Statistica Sinica

Propensity score regression for causal inference with treatment heterogeneity

Peng Wu[a,b]†, Shasha Han[c,d]†, Xingwei Tong[b] and Runze Li[e]∗

[a]*Beijing Technology and Business University,* [b]*Beijing Normal University,* [c]*Chinese Academy of Medical Sciences and Peking Union Medical College,* [d]*Peking University, and* [e]*The Pennsylvania State University*

This Supplementary Material consists of Sections 1-8, where Sections 1-5 give the technical proofs of Propositions 1-2 and Theorems 1-2 in the manuscript. Sections 6-7 extend the proposed method to cases of multidimensional covariate x^l and discrete x^l. Sections 8 contain the summary statistics of the datasets in the empirical application.

1. Proofs of Proposition 1 and Proposition 2

Proof of Proposition 1. Since $e(X)$ is a balancing score (Rosenbaum and Rubin, 1983), we have $D \perp\!\!\!\perp (X^{-l}, X^l) \mid e(X)$. Combine it with the ignorability assumption $D \perp\!\!\!\perp \{Y(1), Y(0)\}|e(X)$, we have $D \perp\!\!\!\perp \{Y(1), Y(0)\}|X^l, e(X)$ by applying

†contributed equally.

∗correspond to: rzli@psu.edu

the corollary 15.2.1 of Anderson (2003). Then,

$$
\begin{aligned}
\mathbb{E}[Y|D, X^l, e] &= \mathbb{E}[Y|D, X^l, e] \\
&= \mathbb{E}[DY(1) + (1-D)Y(0)|D, X^l, e] \\
&= \mathbb{E}[D\{Y(1) - Y(0)\} + Y(0)|D, X^l, e] \\
&= \mathbb{E}[D\{Y(1) - Y(0)\}|D, X^l, e] + \mathbb{E}[Y(0)|D, X^l, e] \\
&= D\mathbb{E}[Y(1) - Y(0)|X^l, e] + \mathbb{E}[Y(0)|X^l, e] \\
&= D\beta(X^l, e) + \mathbb{E}[Y(0)|X^l, e].
\end{aligned}
$$

□

Proof of Proposition 2.

$$
\begin{aligned}
\tau(x^l) &= \mathbb{E}[Y(1) - Y(0)|X^l = x^l] \\
&= \mathbb{E}[\mathbb{E}\{Y(1) - Y(0)|X^l, e\}|X^l = x^l] \\
&= \mathbb{E}[\beta(X^l, e)|X^l = x^l].
\end{aligned}
$$

□

2. Regularity Assumptions

For technical proof, we make the following assumptions.

Assumption 1. The propensity score model can be written as $e(X) = g(X^\intercal\alpha)$. α is the true unknown parameter. $g(\cdot)$ can be a known function (e.g., generalized linear model) or an unknown smooth function (e.g., single index model).

(i) The estimates of α, $\hat{\alpha}$, satisfies $\hat{\alpha} - \alpha = O_p(N^{-1/2})$;

(ii) The second-order derivative of g is uniformly bounded, i.e., $\sup_t |g''(t)|$ is bounded.

Assumption 2. (Conditions for nonparametric estimation of $\beta(x^l, e)$ in Step 1)

(i) $h_1 \to 0$, $h_2 \to 0$, $Nh_1h_2 \to \infty$, $\sqrt{Nh_1h_2}(h_1 + h_2)^3 \to 0$, and $Nh_2^4 \to \infty$ as $N \to \infty$.

(ii) Denote $f(x^l, e)$ to be the joint probability density for (X^l, e). Assume that $f(x^l, e)$ has continuous partial derivative of order one with respect to x^l and e, and $f(x^l, e) > \delta > 0$ for some positive constant δ. This condition implies that $f(x^l, e)$ have a bounded support.

(iii) $\beta(x^l, e)$ and $\mathbb{E}[Y(0)|X^l = x^l, e(X) = e]$ are twice differentiable.

(iv) $\sup_{d,z} \mathbb{E}(|\xi|^{2+\eta}\big|D = d, X^l = x^l, e = e) < \infty$ for some $\eta > 0$.

(v) Let $\sigma^2(d, x^l, e) := \text{var}(\xi|D = d, X^l = x^l, e = e)$ is differentiable with respect to (x^l, e), $\sigma^2(d, x^l, e) > 0$.

(vi) The kernel function K is a symmetric probability density function, whose second order derivative K'' is Lipschitz continuous and $||K''(t)||_\infty := \sup_t |K''(t)| < \infty$.

Conditions $2(i)$—$2(v)$ are commonly used in standard local linear estimation (e.g., Li and Racine, 2007), condition $2(vi)$ holds generally for symmetric kernel functions.

Assumption 3. (Conditions for nonparametric estimation of $\tau(x^l)$ in Step 2)

(i) $h_3 \to 0$, $Nh_3 \to \infty$, $\tau(x^l)$ is twice differentiable.

(ii) $\mathbb{E}[\beta''_{x^l x^l}(X^l, e)|X^l = x^l]$ and $\mathbb{E}[\beta''_{ee}(X^l, e)|X^l = x^l]$ are differentiable.

(iii) $\text{var}(\beta(X^l, e)|X^l = x^l)$ and $\mathbb{E}\big[(D - e)^2\xi^2/e^2(1 - e)^2|X^l = x^l\big]$ are finite and differentiable.

(iv) $\sqrt{Nh_3}(h_1^2 + h_2^2) \to 0$, $\sqrt{Nh_3}h_3^2 \to 0$.

Conditions 3(i)—3(iii) are regular conditions. Condition 3(iv) guarantees that the bias term of $\hat{\beta}(x^l, e)$ and $\hat{\tau}(x^l, e)$ is ignorable when discussing the asymptotic properties of $\hat{\tau}(x^l, e)$.

3. Preliminaries

Let $\mu_j = \int t^j K(t)dt$ for $j = 0, 1, 2$. Denote β'_{x^l}, β'_e, $\beta''_{x^l x^l}$ and β''_{ee} $(j = 1, 2)$ to be the first and second derivatives of $\beta(x^l, e)$ with respect to x^l and e, respectively. For ease of exposition, we define $Z = (X^l, e)^\intercal$, $D^\dagger = (D, 1)^\intercal$. Let $h = (h_1, h_2)$, $\hat{K}_h(Z_i - z) = K_{h_1}(X_i^l - x^l)K_{h_2}(\hat{e}_i - e)$, $i = 1, \cdots, n$, $\hat{\boldsymbol{W}} = \text{diag}\{\hat{K}_h(Z_1 - z), \cdots, \hat{K}_h(Z_N - z)\}$, and $\hat{\boldsymbol{\Gamma}} = (\hat{\Gamma}_1, \cdots, \hat{\Gamma}_n)^\intercal$ with $\hat{\Gamma}_i = (D_i, 1, D_i(X_i^l - x^l)/h_1, (X_i^l -$

$x^l)/h_1, D_i(\hat{e}_i - e_0)/h_2, (\hat{e}_i - e_0)/h_2)^\intercal = (D_i^{\dagger\intercal}, D_i^{\dagger\intercal}(X_i^l - x^l)/h_1, D_i^{\dagger\intercal}(\hat{e}_i - e_0)/h_2)$. Correspondingly, $K_h(Z_i - z)$ and $\boldsymbol{W}$ are defined similarly as $\hat{K}_h(Z_i - z)$ and $\hat{\boldsymbol{W}}$, but substitute $\hat{e}_i$ with e_i. In addition, $\tilde{\beta}(x^l, e)$, $\hat{\beta}(x^l, e)$, $\tilde{\tau}(x^l, e)$, $\hat{\tau}(x^l, e)$ and $\beta_2(x^l, e)$ can be written as $\tilde{\beta}(z)$, $\hat{\beta}(z)$, $\tilde{\tau}(z)$, $\hat{\tau}(z)$ and $\beta_2(z)$. Similarly, we can define β'_{2,x^l}, $\beta'_{2,e}$, β''_{2,x^lx^l} and $\beta''_{2,ee}$ $(j = 1, 2)$ as the first and second derivatives of $\beta_2(z)$.

3.1 Decompositions of $\tilde{\beta}(x^l, e)$, $\hat{\beta}(x^l, e)$, $\tilde{\tau}(x^l, e)$ and $\hat{\tau}(x^l, e)$

$\tilde{\beta}(z)$ and $\hat{\beta}(z)$ can be decomposed as

$$\tilde{\beta}(z) - \beta(z) = (1, 0, 0, 0, 0, 0) \cdot S_N^{-1}(z)\{A_{1N}(z) + A_{2N}(z)\}, \tag{S.1}$$

$$\hat{\beta}(z) - \beta(z) = (1, 0, 0, 0, 0, 0) \cdot \hat{S}_N^{-1}(z)\{\hat{A}_{1N}(z) + \hat{A}_{2N}(z)\}, \tag{S.2}$$

where

$$\hat{S}_N(z) = N^{-1}\hat{\boldsymbol{\Gamma}}^\intercal\hat{\boldsymbol{W}}\hat{\boldsymbol{\Gamma}} = N^{-1}\sum_{i=1}^N \hat{\Gamma}_i\hat{K}_h(Z_i - z)\hat{\Gamma}_i^\intercal,$$

$$\hat{A}_{1N}(z) = N^{-1}\sum_{i=1}^N \hat{\Gamma}_i\hat{K}_h(Z_i - z) \cdot \{D_i\beta(Z_i) + \beta_2(Z_i) -$$

$$\hat{\Gamma}_i^\intercal(\beta(z), \beta_2(z), h_1\beta'_{x^1}(z), h_1\beta'_{2,x^l}(z), h_2\beta'_e(z), h_2\beta'_{2,e}(z))^\intercal\},$$

$$\hat{A}_{2N}(z) = N^{-1}\sum_{i=1}^N \hat{\Gamma}_i\hat{K}_h(Z_i - z)\xi_i,$$

where $S_N(z)$, $A_{1N}(z)$, $A_{2N}(z)$ are defined similarly as $\hat{S}_N(z)$, $\hat{A}_{1N}(z)$, $\hat{A}_{2N}(z)$, but replace $\hat{e}_i$ with e_i.

The local linear estimator of $\tau(x^l)$ can be written as

$$\tilde{\tau}(x^l) - \tau(x^l) = (1,0)T_N^{-1}(x^l)\{A_{3N}(x^l) + A_{4N}(x^l) + A_{5N}(x^l)\}, \tag{S.3}$$

$$\hat{\tau}(x^l) - \tau(x^l) = (1,0)T_N^{-1}(x^l)\{A_{3N}(x^l) + A_{4N}(x^l) + A_{5N}(x^l) + A_{6N}(x^l)\}, \tag{S.4}$$

where

$$T_N(x^l) = N^{-1}\sum_{i=1}^{N} G_i K_{h_3}(X_i^l - x^l)G_i^\intercal,$$

$$A_{3N}(x^l) = N^{-1}\sum_{i=1}^{N} G_i K_{h_3}(X_i^l - x^l)\{\tilde{\beta}(X_i^l, e_i) - \beta(X_i^l, e_i)\},$$

$$A_{4N}(x^l) = N^{-1}\sum_{i=1}^{N} G_i K_{h_3}(X_i^l - x^l)\{\beta(X_i^l, e_i) - \tau(X_i^l)\},$$

$$A_{5N}(x^l) = N^{-1}\sum_{i=1}^{N} G_i K_{h_3}(X_i^l - x^l)\{\tau(X_i^l) - \tau(x^l) - \tau'(x^l)(X_i^l - x^l)\},$$

$$A_{6N}(x^l) = N^{-1}\sum_{i=1}^{N} G_i K_{h_3}(X_i^l - x^l)\{\hat{\beta}(X_i^l, e_i) - \tilde{\beta}(X_i^l, e_i)\}.$$

3.2 Asymptotic properties of $\tilde{\beta}(x^l, e)$

The following Lemma 3.1 presents the asymptotic properties of $\tilde{\beta}(z)$, which are are parallel to asymptotic results of standard local linear estimator. Results in Lemma 3.1 will be used to deal with $A_{3N}(x^l)$ in (S.3) and (S.4).

Lemma 3.1. Under Assumption 2, we have

$$\mathbb{V}(z)^{-1/2}\{\tilde{\beta}(z) - \beta(z) - \text{bias}(\tilde{\beta}(z))\} \xrightarrow{d} N(0,1),$$

where $\text{bias}(\tilde{\beta}(z)) = \frac{1}{2}\mu_2\{\beta''_{x^l x^l}(z)h_1^2 + \beta''_{ee}(z)h_2^2\}$, and $\mathbb{V}(z) = (1,0)\{\nu^2/Nh_1h_2f(z)\}\boldsymbol{\Omega}^{-1}\boldsymbol{M}\boldsymbol{\Omega}^{-1}(1,0)^\intercal$,

$\boldsymbol{\Omega} = \mathbb{E}[D^\dagger D^{\dagger\intercal}|Z = z] = \begin{pmatrix} e & e \\ e & 1 \end{pmatrix}$, $\boldsymbol{M} = \mathbb{E}[D^\dagger D^{\dagger\intercal}\sigma^2(D, Z)|Z = z]$, $f(z)$ is the density function of Z.

Proof. According to the decomposition (S.1), we analyze $S_N(z)$, $A_{1N}(z)$ and $A_{2N}(z)$ one-by-one. We first deal with $S_N(z)$.

$$S_N(z) = \begin{pmatrix} S_{N,0}(z) & S_{N,1x^l}(z) & S_{N,1e}(z) \\ S_{N,1x^l}(z) & S_{N,2x^l}(z) & S_{N,2x^le}(z) \\ S_{N,1e}(z) & S_{N,2x^le}(z) & S_{N,2e}(z) \end{pmatrix},$$

where

$$S_{N,0}(z) = N^{-1}\sum_{i=1}^N D^\dagger D^{\dagger\intercal} K_h(Z_i - z),\ S_{N,2w}(z) = N^{-1}\sum_{i=1}^N D^\dagger D^{\dagger\intercal} K_h(Z_i - z)(\frac{X_i^l - x^l}{h_1})^2,$$
$$S_{N,2e}(z) = N^{-1}\sum_{i=1}^N D^\dagger D^{\dagger\intercal} K_h(Z_i - z)(\frac{e_i - e}{h_2})^2,\ S_{N,1e}(z) = N^{-1}\sum_{i=1}^N D^\dagger D^{\dagger\intercal} K_h(Z_i - z)(\frac{e_i - e}{h_2}),$$
$$S_{N,1x^l}(z) = N^{-1}\sum_{i=1}^N D^\dagger D^{\dagger\intercal} K_h(Z_i - z)(\frac{X_i^l - x^l}{h_1}),\ S_{N,2x^le}(z) = N^{-1}\sum_{i=1}^N D^\dagger D^{\dagger\intercal} K_h(Z_i - z)(\frac{X_i^l - x^l}{h_1})(\frac{e_i - e}{h_2}).$$

Under conditions 2(i) and 2(ii), and by observing that $\mathbb{E}(D_i|Z_i) = \mathbb{E}(D_i|e_i) = e_i$ and calculating the mean and the variance of $S_{N,0}(z)$, it yields that $\mathbb{E}[S_{N,0}(z)] = f(z)\boldsymbol{\Omega} + O(h_1 + h_2)$ and $\text{var}[S_{N,0}(z)] = O((Nh_1h_2)^{-1})$, Thus, $S_{N,0}(z) = f(z)\boldsymbol{\Omega} + O_p(h_1 + h_2 + 1/\sqrt{Nh_1h_2})$. Similarly we can derive the asymptotical order of other elements in $S_N(z)$, and show that

$$S_N(z) = f(z) \begin{pmatrix} \mathbf{\Omega} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mu_2 \mathbf{\Omega} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mu_2 \mathbf{\Omega} \end{pmatrix} + O_p(h_1 + h_2 + \frac{1}{\sqrt{N h_1 h_2}}). \tag{S.5}$$

Next we consider $A_{1N}(z)$ and $A_{2N}(z)$. By calculating the mean and the variance of $A_{1N}(z)$, it follows that under condition 2(iii),

$$(1,1,0,0,0,0)A_{1N}(z) = \frac{1}{2}\mu_2 f(z)\mathbf{\Omega} \begin{pmatrix} \beta''_{x^l x^l}(z)h_1^2 + \beta''_{ee}(z)h_2^2 \\ \beta''_{2,x^l x^l}(z)h_1^2 + \beta''_{2,ee}(z)h_2^2 \end{pmatrix} + O_p\left(h_1^3 + h_2^3 + \sqrt{\frac{h_1^4 + h_2^4}{N h_1 h_2}}\right). \tag{S.6}$$

Under conditions 2(iv) and 2(v), it follows by using Lyapunov central limit theorem that $(I_2, 0_{2\times 4}) \cdot \sqrt{N h_1 h_2} A_{2N}(z)$ is asymptotically normal distribution, it has zero mean and its variance is $\nu^2 f(z) \mathbb{E}[D_i^\dagger D_i^{\dagger\intercal} \sigma^2(D_i, Z_i) | Z_i = z] + o(1)$. That is,

$$\sqrt{N h_1 h_2}(I_2, 0_{2\times 4}) A_{2N}(z) \xrightarrow{d} N(0, \nu^2 f(z) \boldsymbol{M}), \tag{S.7}$$

where $\boldsymbol{M} = E[D^\dagger D^{\dagger\intercal} \sigma^2(D, Z) | Z = z]$. Under condition 2(i), Lemma 3.1 follows from (S.5)—(S.7).

□

3.3 Taylor expansion with an integral remainder

The Taylor expansion with an integral remainder is given by: if $f(x) \in C^m[a,b]$, $m \geq 1$, then

$$f(x_1) = f(x_2) + f'(x_2)(x_1 - x_2) + \cdots + \frac{f^{(m-1)}(x_2)}{(m-1)!}(x_1 - x_2)^{m-1} + R_{m,f}(x_1, x_2),$$

where

$$R_{m,f}(x_1, x_2) = \int_{x_2}^{x_1} \frac{f^{(m)}(t)}{(m-1)!}(x_1 - t)^{m-1} dt.$$

Denote $\hat{\alpha}$ as the $\sqrt{N}$-consistent estimator of α, $e(\alpha) = g(X^\intercal \alpha)$. We give statements of Taylor expansion of kernel function and propensity score function here, which will be used in the proof of Theorem 1. Under condition 2(vi), for a given value of propensity score e_0,

$$\begin{aligned} &K(\frac{e(\alpha) - e_0}{h_2}) - K(\frac{e(\alpha) - e_0}{h_2}) \\ &= K'(\frac{e(\alpha) - e_0}{h_2})\frac{e(\alpha) - e(\alpha)}{h_2} + \frac{1}{2}K''(\frac{e(\alpha) - e_0}{h_2})\Big(\frac{e(\alpha) - e(\alpha)}{h_2}\Big)^2 + R_e, \end{aligned} \quad \text{(S.8)}$$

where

$$R_e = \int_{(e(\alpha)-e_0)/h_2}^{(e(\alpha)-e_0)/h_2} \Big(K''(t) - K''(\frac{e(\alpha) - e_0}{h_2})\Big)\Big(\frac{e(\alpha) - e_0}{h_2} - t\Big) dt$$

By the Lipschitz continuity of K'',

$$\sup_e |R_e| \leq M_1 h_2^{-3} |e(\alpha) - e(\alpha)|^3.$$

M_1 is a finite constant. Likewise, under condition 1(ii), we have

$$e(\alpha) - e(\alpha) = g'(X^\intercal \alpha) X^\intercal (\alpha - \alpha) + R_\alpha, \quad \text{(S.9)}$$

where

$$\sup_{\alpha} |R_\alpha| \leq M_2\{X^\intercal(\alpha - \alpha)\}^2,$$

M_2 is a finite constant.

4. Proof of Theorem 1

For functions $\varphi_N(z)$, $\psi_N(z)$, $\varphi_N(z) = u_p(\psi_N(z))$ stands for $\varphi_N(z)/\psi_N(z) = o_p(1)$ as $N \to \infty$ uniformly for $z \in \mathcal{X}^l \times (0,1)$, and $\varphi_N(z) = U_p(\psi_N(z))$ means $\varphi_N(z)/\psi_N(z) = O_p(1)$ as $N \to \infty$ uniformly for $z \in \mathcal{X}^l \times (0,1)$.

Lemma 4.1. Under Assumptions 1—2, the following equations are valid.

$$\sup_{z\in\mathcal{X}^l\times(0,1)} \left|\hat{S}_N(z) - S_N(z)\right| = O_p(N^{-1/2}), \tag{S.10}$$

$$(1,0,0,0,0,0) \sup_{z\in\mathcal{X}^l\times(0,1)} \left|\hat{A}_{1N}(z) - A_{1N}(z)\right| = O_p(N^{-1/2}), \tag{S.11}$$

$$(1,0,0,0,0,0) \sup_{z\in\mathcal{X}^l\times(0,1)} \left|\hat{A}_{2N}(z) - A_{2N}(z)\right| = O_p(N^{-1/2}). \tag{S.12}$$

Lemma 4.1 is consistent with the results in Gu and Yang (2015). Based on Lemma 4.1, we can finish the proof of Theorem 1.

Proof of Theorem 1. According to (S.10), (S.11) and (S.12), we have

$$
\begin{aligned}
&\hat{\beta}(z) - \beta(z) \\
&= (1,0,0,0,0,0)\hat{S}_N^{-1}(z)(\hat{A}_{1N}(z) + \hat{A}_{2N}(z)) \\
&= (1,0,0,0,0,0)\big\{S_N(z) + U_p(N^{-1/2}h_2^{-1})\big\}^{-1}(\hat{A}_{1N}(z) + \hat{A}_{2N}(z)) \\
&= (1,0,0,0,0,0)S_N^{-1}(z)(A_{1N}(z) + A_{2N}(z)) + U_p(N^{-1/2})\big\{(1,0,0,0,0,0)(A_{1N}(z) + A_{2N}(z))\big\} + U_p(N^{-1/2}) \\
&= \tilde{\beta}(z) - \beta(z) + U_p\big(N^{-1/2}(h_1^2 + h_2^2) + N^{-1/2}u_p(1) + N^{-1}h_2^{-1} + N^{-1/2}\big) \\
&= \tilde{\beta}(z) - \beta(z) + U_p(N^{-1/2}).
\end{aligned}
$$

This completes the proof of Theorem 1. □

We present the proof of Lemma 4.1 below.

Proof of Lemma 4.1. . Let $e(\alpha) = g(X^\intercal\alpha)$, $e(\alpha) = g(X^\intercal\alpha)$, $e_i = g(X_i^\intercal\alpha)$, $\hat{e}_i = g(X_i^\intercal\hat{\alpha})$, and e the value of propensity scores at a given point. We split the proof into three steps.

Step 1. We first show (S.10). Note that

$$
\hat{S}_N(z) = \begin{pmatrix} \hat{S}_{N,0}(z) & \hat{S}_{N,1x^l}(z) & \hat{S}_{N,1e}(z) \\ \hat{S}_{N,1x^l}(z) & \hat{S}_{N,2x^l}(z) & \hat{S}_{N,2x^le}(z) \\ \hat{S}_{N,1e}(z) & \hat{S}_{N,2x^le}(z) & \hat{S}_{N,2e}(z) \end{pmatrix}.
$$

We consider only $\hat{S}_{N,0}(z)$ and other elements in $\hat{S}_N(z)$ can be treated similarly.

$$\hat{S}_{N,0}(z) - S_{N,0}(z) = N^{-1}\sum_{i=1}^{N} D^{\dagger}D^{\dagger\intercal}K_{h_1}(X_i^l - x^l)\{K_{h_2}(\hat{e}_i - e) - K_{h_2}(e_i - e)\}.$$

Let $g_\alpha(D,X) := D^{\dagger}D^{\dagger\intercal}K_{h_1}(X^l - x^l)K_{h_2}(e(\alpha) - e)$. According to Theorem 2.11.22 of van der Vaart and Wellner (1996),

$$\{g_a : \alpha \in B_{\epsilon_N}(\alpha), z \in \mathcal{X}^l \times (0,1), \epsilon_N = CN^{-1/2}\} \text{ is } P\text{-Donsker class},$$

where C is a finite constant and $B_{\epsilon_N}(\alpha) = \{\alpha : ||\alpha - \alpha|| \le \epsilon_N\}$, $||\cdot||$ denotes L_2-norm. By Lemma 19.24 of van der Vaart (1998),

$$\sup_{\alpha \in B_{\epsilon_n}(\alpha)} |\mathbb{G}_N(g_\alpha - g_\alpha)| = u_p(1),$$

where $\mathbb{G}_N g_\alpha := N^{-1/2}\sum_{i=1}^{N}\{g_\alpha(D_i, X_i) - \mathbb{E}(g_\alpha)\}$. Since

$$\begin{aligned}\sup_{\alpha \in B_{\epsilon_N}(\alpha)} |N^{-1}\sum_{i=1}^{N}(g_\alpha - g_{\alpha_0})| &\le N^{-1/2}\cdot \sup_{\alpha \in B_{\epsilon_n}(\alpha)} |\mathbb{G}_N(g_\alpha - g_\alpha)| + \sup_{\alpha \in B_{\epsilon_N}(\alpha)} |\mathbb{E}(g_\alpha - g_{\alpha_0})| \\ &= u_p(N^{-1/2}) + \sup_{\alpha \in B_{\epsilon_N}(\alpha)} |\mathbb{E}(g_\alpha - g_{\alpha_0})|.\end{aligned}$$

It suffices to show that $\sup_{\alpha \in B_{\epsilon_N}(\alpha)} |\mathbb{E}(g_\alpha - g_{\alpha_0})| = U_p(N^{-1/2})$. Indeed, for any $\alpha \in B_{\epsilon_N}(\alpha)$, by Taylor expansion (S.8),

$$\begin{aligned}\mathbb{E}(g_\alpha - g_\alpha) &= \mathbb{E}[D^{\dagger}D^{\dagger\intercal}K_{h_1}(X^l - x^l)\{K_{h_2}(e(\alpha) - e) - K_{h_2}(e(\alpha) - e)\}] \\ &\le \mathbb{E}[D^{\dagger}D^{\dagger\intercal}K_{h_1}(X^l - x^l)h_2^{-2}K'(\frac{e(\alpha) - e}{h_2})(e(\alpha) - e(\alpha))] \\ &\quad + \mathbb{E}[D^{\dagger}D^{\dagger\intercal}K_{h_1}(X^l - x^l)K''(\frac{e(\alpha) - e}{h_2})h_2^{-3}(e(\alpha) - e(\alpha)^2] \\ &\quad + M_1 \cdot \mathbb{E}[D^{\dagger}D^{\dagger\intercal}K_{h_1}(X^l - x^l)h_2^{-4}|e(\alpha) - e(\alpha)|^3] \\ &:= S_{n,01}(z) + S_{n,02}(z) + S_{n,03}(z).\end{aligned}$$

and by Taylor expansion (S.9) and condition 2(i),

$$S_{n,01}(z) \leq \mathbb{E}[D^{\dagger}D^{\dagger\intercal}K_{h_1}(X^l - x^l)h_2^{-2}K'(\frac{e(\alpha)-e}{h_2})g'(X^{\intercal}\alpha)X^{\intercal}](\alpha - \alpha)$$

$$+M_2 \cdot \mathbb{E}[D^{\dagger}D^{\dagger\intercal}K_{h_1}(X^l - x^l)h_2^{-2}(\alpha-\alpha)^{\intercal}XX^{\intercal}(\alpha-\alpha)]$$

$$= U_p(||\alpha-\alpha||) + U_p(h_2^{-2}||\alpha-\alpha||^2)$$

$$= U_p(N^{-1/2})$$

The last second equality holds because $\mathbb{E}[D^{\dagger}D^{\dagger\intercal}K_{h_1}(X^l - x^l)h_2^{-2}K'(\frac{e(\alpha)-e}{h_2})g'(X^{\intercal}\alpha)X^{\intercal}] = O(1)$ by the symmetry of $K(t)$, namely, $\int_t K'(t)dt = 0$. Likewise, $S_{n,02}(z) = U_p(h_2^{-2}||\alpha-\alpha||^2) = u_p(N^{-1/2})$, $S_{n,03}(z) = U_p(h_2^{-4}||\alpha-\alpha||^{3/2}) = u_p(N^{-1/2})$. Thus, (S.10) holds.

Step 2. Next we prove (S.11). Define $\hat{R}_i := D_i\beta(Z_i) + \beta_2(Z_i) - D_i[\beta(z) + \beta'_{x^1}(z)(X_i^l - x^l) + \beta'_e(z)(\hat{e}_i - e)] - [\beta_2(z) + \beta'_{2,x^l}(z)(X_i^l - x^l) + \beta'_{2,e}(z)(\hat{e}_i - e)]$. Similarly, we define R_i by replacing $\hat{e}_i$ in $\hat{R}_i$ with e_i.

$$(1,0,0,0,0,0)\{\hat{A}_{1N}(z) - A_{1N}(z)\} = A_{1N1}(z) + A_{1N2}(z) + A_{1N3}(z), \qquad \text{(S.13)}$$

where

$$A_{1N1}(z) := N^{-1}\sum_{i=1}^{N} K_{h_1}(X_i^l - x^l)[K_{h_2}(\hat{e}_i - e) - K_{h_2}(e_i - e)]R_i,$$

$$A_{1N2}(z) := N^{-1}\sum_{i=1}^{N} K_{h_1}(X_i^l - x^l)K_{h_2}(e_i - e)(\hat{R}_i - R_i),$$

$$A_{1N3}(z) := N^{-1}\sum_{i=1}^{N} K_{h_1}(X_i^l - x^l)[K_{h_2}(\hat{e}_i - e) - K_{h_2}(e_i - e)](\hat{R}_i - R_i).$$

For $A_{1N1}(z)$, we claim that

$$\sup_{z\in\mathcal{X}^l\times(0,1)} |A_{1N1}(z)| = o_p(N^{-1/2}). \tag{S.14}$$

By conditions 3(ii) and 3(iii), $\max_i |R_i|$ is uniformly bounded. Then according to Theorem 2.10.6 of van der Vaart and Wellner (1996),

$\{h_\alpha := K_{h_1}(X^l - x^l)K_{h_2}(e(\alpha) - e)R, \alpha \in B_{\epsilon_N}(\alpha), z \in \mathcal{X}^l \times (0,1)\}$ is P-Donsker class,

Thus,

$$|A_{1N1}(z)| \leq u_p(N^{-1/2}) + \sup_{\alpha\in B_{\epsilon_N}(\alpha)} |\mathbb{E}(h_\alpha - h_\alpha)|.$$

We focus on discussing $\mathbb{E}(h_\alpha - h_\alpha)$. Again, by (S.8),

$$\begin{aligned}
\mathbb{E}(h_\alpha - h_\alpha) \leq\ & \mathbb{E}[K_{h_1}(X^l - x^l)h_2^{-2}K'(\frac{e(\alpha) - e}{h_2})(e(\alpha) - e(\alpha))R] \\
& +\mathbb{E}[K_{h_1}(X^l - x^l)h_2^{-3}K''(\frac{e(\alpha) - e}{h_2})(e(\alpha) - e(\alpha))^2 R] \\
& +M_1 \cdot \mathbb{E}[K_{h_1}(X^l - x^l)h_2^{-4}|e(\alpha) - e(\alpha)|^3|R|] \\
:=\ & A_{1n1,1}(z) + A_{1n1,2}(z) + A_{1n1,3}(z).
\end{aligned}$$

By a similar argument of $S_{n,01}(z)$, we obtain that

$$A_{1n1,1}(z) = u_p(N^{-1/2}), \quad A_{1n1,2}(z) = u_p(N^{-1/2}),$$

Note that R is uniformly bounded, $A_{1n1,3}(z) = U_p(h_2^{-4}N^{-3/2}) = u_p(N^{-1/2})$. Thus, (S.14) holds.

Next, we prove

$$\sup_{z\in\mathcal{X}^l\times(0,1)}|A_{1N2}(z)| = O_p(N^{-1/2}). \tag{S.15}$$

We decompose $A_{1N2}(z)$ into two parts.

$$A_{1N2}(z) = N^{-1}\sum_{i=1}^{N}\mathbb{E}[K_{h_1}(X_i^l - x^l)K_{h_2}(e_i - e)](\hat{R}_i - R_i) +$$
$$N^{-1}\sum_{i=1}^{N}\Big\{K_{h_1}(X_i^l - x^l)K_{h_2}(e_i - e) - \mathbb{E}[K_{h_1}(X_i^l - x^l)K_{h_2}(e_i - e)]\Big\}(\hat{R}_i - R_i)$$
$$:= A_{1N2,1}(z) + A_{1N2,2}(z).$$

Since $\sup_{z\in\mathcal{X}^l\times(0,1)}|\mathbb{E}[K_{h_1}(X_i^l - x^l)K_{h_2}(e_i - e)]| = O(1)$,

$$\sup_{z\in\mathcal{X}^l\times(0,1)}|A_{1N2,1}(z)| = O(1)\cdot\sup_{z}\Big|N^{-1}\sum_{i=1}^{N}(\hat{R}_i - R_i)\Big| = O_p(N^{-1/2}).$$

By Cauchy-Schwarz inequality,

$$\sup_{z\in\mathcal{X}^l\times(0,1)}|A_{1N2,2}(z)| \le \sup_{z\in\mathcal{X}^l\times(0,1)}\Big|N^{-1}\sum_{i=1}^{N}(\hat{R}_i - R_i)^2\Big|\cdot$$
$$\sup_{z\in\mathcal{X}^l\times(0,1)}\Big|N^{-1}\sum_{i=1}^{N}\big\{K_{h_1}(X_i^l - x^l)K_{h_2}(e_i - e) - \mathbb{E}[X_iK_{h_1}(X_i^l - x^l)K_{h_2}(e_i - e)]\big\}^2\Big|$$
$$= o_p(N^{-1/2}).$$

Thus, (S.15) holds.

We consider $A_{1N3}(z)$. By the mean value theorem,

$$|K_{h_2}(\hat{e}_i - e) - K_{h_2}(e_i - e)| \le M_3h_2^{-2}|\hat{e}_i - e_i|, \quad |\hat{R}_i - R_i| \le M_4|\hat{e}_i - e_i|,$$

where M_3 and M_4 are finite constants. It follows that

$$|A_{1N3}(z)| \leq M_3 M_4 N^{-1} \sum_{i=1}^{N} K_{h_1}(X_i^l - x^l) h_2^{-2} (\hat{e}_i - e_i)^2 = U_p(N^{-1} h_2^{-2}) = u_p(N^{-1/2}). \tag{S.16}$$

Thus, (S.11) follows from (S.13)—(S.16).

Step 3. Finally, we show (S.12). By Taylor expansion (S.8),

$$(1,0,0,0,0,0)\{\hat{A}_{2N}(z) - A_{2N}(z)\} = N^{-1} \sum_{i=1}^{N} K_{h_1}(X_i^l - x^l)[K_{h_2}(\hat{e}_i - e) - K_{h_2}(e_i - e)]\xi_i$$

$$= N^{-1} \sum_{i=1}^{N} K_{h_1}(X_i^l - x^l) h_2^{-2} K'(\frac{e_i - e}{h_2}) \xi_i (\hat{e}_i - e_i) +$$

$$N^{-1} \sum_{i=1}^{N} K_{h_1}(X_i^l - x^l) h_2^{-3} K''(\frac{e_i - e}{h_2}) \xi_i (\hat{e}_i - e_i)^2 + N^{-1} \sum_{i=1}^{N} K_{h_1}(X_i^l - x^l) h_2^{-1} R_{i,e} \xi_i$$

$$:= A_{2N1}(z) + A_{2N2}(z) + A_{2N3}(z).$$

where $|R_{i,e}| \leq M_1 h_2^{-3} |\hat{e}_i - e_i|^3$. Following a similar argument to $S_{n,01}(z)$, it yields that

$$A_{2N1}(z) = U_p(N^{-1/2}), \quad A_{2N2}(z) = u_p(N^{-1/2}), \quad A_{2N2}(z) = u_p(N^{-1/2}).$$

Thus, (S.12) is valid.

□

5. Proof of Theorem 2

Proof of Theorem 2. Under condition 3(i), following a similar argument to the proof of Lemma 3.1, we have

$$T_N(x^l) = f(x^l)\begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix} + O_p(h_3 + \frac{1}{\sqrt{Nh_3}}), \tag{S.17}$$

and

$$(1,0)A_{5N}(x^l) = \frac{1}{2}\mu_2 f(x^l)\tau''(x^l)h_3^2 + O_p(h_3^3 + h_3^2/\sqrt{Nh_3}). \tag{S.18}$$

where $f(x^l)$ is the density function of X^l. Theorem 1 implies that

$$(1,0)A_{6N}(x^l) = U_p(N^{-1/2}). \tag{S.19}$$

Next, we focus on analyzing $A_{3N}((x^l))$. According to the proof of Lemma 3.1, $\tilde{\beta}(X_i^l, e_i) - \beta(X_i^l, e_i)$ can be represented as $(1, 0_{1\times 5})S_N^{-1}(z)\{A_{1N}(z) + A_{2N}(z)\}$, which equals

$$\frac{1}{2}\mu_2(\beta''_{x^l x^l}(z_i)h_1^2 + \beta''_{ee}(z_i)h_2^2) + (1,0)\frac{1}{f(z_i)}\begin{pmatrix} e_i & e_i \\ e_i & 1 \end{pmatrix}^{-1} \frac{1}{N}\sum_{j=1}^{N}(D_j, 1)^\intercal K_h(Z_j - z_i)\xi_j +$$
$$U_p\Big((h_1 + h_2)^3\Big)$$
$$= \frac{1}{2}\mu_2(\beta''_{x^l x^l}(z_i)h_1^2 + \beta''_{ee}(z_i)h_2^2) + \frac{1}{N}\sum_{j=1}^{N}\frac{(D_j - e_i)}{f(z_i)e_i(1-e_i)}K_h(Z_j - z_i)\xi_j + U_p((h_1 + h_2)^3).$$

Therefore, $(1,0)A_{3N}(x^l) = A_{3N1}(x^l) + A_{3N2}(x^l) + U_p((h_1+h_2)^3)$, where

$$\begin{aligned}
A_{3N1}(x^l) &= \frac{1}{2}\mu_2 \cdot \frac{1}{N}\sum_{i=1}^{N} K_{h_3}(X_i^l - x^l)(\beta''_{x^l x^l}(Z_i)h_1^2 + \beta''_{ee}(Z_i)h_2^2), \\
A_{3N2}(x^l) &= \frac{1}{N}\sum_{i=1}^{N} K_{h_3}(X_i^l - x^l) \cdot \frac{1}{N}\sum_{j=1}^{N} \frac{(D_j - e_i)}{f(Z_i)e_i(1-e_i)} K_h(Z_j - Z_i)\xi_j \\
&= \frac{1}{N}\sum_{j=1}^{N} \xi_j \Big\{ \frac{1}{N}\sum_{i=1}^{N} K_{h_3}(X_i^l - x^l)\frac{(D_j - e_i)}{f(Z_i)e_i(1-e_i)} K_h(Z_i - Z_j)\Big\}.
\end{aligned} \tag{S.20}$$

Under Conditions 3(ii) and 3(iii), using the asymptotic mean and variance of $A_{3N1}(x^l)$, we can show that

$$\begin{aligned}
A_{3N1}(x^l) &= \frac{1}{2}\mu_2 f(x^l)[E\{\beta''_{x^l x^l}(Z_i)|X_i^l = x^l\}h_1^2 + E\{\beta''_{ee}(Z_i)|X_i^l = x^l\}h_2^2] + U_p((h_1^2+h_2^2)(h_3^2 + \frac{1}{\sqrt{Nh_3}})) \\
&= \frac{1}{2}\mu_2 f(x^l)[E\{\beta''_{x^l x^l}(Z_i)|X_i^l = x^l\}h_1^2 + E\{\beta''_{ee}(Z_i)|X_i^l = x^l\}h_2^2] + u_p(1/\sqrt{Nh_3}).
\end{aligned} \tag{S.21}$$

It is easy to show that

$$\{K_{h_3}(X^l - x^l)\frac{(D_j - e)}{f(Z)e(1-e)} K_h(Z - Z_j), Z \in \mathcal{X}^l \times (0,1)\} \quad \text{is } P\text{-GC class.} \tag{S.22}$$

Further,

$$\begin{aligned}
&\mathbb{E}_{X^l, e(X)}[K_{h_3}(X^l - x^l) \cdot \frac{(D_j - e(X))}{f(Z)e(X)(1-e(X))} K_h(Z - Z_j)] \\
&= \int_{t_1}\int_{t_2} \frac{1}{h_3} K(\frac{X_j^l + h_1 t_1 - x^l}{h_3}) \cdot \frac{D_j - e_j - h_2 t_2}{(e_j + h_2 t_2)(1 - e_j - h_2 t_2)} K(t_1)K(t_2)dt_1 dt_2 \\
&= \int_{t_2} \frac{D_j - e_j - h_2 t_2}{(e_j + h_2 t_2)(1 - e_j - h_2 t_2)} K(t_2)dt_2 \cdot \int_{t_1} K(t_1)\frac{1}{h_3}K(\frac{X_j^l + h_1 t_1 - x^l}{h_3})dt_1 \\
&= \Big\{\frac{D_j - e_j}{e_j(1-e_j)} + u_p(1)\Big\} \cdot \bar{K}_{h_3}(X_j^l - x^l),
\end{aligned} \tag{S.23}$$

where $\bar{K}_{h_3}(\cdot) := \bar{K}(\cdot/h_3)/h_3 := \int_t K(t)K_{h_3}(\cdot + h_1 t)dt$. According to (S.22) and (S.23),

$$A_{3N2}(x^l) = \left\{\frac{1}{N}\sum_{i=1}^{N}\bar{K}_{h_3}(X_i^l - x^l)\frac{D_i - e_i}{e_i(1-e_i)}\xi_i\right\}(1 + o_p(1)).$$

Combining $A_{3N2(x^l)}$ and $A_{4N}(x^l)$, we have $\sqrt{Nh_3}\{U_{3n2}+(1,0)A_{4N}\} = \tilde{A}_{4N}(x^l)(1+o_p(1))$, with

$$\tilde{A}_{4N}(x^l) = \sqrt{Nh_3}\cdot\frac{1}{N}\sum_{i=1}^{N}\left\{K_{h_3}(X_i^l - x^l)(\beta(Z_i) - \tau(X_i^l)) + \bar{K}_{h_3}(X_i^l - x^l)\frac{D_i - e_i}{e_i(1-e_i)}\xi_i\right\},$$

Note that the above term has zero mean due to

$$\text{cov}(K_{h_3}(X_i^l - x^l)(\beta(Z_i) - \tau(X_i^l)), \bar{K}_{h_3}(X_i^l - x^l)\frac{D_i - e_i}{e_i(1-e_i)}\xi_i) = 0.$$

Its variance is

$$h_3\mathbb{E}\left\{K_{h_3}^2(X_i^l - x^l)(\beta(Z_i) - \tau(X_i^l))^2 + \bar{K}_{h_3}^2(X_i^l - x^l)\frac{(D_i - e_i)^2}{e_i^2(1-e_i)^2}\xi_i^2\right\}$$
$$= \nu f(x^l)\text{var}[\beta(Z_i)|X_i^l = x^l] + f(x^l)\int \bar{K}^2(t)dt \cdot E\left[(\frac{(D_i - e_i)^2}{e_i^2(1-e_i)^2}\xi_i^2|X_i^l = x^l\right] + o(1).$$

This completes the proof of Theorem 2.

□

6. Generalized Version of Theorem 2

Suppose that X^l is q-dimensional, the bandwidths $h_1 = (h_{11}, ..., h_{1q})$, $h_3 = (h_{31}, ..., h_{3q})$. Let $X_{i,j}^l$ be the j-th component of X_i^l for subject i, x_j^l be the j-th element of x^l, and define $K_{h_1}(X_i^l - x^l) = \prod_{j=1}^q K_{h_{1j}}(X_{i,j}^l - x_j^l)$, $K_{h_3}(X_i^l - x^l) =$

$\prod_{j=1}^{q} K_{h_{3j}}(X_{i,j}^l - x_j^l)$, $(X_i^l - x^l)/h_1 = ((X_{i,1}^l - x_1^l)/h_{11}, \cdots, (X_{i,q}^l - x_q^l)/h_{1q})^\intercal$. Then the local linear estimators of $\beta(X^l, e)$ and $\tau(x^l)$ are given as

$$\hat{\beta}(x^l, \hat{e}) = (1, 0_{1\times(2q+3)})(\hat{\boldsymbol{\Gamma}}^\intercal \hat{\boldsymbol{W}} \hat{\boldsymbol{\Gamma}})^{-1} \hat{\boldsymbol{\Gamma}}^\intercal \hat{\boldsymbol{W}} \boldsymbol{Y}, \tag{S.24}$$

$$\hat{\tau}(x^l) = (1, 0_{1\times q})(\boldsymbol{G}^\intercal \boldsymbol{\Lambda} \boldsymbol{G})^{-1} \boldsymbol{G}^\intercal \boldsymbol{\Lambda} \hat{\boldsymbol{\beta}}, \tag{S.25}$$

where $0_{1\times q}$ is a q-dimensional row vector with each element being zero, $\boldsymbol{Y} = (Y_1, \cdots, Y_N)^\intercal$, $\boldsymbol{\Gamma} = (\Gamma_1, \cdots, \Gamma_N)^\intercal$ with $\Gamma_i = (D_i, 1, D_i(X_i^l - x^l)^\intercal/h_1, (X_i^l - x^l)^\intercal/h_1, D_i(e_i - e)/h_2, (e_i - e)/h_2)^\intercal$, $i = 1, \ldots, N$, and $\boldsymbol{W} = \text{diag}\{K_{h1}(X_1^l - x^l)K_{h2}(e_1 - e), \ldots, K_{h1}(X_N^l - x^l)K_{h2}(e_N - e)\}$,

Further, $\tilde{\boldsymbol{\beta}} = (\tilde{\beta}(X_1^l, e_1), \ldots, \tilde{\beta}(X_N^l, e_N))^\intercal$, $\boldsymbol{G} = (G_1, \ldots, G_n)^\intercal$ with $G_i = (1, (X_i^l - x^l)^\intercal/h_3)^\intercal$, and $\boldsymbol{\Lambda} = \text{diag}\{K_{h_3}(X_1^l - x^l), \ldots, K_{h_3}(X_N^l - x^l)\}$.

Assumption 2'. (Conditions for nonparametric estimation of $\beta(x^l, e)$ in Step 1)

(i) The bandwidths $h_{1j} \to 0$ for $j = 1, ..., q$, $h_2 \to 0$, $N(\prod_{j=1}^{q} h_{1j})h_2 \to \infty$, and $\sqrt{N(\prod_{j=1}^{q} h_{1j})h_2}(\sum_{j=1}^{q} h_{1j} + h_2)^3 \to 0$, and $Nh_2^4 \to \infty$ as $N \to \infty$.

(ii) The same as assumptions 2(ii)—2(vi).

Assumption 3'. (Conditions for nonparametric estimation of $\tau(x^l)$ in Step 2)

(i) $h_{3j} \to 0$ for $j = 1, ..., q$, $N(\prod_{j=1}^{q} h_{3j}) \to \infty$, $\tau(x^l)$ is twice differentiable.

(ii) $\mathbb{E}[\beta''_{ee}(X^l, e)|X^l = x^l]$ and $\mathbb{E}[\beta''_{x_j^l x_j^l}(X^l, e)|X^l = x^l]$ for $j = 1, ..., q$ are differentiable.

(iii) The same as assumption 3(iii).

(iv) $\sqrt{N(\prod_{j=1}^{q} h_{3j})}(\sum_{j=1}^{q} h_{1j}+h_2)^3 \to 0$, $\sqrt{N(\prod_{j=1}^{q} h_{3j})}(\sum_{j=1}^{q} h_{1j}^2+h_2^2)(\sum_{j=1}^{q} h_{3j}^2) \to 0$.

Let $\tau''_{x_j^l x_j^l}$ be the second derivative of $\tau(x^l)$ with respect to x_j^l. The following Theorem 6.1 is a generalization of Theorem 2 in the manuscript.

Theorem 6.1. Under Assumptions 1, 2' and 3', $\hat{\tau}(x^l)$ is a consistent estimator of $\tau(x^l)$, and

$$\mathbb{V}\left(x^l\right)^{-1/2}\left\{\hat{\tau}(x^l)-\tau(x^l)-bias(\hat{\tau}(x^l))\right\} \xrightarrow{d} N(0,1),$$

where

$$bias(\hat{\tau}(x^l)) = \frac{1}{2}\mu_2\Big\{\sum_{j=1}^{q}\Big(\tau''(x_j^l)h_{3j}^2+\mathbb{E}(\beta''_{x_j^l x_j^l}(Z)|X^l=x^l)h_{1j}^2\Big)+E(\beta''_{ee}(Z)|X^l=x^l)h_2^2\Big\},$$

$$\mathbb{V}\left(x^l\right) = \frac{1}{n(\prod_{j=1}^{q} h_{3j})f(x^l)}\left\{\nu^q\cdot\text{var}(\beta(Z)|X^l=x^l)+\int\bar{K}^2(t)dt\cdot E\bigg(\frac{(D-e)^2}{e^2(1-e)^2}\xi^2|X^l=x^l\bigg)\right\},$$

where $\bar{K}(x) = \int_{t_1}\cdots\int_{t_q}\left(\prod_{j=1}^{q} K(t_j)K(x+h_{1j}t_j/h_{3j})\right)dt_1\cdots dt_q$, $f(x^l)$ is the density function of X^l.

Proof. The proof is identical to that of Theorem 2. □

7. Extension to Discrete X^l

For discrete X^l, either nominal or ordinal, the common way to estimate $\tau(x^l)$ is through stratification, splitting the samples into different cells based on the values of X^l and estimate $\tau(x^l)$ within each cell (Abrevaya et al., 2015; Lee et al., 2017). As pointed by Li and Racine (2010), the sample-splitting approach may degrade efficiency for finite samples. Another approach is the kernel smoothing method, first proposed by Aitchison and Aitken (1976).

7.1 Estimation

Concretely, if X^l is an unordered discrete variable, Li and Racine (2007) suggested using a variant of the Aitchison and Aitken (1976) kernel function defined as,

$$L_\lambda(X_i^l, x^l) = \begin{cases} 1, & \text{when } X_i^l = x^l, \\ \lambda, & \text{otherwise,} \end{cases}$$

where λ is a tuning parameter, corresponding to the bandwidth of continuous X^l. For an ordered discrete covariate, the kernel function is

$$L_\lambda(X_i^l, x^l) = \begin{cases} 1, & \text{when } X_i^l = x^l, \\ \lambda^{|X_i^l - x^l|}, & \text{otherwise.} \end{cases}$$

Note that for both unordered and ordered cases, $\lambda = 0$ leads to an indicator function whereas $\lambda = 1$ leads to a uniform weight function. Therefore, the range of λ is $[0, 1]$. Observe that the kernel weight function we use here does not add up to one when $\lambda \neq 0$. However, this does not affect the estimate of $\hat{\beta}(x^l, e)$, as the kernel function appears in both the numerator and the denominator.

The kernel function for the mixed type of data (x^l, e) is simply the product of $L_{\lambda_1}(\cdot)$ and $K_{h_2}(\cdot)$, i.e.,

$$L_{\lambda_1}(X_i^l, x^l) \cdot K_{h_2}(\hat{e}_i - e).$$

$\hat{\boldsymbol{\Gamma}} = (\hat{\Gamma}_1, \cdots, \hat{\Gamma}_N)^\intercal$ with $\hat{\Gamma}_i = (D_i, 1, D_i(\hat{e}_i - e)/h_2, (\hat{e}_i - e)/h_2)^\intercal$, $\hat{\boldsymbol{W}} = \text{diag}(L_{\lambda_1}(X_1^l, x^l)K_{h_2}(\hat{e}_1 - e), \cdots, L_{\lambda_1}(X_N^l, x^l)K_{h_2}(\hat{e}_N - e))$. Then the estimators of $\beta(x^l, e)$ and $\tau(x^l)$ are given as

$$\hat{\beta}(x^l, e) = (1, 0, 0, 0)(\boldsymbol{\Gamma}^\intercal \boldsymbol{W} \boldsymbol{\Gamma})^{-1} \boldsymbol{\Gamma}^\intercal \boldsymbol{W} \boldsymbol{Y}. \tag{S.26}$$

$$\hat{\tau}(x^l) = \frac{N^{-1} \sum_{i=1}^N \hat{\beta}(X_i^l, e_i) L_{\lambda_2}(X_i^l, x^l)}{N^{-1} \sum_{i=1}^N L_{\lambda_2}(X_i^l, x^l)}. \tag{S.27}$$

7.2 Asymptotic properties

Assumption 1'. The propensity score model can be written as $e(X) = g(X^\intercal \alpha)$. α is the true unknown parameter, $g(\cdot)$ is a known function.

(i) The estimates of α, $\hat{\alpha}$, has the asymptotic linear representation,

$$\sqrt{N}(\hat{\alpha} - \alpha) = N^{-1/2} \sum_{i=1}^N \psi(X_i) + o_p(1),$$

with $\mathbb{E}[\psi(X_i)] = 0, \mathbb{E}[\psi(X_i)\psi^\intercal(X_i)] < \infty$. That is, $\psi(X_i)$ is the influence function of $\hat{\alpha}$.

(ii) The second-order derivative of g is uniformly bounded, i.e., $\sup_t |g''(t)|$ is bounded.

Assumption 1' usually holds for generalized linear models.

Assumption 4.

(i) $\lambda_1 \to 0$, $h_2 \to 0$, $Nh_2 \to \infty$, and $\sqrt{Nh_2}(\lambda_1 + h_2^4) \to 0$ as $N \to \infty$.

(ii) Denote $f(x^l, e)$ to be the joint probability density for (X^l, e). Assume that $f(x^l, e)$ has continuous partial derivative of order one with respect to x^l and e, and $f(x^l, e) > \delta > 0$ for some positive constant δ.

(iii) $\beta(x^l, e)$ and $\beta_2(x^l, e) := \mathbb{E}[Y(0)|X^l = x^l, e(X) = e]$ are twice differentiable with respect to e.

(iv) $\sup_{d,z} \mathbb{E}(|\xi|^{2+\eta} \big| D = d, X^l = x^l, e = e) < \infty$ for some $\eta > 0$.

(v) Let $\sigma^2(d, x^l, e) := \text{var}(\xi | D = d, X^l = x^l, e = e)$ is differentiable with respect to e, $\sigma^2(d, x^l, e) > 0$.

(vi) The kernel function $K(\cdot)$ is Lipschitz and $\int tK'(t)dt < \infty$.

(vii) $\sqrt{N}h_2^4 \to 0$, $\sqrt{N}\lambda_2 h_2^2 \to 0$, $\sqrt{N}\lambda_2 \to 0$.

The following Theorem 7.1 presents the large sample properties of $\hat{\tau}(x^l)$ in (S.27).

Theorem 7.1. When X^l is discrete (unordered and ordered), under Assumptions 1' and 4, we have

$$\mathbb{V}(x^l)^{-1/2}\big\{\hat{\tau}(x^l) - \tau(x^l) - \text{bias}(\hat{\tau}(x^l))\big\} \xrightarrow{d} N(0, I_2),$$

where

$$\text{bias}(\hat{\tau}(x^l)) = \frac{1}{2}\mu_2 h_2^2 E[\beta''_{ee}(X_i^l, e_i)|X_i^l = x^l],$$

$$\mathbb{V}(x^l) = \frac{1}{N}\text{var}\left[\xi_i \frac{D_i - e_i}{e_i(1-e_i)} \cdot L_{\lambda_1}(X_i^l, x^l) - B^\intercal \psi(X_i) + L_{\lambda_2}(X_i^l, x^l)\{\beta(X_i^l, e_i) - \tau(X_i^l)\}\right]$$

with $B = P(X_i^l = x^l)\Big\{\mathbb{E}[g'(X_i^\intercal \alpha)\beta'_e(X_i^l, e_i)X_i|X_i^l = x^l] + \mathbb{E}[\xi_i \frac{g'(X_i^\intercal \alpha)}{e_i(1-e_i)} X_i|X_i^l = x^l] \cdot \int tK'(t)dt\Big\}$.

As in the case of continuous X^l, the asymptotic variance of $\hat{\tau}(x^l)$ can be estimated by the plug-in method, which is the sample variance of

$$\hat{\xi}_i \frac{D_i - \hat{e}_i}{\hat{e}_i(1-\hat{e}_i)} \cdot L_{\lambda_1}(X_i^l, x^l) - \hat{B}^\intercal \psi(X_i) + L_{\lambda_1}(X_i^l, x^l)\{\hat{\beta}(X_i^l, e_i) - \hat{\tau}(X_i^l)\},\ i = 1, ..., N,$$

where $\hat{\xi}_i = Y_i - D_i \hat{\beta}(X_i^l, e_i) - \hat{\beta}_2(X_i^l, e_i)$,

$$\hat{B} = \frac{1}{N} \sum_{\{i: X_i^l = x^l\}} \left\{ g'(X_i^T \hat{\alpha})\hat{\beta}'_e(X_i^l, e_i) + \int tK'(t)dt \cdot \hat{\xi}_i \frac{g'(X_i^T \hat{\alpha})}{\hat{e}_i(1-\hat{e}_i)} \right\} X_i,$$

$\hat{\beta}'_e(X_i^l, e_i)$ can be obtained by regressing $\hat{\beta}(X^l, e)$ on $\hat{e}$ within the subpopulation $\{i : X_i^l = x^l\}$.

7.3 Proof

Here we only consider the unordered case. The ordered case follows a similar argument. Assume that X^l takes value on $\{1, \cdots, K\}$ and denote $\int dx^l de_i = \sum_{x^l=1}^K \int de_i$. We also define $Z = (X^l, e)^\intercal$, $D^\dagger = (D, 1)^\intercal$, then $\tilde{\beta}(x^l, e)$, $\hat{\beta}(x^l, e)$, $\tilde{\tau}(x^l, e)$ and $\hat{\tau}(x^l, e)$ can be written as $\tilde{\beta}(z)$, $\hat{\beta}(z)$, $\tilde{\tau}(z)$ and $\hat{\tau}(z)$, $K_h(Z_i - z) = L_{\lambda_1}(X_i^l, x^l) \cdot K_{h_2}(\hat{e}_i - e)$. We first present the asymptotic properties of $\tilde{\beta}(z)$.

Lemma 7.2. Under Assumptions 1 and 4, we have

$$\mathbb{V}(z)^{-1/2}\{\tilde{\beta}(z)-\beta(z)-\text{bias}(\tilde{\beta}(z))\} \xrightarrow{d} N(0,1),$$

where $\text{bias}(\tilde{\beta}(z))=\frac{1}{2}\mu_2 h_2^2\beta''_{ee}(z)$, and $\mathbb{V}(z)=(1,0)\{\nu/nh_2 f(z)\}\boldsymbol{\Omega}^{-1}\boldsymbol{M}\boldsymbol{\Omega}^{-1}(1,0)^\intercal$, $\boldsymbol{\Omega}=\mathbb{E}[D^\dagger D^{\dagger\intercal}|Z=z]=\begin{pmatrix} e & e \\ e & 1 \end{pmatrix}$, $\boldsymbol{M}=\mathbb{E}[D^\dagger D^{\dagger\intercal}\sigma^2(D,Z)|Z=z]$, $f(z)$ is the density function of Z.

Proof.

$$\tilde{\beta}(z)-\beta(z)=(1,0,0,0)\cdot H_N^{-1}(z)(A_{7N}(z)+A_{8N}(z)),$$

where

$$H_N(z)=N^{-1}\boldsymbol{\Gamma}^\intercal\boldsymbol{W}\boldsymbol{\Gamma}=N^{-1}\sum_{i=1}^{N}\Gamma_i K_h(Z_i-z)\cdot\Gamma_i^\intercal,$$

$$A_{7N}(z)=N^{-1}\sum_{i=1}^{N}\Gamma_i K_h(Z_i-z)\cdot\{D_i\beta(Z_i)+\beta_2(Z_i)-$$

$$\Gamma_i^\intercal(\beta(z),\beta_2(z),h_2\beta'_e(z),h_2\beta'_{2,e}(z))^\intercal\},$$

$$A_{8N}(z)=N^{-1}\sum_{i=1}^{N}\Gamma_i K_h(Z_i-z)\cdot\xi_i.$$

Consider $H_N(z)$.

$$H_N(z)=\begin{pmatrix} H_{N,0}(z) & H_{N,1e}(z) \\ H_{N,1e}(z) & H_{N,2e}(z) \end{pmatrix},$$

where $H_{N,0}(z) = N^{-1}\sum_{i=1}^{N} D_i^{\dagger} D_i^{T\dagger} K_h(Z_i - z)$, $H_{N,1e}(z) = N^{-1}\sum_{i=1}^{N} D_i^{\dagger} D_i^{T\dagger} K_h(Z_i - z)(e_i - e_0)/h_2$, $H_{N,2e}(z) = N^{-1}\sum_{i=1}^{N} D_i^{\dagger} D_i^{T\dagger} K_h(Z_i - z)((e_i - e_0)/h_2)^2$. Under condition 4(i) and 4(ii),

$$
\begin{aligned}
&\mathbb{E}[e_i K_{h_2}(e_i - e) L_{\lambda_1}(X_i^l, x^l)] \\
&= \int_{X_i^l}\int_{e_i} \lambda_1^{I\{X_i^l \neq x^l\}} e_i K_{h_2}(e_i - e) f(X_i^l, e_i) de_i dX_i^l = \sum_{s=1}^{K} \lambda_1^{I\{s \neq x^l\}} \int e_i K_{h_2}(e_i - e) f(s, e_i) de_i \\
&= \sum_{s=1}^{K} \lambda_1^{I\{s \neq x^l\}} [ef(s,e) + O(h_2)] = ef(x^l, e) + O(h_2) + \sum_{s \neq x^l} \lambda_1 e f(s,e) + O(\lambda_1 h_2) \\
&= ef(x^l, e) + O(h_2 + \lambda_1).
\end{aligned}
$$

Thus,

$$
\mathbb{E}[H_{N,0}(z)] = \mathbb{E}[D_i^{\dagger} D_i^{T\dagger} K_{h_2}(e_i - e) L_{\lambda_1}(X_i^l, x^l)] = E[\begin{pmatrix} e_i & e_i \\ e_i & 1 \end{pmatrix} K_{h_2}(e_i - e) L_{\lambda_1}(X_i^l, x^l)]
$$

$$
= f(z)\mathbf{\Omega} + O(h_2 + \lambda_1).
$$

Similarly, by calculation, $\text{var}\{H_{N,0}(z)\} = O(1/(Nh_2))$. Note that $H_{N,0}(z) = f(z)\mathbf{\Omega} + O_p(h_2 + \lambda_1 + 1/\sqrt{Nh_2})$. It is easy to show that

$$
H_n(z) = f(z)\begin{pmatrix} \mathbf{\Omega} & 0 \\ 0 & \mu_2\mathbf{\Omega} \end{pmatrix} + o_p(1).
$$

In addition, under conditions 4(i)–4(iii), we can show that

$$(1,1,0,0)A_{7N}(z) = \frac{1}{2}\mu_2 f(z)h_2^2 \mathbf{\Omega} \begin{pmatrix} \beta''_{ee}(z) \\ \\ \beta''_{2,ee}(z) \end{pmatrix} + O_p(\lambda_1 + h_2^4 + \sqrt{\frac{\lambda_1^2 + h_2^4}{Nh_2}}).$$

Under conditions 5(iv) and 5(v), $\sqrt{nh_2}(I_2, 0_{2\times 2})A_{8N}(z) \xrightarrow{d} N(0, v_0 f(z)\boldsymbol{M})$.

□

Lemma 7.3. Under Assumptions 1 and 4, the following equations are valid.

$$\tilde{\beta}(x^l, e) - \beta(x^l, e) = \frac{1}{2}\mu_2\beta''_{ee}(z)h_2^2 + \frac{1}{N}\sum_{j=1}^{N}\frac{(D_j - e)}{f(z)e(1-e)}K_h(Z_j - z)\xi_j + u_p(N^{-1/2}).$$

$$\hat{\beta}(x^l, e) - \beta(x^l, e) = \tilde{\beta}(x^l, e) - \beta(x^l, e) + \frac{1}{N}\sum_{j=1}^{N}\frac{(D_j - e)}{f(z)e(1-e)}K_h(Z_j - z)\{\xi_j + \hat{R}_j - R_j\}$$
$$+ \frac{1}{N}\sum_{j=1}^{N}\frac{(D_j - e)}{f(z)e(1-e)}\{\hat{\boldsymbol{K}}_h(Z_j - z) - K_h(Z_j - z)\}\xi_j + u_p(N^{-1/2}),$$

where $\hat{R}_j = D_j\beta(Z_j) + \beta_2(Z_j) - D_j[\beta(z) + \beta'_e(z)(\hat{e}_j - e)] - [\beta_2(z) + \beta'_{2,e}(z)(\hat{e}_j - e)]$, R_j has the same form of $\hat{R}_j$, but replaces $\hat{e}_j$ with e_j.

Proof. The asymptotic linear approximation of $\tilde{\beta}(X_i^l, e_i) - \beta(X_i^l, e_i)$ can be derived directly by Lemma 7.2. In addition, the linear approximation of $\hat{\beta}(X_i^l, e_i) - \beta(X_i^l, e_i)$ holds by an identical argument to the proof of Lemma 4.1.

□

Proof of Theorem 7.1. $\hat{\tau}(x^l)$ can be written as

$$\hat{\tau}(x^l) - \tau(x^l) = \frac{N^{-1}\sum_{i=1}^{N}(\hat{\beta}(X_i^l, e_i) - \tau(x^l))L_{\lambda_2}(X_i^l, x^l)}{N^{-1}\sum_{i=1}^{N} L_{\lambda_2}(X_i^l, x^l)} = \frac{A_{9N}(x^l) + A_{10N}(x^l) + A_{11N}(x^l)}{N^{-1}\sum_{i=1}^{N} L_{\lambda_2}(X_i^l, x^l)},$$

where

$$\begin{aligned} A_{9N}(x^l) &= N^{-1}\sum_{i=1}^{N} L_{\lambda_2}(X_i^l, x^l)\{\hat{\beta}(X_i^l, e_i) - \beta(X_i^l, e_i)\}, \\ A_{10N}(x^l) &= N^{-1}\sum_{i=1}^{N} L_{\lambda_2}(X_i^l, x^l)\{\beta(X_i^l, e_i) - \tau(X_i^l)\}, \\ A_{11N}(x^l) &= N^{-1}\sum_{i=1}^{N} L_{\lambda_2}(X_i^l, x^l)\{\tau(X_i^l) - \tau(x^l)\}. \end{aligned}$$

Step 1. Consider the denominator $N^{-1}\sum_{i=1}^{N} L_{\lambda_2}(X_i^l, x^l)$, since $L_{\lambda_2}(X_i^l, x^l) = I\{X_i^l = x^l\} + \lambda_2 I\{X_i^l \neq x^l\}$, we have

$$\begin{aligned} \mathbb{E}(L_{\lambda_2}(X_i^l, x^l)) &= \mathbb{E}[I\{X_i^l = x^l\}] + \lambda_2 \mathbb{E}[I\{X_i^l \neq x^l\}] \\ &= P(X_i^l = x^l) + \lambda_2(1 - P(X_i^l = x^l)) \\ &= P(X_i^l = x^l) + O(\lambda_2), \\ \mathrm{var}(L_{\lambda_2}(X_i^l, x^l)) &= \mathbb{E}[L_{\lambda_2}^2(X_i^l, x^l)] - \{\mathbb{E}[L_{\lambda_2}(X_i^l, x^l)]\}^2 \\ &= P(X_i^l = x^l) + O(\lambda_2^2) - \{P(X_i^l = x^l) + O(\lambda_2)\}^2 \\ &= P(X_i^l = x^l)(1 - P(X_i^l = x^l)) + O(\lambda_2). \end{aligned}$$

Therefore, $N^{-1}\sum_{i=1}^{N} L_{\lambda_2}(X_i^l, x^l) = P(X_i^l = x^l) + O_p(\lambda_2 + N^{-1/2})$.

Step 2. According to Lemma 7.3,

$$A_{9N}(x^l) = \frac{1}{2}\mu_2 \cdot \frac{1}{N}\sum_{i=1}^{N} L_{\lambda_2}(X_i^l, x^l)\beta_{ee}''(Z_i)h_2^2 + A_{9N1}(x^l) + A_{9N2}(x^l) + A_{9N3}(x^l) + u_p(N^{-1/2}),$$

where

$$A_{9N1}(x^l) = \frac{1}{N}\sum_{i=1}^{N}\left\{L_{\lambda_2}(X_i^l, x^l)\cdot\frac{1}{N}\sum_{j=1}^{N}\frac{(D_j - e_i)}{f(Z_i)e_i(1-e_i)}K_h(Z_j - Z_i)\xi_j\right\}$$

$$A_{9N2}(x^l) = \frac{1}{N}\sum_{i=1}^{N}\left\{L_{\lambda_2}(X_i^l, x^l)\cdot\frac{1}{N}\sum_{j=1}^{N}\frac{(D_j - e_i)}{f(Z_i)e_i(1-e_i)}\boldsymbol{K}_h(Z_j - Z_i)\{\hat{R}_j - R_j\}\right\},$$

$$A_{9N3}(x^l) = \frac{1}{N}\sum_{i=1}^{N}\left\{L_{\lambda_2}(X_i^l, x^l)\cdot\frac{1}{N}\sum_{j=1}^{N}\frac{(D_j - e_i)}{f(Z_i)e_i(1-e_i)}\{\hat{\boldsymbol{K}}_h(Z_j - Z_i) - \boldsymbol{K}_h(Z_j - Z_i)\}\xi_j\right\}.$$

By a similar argument to the proof of Theorem 2, it can be shown that

$$\begin{aligned} A_{9N1}(x^l) =& \frac{1}{N}\sum_{i=1}^{N}\left[\xi_i\frac{D_i - e_i}{e_i(1-e_i)}\cdot L_{\lambda_1}(X_i^l, x^l)\right] + o_p(N^{-1/2}),\\ A_{9N2}(x^l) =& -B_1^{\intercal}(\hat{\alpha} - \alpha) + o_p(N^{-1/2}),\\ A_{9N3}(x^l) =& -B_2^{\intercal}(\hat{\alpha} - \alpha) + o_p(N^{-1/2}), \end{aligned}$$

where $B_1 = P(X_i^l = x^l)\cdot\mathbb{E}[g'(X_i^{\intercal}\alpha)\beta_e'(Z_i)X_i|X_i^l = x^l]$, $B_2 = P(X_i^l = x^l)\cdot\mathbb{E}[\xi_i\frac{g'(X_i^{\intercal}\alpha)}{e_i(1-e_i)}X_i|X_i^l = x^l]\cdot\int tK'(t)dt$.

Step 3. Let $B = B_1 + B_2$. By combining $A_{9N1}(x^l), A_{9N2}(x^l), A_{9N3}(x^l)$ with $A_{10N}(x^l)$, under condition 1'(i), we have $\sqrt{n}\{A_{9N1}(x^l) + A_{9N2}(x^l) + A_{9N3}(x^l) + A_{10N}(x^l)\} \xrightarrow{d} N(0, \tilde{\boldsymbol{V}})$, where

$$\tilde{\boldsymbol{V}} = \text{var}\left[\xi_i\frac{D_i - e_i}{e_i(1-e_i)}\cdot L_{\lambda_1}(X_i^l, x^l) - B^{\intercal}\psi(X_i) + L_{\lambda_2}(X_i^l, x^l)\{\beta(Z_i) - \tau(X_i^l)\}\right].$$

Step 4. One can easily get that $\mathbb{E}(A_{11N}(x^l)) = O(\lambda_2)$, $\text{var}(A_{11N}(x^l)) = \lambda_2^2/N$, which implies that $A_{11N}(x^l) = o_p(N^{-1/2})$.

□

8. Description of the dataset in application

Table S1: Descriptions of variables in illustration 1, 2442 participants, 65 covariates.

Name	Summary Statistics			Name	Summary Statistics		
			Continuous Variables				
	Mean	Median	SD		Mean	Median	SD
VACC_H1N1_COUNT	0.2	0.0	0.4	INC_REF	2008.6	2009.0	0.5
VACC_PNEU_COUNT	0.5	0.0	0.8	N_ADULT_R	1.9	2.0	0.8
VACC_SEAS_COUNT	0.5	0.0	0.5	N_PEOPLE_R	2.7	2.0	1.4
ILI_TIME_OFF	-1.9	-4.6	2.9	FLUWT	4097.7	1910.2	7046.9
HH_CHILD_R	0.7	0.0	1.0	AGE	46.2	47.0	15.0

Name	Summary Statistics	Name	Summary Statistics
	Discrete Variables		
	Number of Categories		Number of Categories
INT_MONTH	9	Q24	5
SAMP_DESIG	2	Q24_B	5
VACC_H1N1_F	2	DOCREC_BOTH_F	3
VACC_PNEU_F	2	DOCREC_DKNW_F	3
VACC_SEAS_F	2	DOCREC_H1N1_F	3
B_H1N1_ANTIV	2	DOCREC_NTHR_F	3
B_H1N1_AVOID	2	DOCREC_REFD_F	2
B_H1N1_FMASK	2	DOCREC_SEAS_F	3
B_H1N1_HANDS	2	ILI_OTHER_F	2
B_H1N1_LARGE	2	ILI_TREAT_F	2
B_H1N1_RCONT	2	CHRONIC_MED_F	2
B_H1N1_TOUCH	2	CLOSE_UNDER6MO_F	2
CONCERN_NONE_F	2	HEALTH_WORKER_F	2
CONCERN_NOTV_F	2	PATIENT_CONTACT_F	3
CONCERN_SOME_F	2	EDUCATION_COMP	4
CONCERN_VERY_F	2	HISP_I	2
HQ23	5	INC_CAT1	8
HQ24	5	INC_POV	4
HQ24_B	5	MARITAL	2
INT_H1N1_DKNW_F	3	Q95	4
INT_H1N1_DNOT_F	3	RACEETH4_I	4
INT_H1N1_DYES_F	3	RACE_I_R	3
INT_H1N1_PNOT_F	3	RENT_OWN_R	3
INT_H1N1_PYES_F	3	SEX_I	2
INT_H1N1_REFD_F	3	CEN_REG	4
KNOW_H1N1_ALOT_F	2	HHS_REGION	10
KNOW_H1N1_LITL_F	2	MSA3_I	3
KNOW_H1N1_NONE_F	2	STATE	51
Q23	5		

Notes: Variable names are given in Centers for Disease Control and Prevention (2010).

Table S2: Descriptions of variables in illustration 2, 8425 participants, 62 covariates.

Name	Summary Statistics			Name	Summary Statistics		
Continuous Variables							
	Mean	Median	SD		Mean	Median	SD
VACC_H1N1_COUNT	0.3	0.0	0.5	N_ADULT_R	1.9	2.0	0.7
VACC_PNEU_COUNT	0.3	0.0	0.6	N_PEOPLE_R	2.5	2.0	1.3
VACC_SEAS_COUNT	0.5	0.0	0.5	FLUWT	4037.1	2078.5	6093.7
Q9_NUM	3.2	2.0	4.2	AGE	46.9	48.0	13.7
HH_CHILD_R	0.6	0.0	0.9				
Discrete Variables							
	Number of Categories				Number of Categories		
INT_MONTH	6			DOCREC_H1N1_F	2		
INT_MONTH	6			DOCREC_H1N1_F	2		
SAMP_DESIG	2			DOCREC_NTHR_F	2		
VACC_H1N1_F	2			DOCREC_SEAS_F	2		
VACC_PNEU_F	2			ILI_F	2		
VACC_SEAS_F	2			ILI_OTHER_F	2		
B_H1N1_ANTIV	2			PSL_1	2		
B_H1N1_AVOID	2			CHRONIC_MED_F	2		
B_H1N1_FMASK	2			CLOSE_UNDER6MO_F	2		
B_H1N1_HANDS	2			HEALTH_WORKER_F	2		
B_H1N1_LARGE	2			PATIENT_CONTACT_F	2		
B_H1N1_RCONT	2			EDUCATION_COMP	4		
B_H1N1_TOUCH	2			HISP_I	2		
CONCERN_NONE_F	2			INC_CAT1	7		
CONCERN_NOTV_F	2			INC_POV	3		
CONCERN_SOME_F	2			INSURE	3		
CONCERN_VERY_F	2			MARITAL	2		
HQ23	5			Q95_INDSTR	23		
HQ24	5			Q95_OCCPN	25		
HQ24_B	5			RACEETH4_I	4		
KNOW_H1N1_ALOT_F	2			RACE_I_R	3		
KNOW_H1N1_LITL_F	2			RENT_OWN_R	3		
KNOW_H1N1_NONE_F	2			SEX_I	2		
Q23	5			CEN_REG	4		
Q24	5			HHS_REGION	10		
Q24_B	5			MSA3_I	3		
DOCREC_BOTH_F	2			STATE	51		
DOCREC_DKNW_F	2						

References

Abrevaya, J., Y. C. Hus, and R. P. Lieli (2015). Estimating conditional average treatment effect. *Journal of Business and Economic Statistics 33*, 485–505.

Aitchison, J. and C. Aitken (1976). Multivariate binary discrimination by the kernel method. *Biometrika 63*(3), 413–420.

Anderson, T. W. (2003). *An introduction to multivariate statistical analysis* (The third ed.). John Wiley & Sons, Inc.

Centers for Disease Control and Prevention (2010). *National 2009 H1N1 Flu Survey Questionnaire [2010 Q1]*. https://www.cdc.gov/nchs/data/nis/h1n1/: Centers for Disease Control and Prevention.

Gu, L. and L. Yang (2015). Oracally efficient estimation for single-index link function with simultaneous confidence band. *Electronic Journal of Statistics 9*, 1540–1561.

Lee, S., R. Okui, and Y. J. Whang (2017). Doubly robust uniform confidence band for the conditional average treatment effect function. *Journal of Applied Econometrics 32*, 1207–1225.

Li, Q. and J. S. Racine (2007). *Nonparametric econometrics*. Princeton University Press.

Li, Q. and J. S. Racine (2010). Smooth varying-coefficient estimation and inference for qualitative and quantitative data. *Ecomometric Theory 26*(6), 1607–1637.

Rosenbaum, P. R. and D. B. Rubin (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika 70*, 41–55.

van der Vaart, A. W. (1998). *Asymptotic statistics.* Cambridge University Press.

van der Vaart, A. W. and J. A. Wellner (1996). *Weak convergence and empirical processes: with application to statistics.* Springer.