2 since $\|X_t - X_i\|_\infty \le h$. Therefore it is possible to write

$$
\left|e_1'\widehat{\mathcal{K}}^{-1}(X_t)H^{-1}[\tilde{X}_t^{(i)} - \tilde{X}_i^{(t)}]\right| \le \left|\frac{1}{f(X_t)}e_1'\mathcal{K}^{-1}H^{-1}[\tilde{X}_t^{(i)} - \tilde{X}_i^{(t)}]\right| + R_n,
$$

where $R_n = o_p(1)$ and does not depend on X_i or X_t. By the symmetry properties of the kernel (Assumption 3.4), the first row of the matrix $\mathcal{K}$ has zeros precisely at those positions at which the vector $H^{-1}[\tilde{X}_t^{(i)} - \tilde{X}_i^{(t)}]$ is nonzero. A straightforward linear algebra argument (available on request) shows that the first row of the matrix $\mathcal{K}^{-1}$ also has zeros at the same positions. Hence, the second term in the braces in (A.3) is simply bounded by R_n. Combining the bounds on the components of (A.3) shows that

$$
|\omega_{it} - \omega_{ti}| \le \frac{\tilde{M}_n + R_n}{nh^r}\left|K\left(\frac{X_t - X_i}{h}\right)\right|, \tag{A.6}
$$

where $\tilde{M}_n + R_n = o_p(1)$ and does not depend on X_i or X_t. Finally, observe that the inequality holds for any value of h, not just for $\|X_t - X_i\|_\infty \le h$. This is the bound we will need.

The second property of $\hat{q}(X_i)$ that we will make use of is its uniform convergence rate. As shown by Masry (1996), the "include all" version of the estimator satisfies

$$\sup_{x \in \mathcal{X}} |\hat{q}(x) - q(x)| = O_p\left(h^{r+1} + \sqrt{\frac{\log n}{nh^r}}\right).$$

Given the range of h in Assumption 3.5, it follows that $\sup_i |\hat{q}(X_i) - q(X_i)| = o_p(n^{-1/4})$ for the include all as well as the leave-one-out version.

STEP 2 (Expanding $\hat{\Delta}$). We define notation similar to that in Ichimura and Linton (2005). Let $w = (y, d, z, x)$ and

$$\Psi(w, \Delta, q) \equiv \frac{zy}{q} - \frac{(1-z)y}{1-q} - \Delta.$$

Let Ψ_q and Ψ_{qq} denote the partial derivative of Ψ w.r.t. the argument q, and let $W_i = (Y_i, D_i, Z_i, X_i)$. Then

$$\Psi(W_i, \Delta, q(X_i)) = \frac{Z_i Y_i}{q(X_i)} - \frac{(1-Z_i)Y_i}{1-q(X_i)},$$
$$\Psi_q(W_i, \Delta, q(X_i)) = -\left(\frac{Z_i Y_i}{q^2(X_i)} + \frac{(1-Z_i)Y_i}{(1-q(X_i))^2}\right),$$
$$\Psi_{qq}(W_i, \Delta, q(X_i)) = \frac{2Z_i Y_i}{q^3(X_i)} - \frac{2(1-Z_i)Y_i}{(1-q(X_i))^3},$$

and we further define

$$S_q(X_i) = E[\Psi_q(W_i, \Delta, q(X_i))|X_i] = -\left(\frac{m_1(X_i)}{q(X_i)} + \frac{m_0(X_i)}{1-q(X_i)}\right),$$
$$\zeta_i = \Psi_q(W_i, \Delta, q(X_i)) - S_q(X_i),$$
$$\epsilon_i = Z_i - q(X_i),$$
$$\beta_n(X_i) = E[\hat{q}(X_i)|X, ..., X_n] - q(X_i) = \sum_{j:i\neq j} \omega_{ij} q(X_j) - q(X_i),$$

where the last quantity is the bias of the estimator conditional on $X, \ldots, X_n$.

By a Taylor series expansion around $q(X_i)$,

$$
\begin{aligned}
\sqrt{n}(\widehat{\Delta}-\Delta) =& \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Psi(W_i,\Delta,\hat{q}(X_i)) \\
=& \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Psi(W_i,\Delta,q(X_i)) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Psi_q(W_i,\Delta,q(X_i))(\hat{q}(X_i)-q(X_i)) \\
& + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Psi_{qq}(W_i,\Delta,q^*(X_i))(\hat{q}(X_i)-q(X_i))^2 \\
=& \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Psi(W_i,\Delta,q(X_i)) \\
& + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}S_q(X_i)(\hat{q}(X_i)-q(X_i)) \\
& + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\zeta_i(\hat{q}(X_i)-q(X_i)) \\
& + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Psi_{qq}(W_i,\Delta,q^*(X_i))(\hat{q}(X_i)-q(X_i))^2 \\
\equiv& J_0+J_1+J_2+J_3,
\end{aligned}
$$

where $q^*(X_i)$ is a value between $\hat{q}(X_i)$ and $q(X_i)$ for all i, and the J's are defined line by line. We further expand the J_1 term as

$$
\begin{aligned}
J_1 &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}S_q(X_i)(\hat{q}(X_i)-q(X_i)) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}S_q(X_i)\Big(\sum_{j:j\neq i}\omega_{ij}Z_j - q(X_i)\Big) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}S_q(X_i)\Big(\sum_{j:j\neq i}\omega_{ij}(\epsilon_j+q(X_j)) - q(X_i)\Big) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}S_q(X_i)\epsilon_i + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}S_q(X_i)\Big(\sum_{j:j\neq i}\omega_{ij}\epsilon_j - \epsilon_i\Big) \\
&\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}S_q(X_i)\Big(\sum_{j:j\neq i}\omega_{ij}q(X_j) - q(X_i)\Big) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}S_q(X_i)\epsilon_i + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\epsilon_i\Big(\sum_{j:j\neq i}\omega_{ji}S_p(X_j) - S_p(X_i)\Big) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}S_q(X_i)\beta_n(X_i) \\
&\equiv J_{11}+J_{12}+J_{13}.
\end{aligned}
$$

STEP 3 (Evaluating J_0, J_{11}, J_{12}, J_{13}, J_2 and J_3). By the central limit theorem, J_0 and J_{11} are $O_p(1)$ and together they give the influence function representation in Lemma A.1. We will show that the rest of the terms are $o_p(1)$.

For J_{12}, we claim that $\omega_{ij} \approx \omega_{ji}$ in that $\sup_i \left| \sum_{j:i\neq j} (\omega_{ji} - \omega_{ij}) S_p(X_j) \right| = o_p(1)$. By the bound in (A.6),

$$\sup_i \Big| \sum_{j:i\neq j} (\omega_{ji} - \omega_{ij}) S_p(X_j) \Big| \leq \sup_i \sum_{j:i\neq j} |(\omega_{ji} - \omega_{ij})||S_p(X_j)|$$
$$\leq C(\tilde{M}_n + R_n) \sup_i \sum_{j:j\neq i} \frac{1}{nh^r} \Big| K\Big(\frac{X_j - X_i}{h}\Big) \Big| = o_p(1) \cdot O_p(1) = o_p(1).$$

The second inequality holds since $S_p(x)$ is bounded on $\mathcal{X}$. Further,

$$\sup_{x\in\mathcal{X}} \left| \sum_{i=1}^{n} \frac{1}{nh^r} \Big| K\Big(\frac{X_j - X_i}{h}\Big) \Big| - f(x) \int |K(u)| du \right| = o_p(1),$$

which implies $\sup_i \sum_{j:j\neq i} \frac{1}{nh^r} \left| K\left(\frac{X_j - X_i}{h}\right) \right| = O_p(1)$. Given

$$\sup_i \Big| \sum_{j:j\neq i} \omega_{ij} S_p(X_j) - S_p(X_i) \Big| = o_p(1),$$

it is true that conditional on the sample path of the X_i, with probability approaching one, $\sum_{j:j\neq i} \omega_{ji} S_p(X_j) - S_p(X_i)$ is uniformly bounded over i and converges to zero uniformly over i. Also, the ϵ_i are mutually independent conditional on the sample path of the X_i. Hence, conditional on the sample path of the X_i with probability approaching one,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \epsilon_i \Big(\sum_{j:j\neq i} \omega_{ji} S_p(X_j) - S_p(X_i) \Big) = o_p(1),$$

which is sufficient to show that $J_{12} = o_p(1)$.

For J_{13}, observe that $\beta_n(x)$ is the (conditional) bias of $\hat{q}(x)$, which is of order h_n^{r+1} uniformly (?). By the assumptions on h_n, we have $\sup_{x\in\mathcal{X}} |\beta_n(x)| = o_p(n^{-1/2})$. It follows that

$$|J_{13}| = \Big| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} S_p(X_i) \beta_n(X_i) \Big| \leq \sup_{x\in\mathcal{X}} \big| \sqrt{n} \beta_n(x) \big| \frac{1}{n} \sum_{i=1}^{n} |S_p(X_i)| = o_p(1) \cdot O_p(1) = o_p(1).$$

For J_2, observe that $\sup_{x\in\mathcal{X}} |\hat{q}(x) - q(x)| = o_p(1)$ and argue similarly as in showing $J_{12} = o_p(1)$.

Finally, for J_3. Given that $\hat{q}(x)$ is uniformly bounded in probability on $\mathcal{X}$, and $q^*(X_i)$ is between $\hat{q}(X_i)$ and $q(X_i)$, it follows that $q^*(X_i)$ is uniformly bounded and also bounded away from zero in probability. Also, $\sup_i |\hat{q}(X_i) - q(X_i)| = o_p(n^{-1/4})$, so $\sup_i n^{1/2} (\hat{q}(X_i) - q(X_i))^2 = o_p(1)$. Hence,

$$\Big| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Psi_{qq}(W_i, \Delta, q^*(X_i)) (\hat{q}(X_i) - q(X_i))^2 \Big|$$
$$\leq \Big(\sup_i \frac{1}{\sqrt{n}} (\hat{q}(X_i) - q(X_i))^2 \Big) \frac{1}{n} \sum_{i=1}^{n} \Big| \Psi_{qq}(W_i, \Delta, q^*(X_i)) \Big| = o_p(1) \cdot O_p(1) = o_p(1).$$

As a result, we have

$$\sqrt{n}(\widehat{\Delta} - \Delta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Psi(W_i, \Delta, q(X_i)) + S_q(X_i)\epsilon_i + o_p(1) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \delta(Y_i, D_i, Z_i, X_i) + o_p(1),$$

where the function $\delta(\cdot)$ is as defined in Lemma A.1. ∎

References

Fan, J. and I. Gijbels (1996): "Local Polynomial Modelling and Its Applications." Chapman & Hall.

Ichimura, H. and O. Linton (2005): "Asymptotic Expansions for Some Semiparametric Program Evaluation Estimators," in Identification and Inference for Econometric Models: essays in honor of Thomas Rothenberg Andrews, ed. by D.W.K. and Stock, J., Cambridge University Press.

Masry, E. (1996): "Multivariate Local Polynomial Regression for Time Series: Uniform Strong Consistency and Rates." *Journal of Time Series Analysis*, **17**, 571-599.