A Sharp Test for the Judge Leniency Design

Online Supplementary Materials

Mohamed Coulibaly Yu-Chin Hsu Ismael Mourifié Yuanyuan Wan

B Proof of Theorem 1

Proof. Theorem 1-(i) is a direct application of Heckman and Vytlacil (2005)'s testable implications where $g(Y) = 1\{Y \in (y, y']\}$ for $y \le y'$. We focus on part (ii).

We define some notation. Let $\mathcal{L}(\mathcal{P})$ be the set of limit points of $\mathcal{P}$, $\mathcal{L}^o(\mathcal{P})$ be a set of interior point of $\mathcal{P}$, and $\mathcal{C}(\mathcal{P})$ be the closure of $\mathcal{P}$. Furthermore, let $I(\mathcal{P}) = \mathcal{C}(\mathcal{P})/\mathcal{L}^o(\mathcal{P})$ be the complement of $\mathcal{L}^o(\mathcal{P})$ in the closure of $\mathcal{P}$. So $I(\mathcal{P})$ also contains isolation points. Note that $\mathcal{L}^o(\mathcal{P})$ can be written as a union of countable or finite exclusive open intervals: $\mathcal{L}^o(\mathcal{P}) = \cup_{j=1}^{J}(a_j, b_j)$, where $(a_j, b_j) \subseteq \mathcal{P}$, $b_j < a_{j+1}$, and J can be infinity. Let $\Omega(\mathcal{P})$ be a collection of intervals belonging to $(0, 1]$ defined as follows:

$$\Omega(\mathcal{P}) \equiv \big\{(p, p'] : p, p' \in I(\mathcal{P}) \cup \{0, 1\} \text{ and for all } \tilde{p} \text{ such that } p < \tilde{p} < p', \tilde{p} \notin \mathcal{P}\big\}.$$

So the interior of each interval does not intersect with $\mathcal{P}$. $\Omega(\mathcal{P})$ contains a generic element $(c_k, d_k]$, where c_k, $d_k \in I(\mathcal{P})$, $d_k \le c_{k+1}$, $k = 1, 2, \cdots, K$ with K possibly equals to ∞, depending on how many isolation points there are in $\mathcal{P}$. Note that with above notation, for any $v \in (0, 1]$, v must belongs to one of the following categories: (i) an element of $\mathcal{L}^o(\mathcal{P})$ so that $v \in (a_j, b_j)$ for some j, (ii) $v \in \mathcal{L}(\mathcal{P})/\mathcal{L}^o(\mathcal{P})$, and (iii) there exist an integer k such that $v \in (c_k, d_k]$. The following figure illustrates the partition of the unit interval.

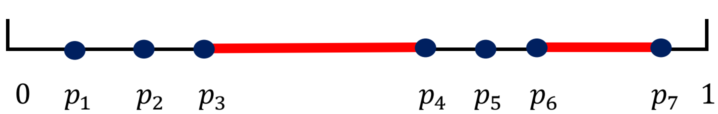

Figure 7: An illustration: $\mathcal{P} = \{p_1, p_2, p_5\} \cup [p_3, p_4] \cup [p_6, p_7]$, $\mathcal{L}^o(\mathcal{P}) = (p_3, p_4) \cup (p_6, p_7)$, and $\Omega(\mathcal{P}) = \{(0, p_1], (p_1, p_2], (p_4, p_5], (p_5, p_6], (p_7, 1]\}$.

We will assume that $\mathbb{P}(y < Y \leq y', D = 1|P = p)$ and $\mathbb{P}(y < Y \leq y', D = 0|P = p)$ are continuously differentiable over $\mathcal{L}^o$ as a regularity condition under which the local instrumental variable (LIV) estimand is well defined.

First, we construct $\tilde{V}$ and $\tilde{D}$ as follows:

$$\mathbb{P}(\tilde{V} \leq t|P = p) = t, \forall (t, p) \in [0, 1] \times \mathcal{P}, \text{ and } \tilde{D} = 1\{P(Z) \geq \tilde{V}\}.$$

By construction, Assumption 2.4 is satisfied. Next, we propose the following distribution for $\tilde{Y}_1|\tilde{V}, P$. For any arbitrary $p \in \mathcal{P}$ and $v \in (0, 1]$, we define

$$\mathbb{P}(\tilde{Y}_1 \leq y|\tilde{V} = v, P = p) = \begin{cases} \frac{\partial}{\partial t}\mathbb{P}(Y \leq y, D = 1|P = t)|_{t=v} & \text{if } v \in \mathcal{L}^o(\mathcal{P}) \\ \lim_{\tilde{v}\to v} \frac{\partial}{\partial t}\mathbb{P}(Y \leq y, D = 1|P = t)|_{t=\tilde{v}} & \text{if } v \in \mathcal{L}(\mathcal{P})/\mathcal{L}^o(\mathcal{P}) \\ \frac{\mathbb{P}(Y \leq y, D=1|P=d_k) - \mathbb{P}(Y \leq y, D=1|P=c_k)}{d_k - c_k} & \text{if } v \notin L(P) \text{ but } v \in (c_k, d_k] \in \Omega(\mathcal{P}). \end{cases}$$

$$\mathbb{P}(\tilde{Y}_0 \leq y|\tilde{V} = v, P = p) = \begin{cases} -\frac{\partial}{\partial v}\mathbb{P}(Y \leq y, D = 0|P = t)|_{t=v} & \text{if } v \in \mathcal{L}^o(\mathcal{P}) \\ -\lim_{\tilde{v}\to v} \frac{\partial}{\partial v}\mathbb{P}(Y \leq y, D = 0|P = t)|_{t=\tilde{v}} & \text{if } v \in \mathcal{L}^o(\mathcal{P}) \\ \frac{\mathbb{P}(Y \leq y, D=0|P=c_k) - \mathbb{P}(Y \leq y, D=0|P=d_k)}{d_k - c_k} & \text{if } v \notin L^o(P) \text{ but } v \in (c_k, d_k] \in \Omega(\mathcal{P}). \end{cases}$$

Note that the conditioning on $\tilde{V} = v$ and $P = p$, the distribution of $\tilde{Y}_1$ does not depend on p. Hence, Assumption 2.1 is satisfied by construction.

We now show that the distribution function constructed above is well defined. We focus on $\mathbb{P}(\tilde{Y}_1 \leq y|\tilde{V} = v, P = p)$ and the verification for $\mathbb{P}(\tilde{Y}_0 \leq y|\tilde{V} = v, P = p)$ is analogous. Let $\underline{y}$ and $\overline{y}$ be the lower and upper bounds of the support of Y, respectively.

1. $\mathbb{P}(\tilde{Y}_1 < \underline{y} - \epsilon|\tilde{V} = v, P = p) = 0$ for all $v \in [0, 1]$ and for any arbitrarily small $\epsilon > 0$. To see this, suppose $v \notin \mathcal{L}(\mathcal{P})$, then there exists $(c_k, d_k] \in \Omega(\mathcal{P})$ such that $v \in (c_k, d_k]$, therefore,

$$\begin{aligned} &\mathbb{P}(\tilde{Y}_1 \leq \underline{y} - \epsilon|\tilde{V} = v, P = p) \\ =&\frac{\mathbb{P}(Y \leq \underline{y} - \epsilon, D = 1|P = d_k) - \mathbb{P}(Y \leq \underline{y} - \epsilon, D = 1|P = c_k)}{d_k - c_k} = \frac{0 - 0}{d_k - c_k} = 0. \end{aligned}$$

On the other hand, if $v \in \mathcal{L}^o(\mathcal{P})$, then $\mathbb{P}(Y \leq \underline{y} - \epsilon, D = 1|P = \tilde{v}) = 0$ for all $\tilde{v}$ in a small neighborhood of v, which implies $\frac{\partial}{\partial v}\mathbb{P}(Y \leq \underline{y} - \epsilon, D = 1|P = v) = 0$. The case that $v \in \mathcal{L}^o(\mathcal{P})$ follows straightforwardly.

2. $\mathbb{P}(\tilde{Y}_1 \leq \overline{y}|\tilde{V} = v, P = p) = 1$. First, if $v \in \mathcal{L}^o(\mathcal{P})$, then

$$\mathbb{P}(Y \leq \overline{y}, D = 1|P = v) = \mathbb{P}(D = 1|P = v) = v \Rightarrow \frac{\partial}{\partial v}\mathbb{P}(Y \leq \overline{y}, D = 1|P = v) = 1.$$

On the other hand, if $v \notin \mathcal{L}(\mathcal{P})$, then

$$\mathbb{P}(\tilde{Y}_1 \leq \overline{y}|\tilde{V} = v, P = p) = \frac{\mathbb{P}(Y \leq \overline{y}, D = 1|P = d_k) - \mathbb{P}(Y \leq \overline{y}, D = 1|P = c_k)}{p' - p} = \frac{d_k - c_k}{d_k - c_k} = 1.$$

3. $\mathbb{P}(\tilde{Y}_1 \leq y|\tilde{V} = v, P = p)$ is nondecreasing in y. For $y < y'$ we have

$$\begin{aligned}
&\mathbb{P}(\tilde{Y}_1 \leq y'|\tilde{V} = v, P = p) - \mathbb{P}(\tilde{Y}_1 \leq y|\tilde{V} = v, P = p) \\
&= \begin{cases} \frac{\partial}{\partial t}\mathbb{P}(y < Y \leq y', D = 1|P = t)|_{t=v} \geq 0 \quad \text{if } v \in \mathcal{L}^o(\mathcal{P}), \\ \lim_{\tilde{v}\to v} \frac{\partial}{\partial t}\mathbb{P}(y < Y \leq y, D = 1|P = t)|_{t=\tilde{v}} \geq 0 \quad \text{if } v \in \mathcal{L}(\mathcal{P})/\mathcal{L}^o(\mathcal{P}) \\ \frac{\mathbb{P}(y<Y\leq y', D=1|P=d_k) - \mathbb{P}(y<Y\leq y', D=1|P=c_k)}{d_k - c_k} \geq 0 \text{ if } v \notin L^o(P) \text{ but } v \in [c_k, d_k] \in \Omega(\mathcal{P}), \end{cases}
\end{aligned}$$

where the last inequalities hold whenever the testable implications hold, i.e. $\mathbb{P}(y < Y \leq y', D = 1|P = p)$ is a non-decreasing function for all $p \in \mathcal{P}$ and all $y < y'$, and by the continuous differentiability of $\mathbb{P}(y < Y \leq y', D = 1|P = p)$ over $\mathcal{L}(\mathcal{P})$.

Finally, we show that $(\tilde{V}, \tilde{Y}_d, P(Z))$, $d \in \{0, 1\}$ is observationally equivalent to $(V, Y_d, P(Z))$ $d \in \{0, 1\}$. For this, we show that the conditioning distribution of $(\tilde{Y}, \tilde{D})$ given $P(Z)$ is the same as the conditioning of (Y, D) given $P(Z)$. Take an arbitrary $p \in \mathcal{P}$.

Suppose first $p \notin \mathcal{L}^o(\mathcal{P})$, then $(0, p]$ can be expressed as unions of exclusive intervals $\left(\cup_{j=1}^{J^*}(a_j, b_j)\right) \cup \left(\cup_{k=1}^{K^*}(c_k, d_k]\right)$ for some J^* and K^*, where $(a_j, b_j)s$ are connected subsets of

$\mathcal{P}$. Therefore,

$$\begin{aligned}\mathbb{P}(\tilde{Y} \leq y, \tilde{D} = 1|P = p) = \mathbb{P}(\tilde{Y}_1 \leq y, \tilde{V} \leq p|P = p) = \int_0^p \mathbb{P}(\tilde{Y}_1 \leq y|\tilde{V} = v, P = p)dv \\ = \sum_{j=1}^{J^*} \int_{a_j}^{b_j} \mathbb{P}(\tilde{Y}_1 \leq y|\tilde{V} = v, P = p)dv + \sum_{k=1}^{K^*} \int_{c_k}^{d_k} \mathbb{P}(\tilde{Y}_1 \leq y|\tilde{V} = v, P = p)dv \\ = \sum_{j=1}^{J^*} (\mathbb{P}(Y \leq y, D = 1|P = b_j) - \mathbb{P}(Y \leq y, D = 1|P = a_j)) \\ + \sum_{k=1}^{K^*} (\mathbb{P}(Y \leq y, D = 1|P = d_k) - \mathbb{P}(Y \leq y, D = 1|P = c_k)) \\ = \mathbb{P}(Y \leq y, D = 1|P = p) - \mathbb{P}(Y \leq y, D = 1|P = 0) = \mathbb{P}(Y \leq y, D = 1|P = p),\end{aligned}$$

where the first equality is by construction that $\tilde{V}$ satisfies Assumption 2.4, the third equality holds because $(0, p]$ can be expressed as unions of exclusive intervals $\left(\cup_{j=1}^{J^*}(a_j, b_j)\right) \cup \left(\cup_{k=1}^{K^*}(c_k, d_k]\right)$, the fourth equality is obtained by inserting the constructed counterfactural distributions, and the last one holds because $\mathbb{P}(Y \leq y, D = 1|P = 0) = 0$.

Suppose that $p \in (a_{j^*}, b_{j^*}) \subseteq \mathcal{L}^0(\mathcal{P})$ for some j^*, then the right hand side equals to

$$\begin{aligned}\mathbb{P}(\tilde{Y} \leq y, \tilde{D} = 1|P = p) = \mathbb{P}(\tilde{Y}_1 \leq y, \tilde{V} \leq p|P = p) = \int_0^p \mathbb{P}(\tilde{Y}_1 \leq y|\tilde{V} = v, P = p)dv \\ = \int_0^{a_{j^*}} \mathbb{P}(\tilde{Y}_1 \leq y|\tilde{V} = v, P = p)dv + \int_{a_{j^*}}^p \mathbb{P}(\tilde{Y}_1 \leq y|\tilde{V} = v, P = p)dv \\ = \mathbb{P}(Y \leq y, D = 1|P = a_{j^*}) + \int_{a_{j^*}}^p \frac{\partial}{\partial v}\mathbb{P}(Y \leq y, D = 1|P = v)dv \\ = \mathbb{P}(Y \leq y, D = 1|P = a_{j^*}) + \mathbb{P}(Y \leq y, D = 1|P = p) - \mathbb{P}(Y \leq y, D = 1|P = a_{j^*}) \\ = \mathbb{P}(Y \leq y, D = 1|P = p),\end{aligned}$$

where the $\int_0^{a_{j^*}} \mathbb{P}(\tilde{Y}_1 \leq y|\tilde{V} = v, P = p)dv = \mathbb{P}(Y \leq y, D = 1|P = a_{j^*})$ holds by the above argument and the fifth equality holds by inserting the constructed counterfactural distributions. This completes the proof.□

C Proof of Theorem 2

We begin by listing a few regularity conditions for the proof of Theorem 2.

Assumption C.1 *The observations* $\{(Y_i, D_i, Z_i, X_i)\}_{i=1}^n$ *are i.i.d. across* i.

Assumption C.2 *We impose the following smoothness conditions:*

1. *The conditional density of* (Y, D) *given* $P(Z, \theta_0) = p$, *denoted by* $f_{Y,D|P}(y, d|p)$, *is Lipschitz continuous both in* p *on* $\mathcal{P}$ *and in* y *on* $\mathcal{Y}$ *for* $d = 0, 1$.

2. *For all* $z \in \mathcal{Z}$, $P(z, \theta)$ *is continuously differentiable in* θ *at* θ_0 *with bounded derivatives.*

Note that Assumption C.2-(1) does not exclude the case of discrete propensity score. When P is discrete and $\mathcal{P}$ contains finite many distinguished elements, any convergent sequence in $\mathcal{P}$ must be a constant sequence eventually, and in that case Assumption C.2-(1) holds automatically. Assumption C.2-(1) implies that the functions m_d and ω, defined in Equations (A.1) to (A.3), are continuous functions of ℓ. Assumption C.2-(2) implies that the class of functions $\{1(p \leq P(Z, \theta) \leq p + r_p) : \theta \in \Theta, p \in [0, 1], r_p \in [0, 1]\}$ is a Vapnik-Chervonenkis (VC) class of function.

Assumption C.3 *The parameter space* Θ *for* θ_0 *is compact, and* θ_0 *is in the interior of* Θ. *The estimator* $\hat{\theta}$ *admits an influence function of the following form,*

$$\sqrt{n}(\hat{\theta} - \theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} s(D_i, Z_i, \theta_0) + o_p(1), \tag{C.1}$$

where $s(\cdot, \cdot, \cdot)$ *is measurable, satisfying* $\mathbb{E}[s(D_i, Z_i, \theta_0)] = 0$, $\mathbb{E}[\sup_\theta |s(D_i, Z_i, \theta)|] < \infty$, *and* $V(\sup_\theta |s(D_i, Z_i, \theta)|) < \infty$.

Assumption C.3 is satisfied for common maximum likelihood estimators and parametric binary response models. For example, if one estimates θ_0 by Probit model $D_i = 1[Z_i'\theta_0 \geq V_i]$, with $V_i \sim N(0, 1)$, then

$$s(D_i, Z_i, \theta_0) = \frac{\phi((2D_i - 1)Z_i'\theta_0)}{\Phi((2D_i - 1)Z_i'\theta_0)} Z_i.$$

If the Logit model is used, then

$$s(D_i, Z_i, \theta_0) = \left(D_i - \frac{\exp(Z_i'\theta_0)}{1 + \exp(Z_i'\theta_0)} \right) Z_i.$$

Assumption C.4 *The estimator* $\hat{\theta}^b$ *satisfies that*

$$\sqrt{n}(\hat{\theta}^b - \hat{\theta}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (W_i - 1) \cdot s(D_i, Z_i, \theta_0) + o_p(1), \tag{C.2}$$

where $s_\theta(\cdot)$ is the same as in Assumption C.3.

Assumption C.4 is satisfied under our weighted bootstrap procedure.

The proof of Theorem Theorem 2 follows from the same arguments as Theorems 5.1 and 5.2 of Hsu (2017) once Lemmas D.1 to D.4 are established, and is omitted for the sake of brevity.

D Lemmas and Intermediary Results

This section collects useful Lemmas, intermediary results, and additional assumptions for establishing the asymptotic results in Theorem 2.

Lemma D.1 *Suppose Assumptions C.2 and C.3 are satisfied, then uniformly in $\ell \in \mathcal{L}$,*

$$\begin{aligned}&\sqrt{n}(\hat{m}_1(y, r_y, p, r_p, \hat{\theta}) - m_1(y, r_y, p, r_p, \theta_0))\\ =&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_{m_1,i}(y, r_y, p, r_p, \theta_0) + o_p(1)\\ \equiv&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(m_{1,i}(y, r_y, p, r_p, \theta_0) - m_1(y, r_y, p, r_p, \theta_0) + \nabla_\theta m_1(y, r_y, p, r_p, \theta_0)\cdot s(D_i, Z_i, \theta_0)) + o_p(1).\end{aligned} \tag{D.1}$$

$$\begin{aligned}&\sqrt{n}(\hat{m}_0(y, r_y, p, r_p, \hat{\theta}) - m_0(y, r_y, p, r_p, \theta_0))\\ =&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_{m_0,i}(y, r_y, p, r_p, \theta_0) + o_p(1)\\ \equiv&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(m_{0,i}(y, r_y, p, r_p, \theta_0) - m_0(y, r_y, p, r_p, \theta_0) + \nabla_\theta m_0(y, r_y, p, r_p, \theta_0)\cdot s(D_i, Z_i, \theta_0)) + o_p(1),\end{aligned} \tag{D.2}$$

$$\begin{aligned}&\sqrt{n}(\hat{w}(p, r_p, \hat{\theta}) - w(p, r_p, \theta_0))\\ =&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_{w,i}(p, r_p, \theta_0) + o_p(1)\\ \equiv&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(w_i(p, r_p, \theta_0) - w(p, r_p, \theta_0) + \nabla_\theta w(p, r_p, \theta_0)\cdot s(D_i, Z_i, \theta_0)) + o_p(1)\end{aligned} \tag{D.3}$$

where functions m_d and w are defined in Equations (A.1) to (A.3) and

$$m_{1i}(y, r_y, p, r_p, \theta) = D_i 1(y \le Y_i \le y + r_y) 1(p \le P(Z_i, \theta) \le p + r_p),$$
$$m_{0i}(y, r_y, p, r_p, \theta) = (D_i - 1) 1(y \le Y_i \le y + r_y) 1(p \le P(Z_i, \theta) \le p + r_p),$$
$$w_i(p, r_p, \theta) = 1(p \le P(Z_i, \theta) \le p + r_p).$$

Proof. *Let $f_P(p)$ denote the density function of $P(Z;\theta_0)$. Following Hsu and Lieli (2021), we calculate the derivatives for $m_d(y, r_y, p, r_p, \cdot)$ and $w(p, r_p, \cdot)$ as:*

$$\nabla_\theta m_1(y, r_y, p, r_p, \theta_0) = \mathbb{E}[D1(y \le Y \le y + r_y)|P(Z, \theta_0) = p] \cdot f_P(p)\mathbb{E}[\nabla_\theta P(Z, \theta_0)|P(Z, \theta_0) = p]$$
$$- \mathbb{E}[D1(y \le Y \le y + r_y)|P(Z, \theta_0) = p + r_p] \cdot f_P(p + r_p)\mathbb{E}[\nabla_\theta P(Z, \theta_0)|P(Z, \theta_0) = p + r_p],$$

$$\nabla_\theta m_0(y, r_y, p, r_p, \theta_0) = \mathbb{E}[(D-1)1(y \le Y \le y+r_y)|P(Z, \theta_0) = p] \cdot f_P(p)\mathbb{E}[\nabla_\theta P(Z, \theta_0)|P(Z, \theta_0) = p]$$
$$- \mathbb{E}[(D - 1)1(y \le Y \le y + r_y)|P(Z, \theta_0) = p + r_p] \cdot f_P(p + r_p)\mathbb{E}[\nabla_\theta P(Z, \theta_0)|P(Z, \theta_0) = p + r_p],$$

$$\nabla_\theta w(p, r_p, \theta_0) = f_P(p)\mathbb{E}[\nabla_\theta P(Z, \theta_0)|P(Z, \theta_0) = p] - f_P(p+r_p)\mathbb{E}[\nabla_\theta P(Z, \theta_0)|P(Z, \theta_0) = p+r_p].$$

Now we prove Equation (D.1), the results for Equations (D.2) and (D.3) are similar. Note that

$$\begin{aligned}
&\sqrt{n}(\hat{m}_1(y, r_y, p, r_p, \hat{\theta}) - m_1(y, r_y, p, r_p, \theta_0)) \\
=&\sqrt{n}(\hat{m}_1(y, r_y, p, r_p, \hat{\theta}) - m_1(y, r_y, p, r_p, \hat{\theta})) + \sqrt{n}(m_1(y, r_y, p, r_p, \hat{\theta}) - m_1(y, r_y, p, r_p, \theta_0)) \\
=&\sqrt{n}(\hat{m}_1(y, r_y, p, r_p, \hat{\theta}) - m_1(y, r_y, p, r_p, \hat{\theta})) + \nabla_\theta m_1(y, r_y, p, r_p, \theta_0)'\sqrt{n}(\hat{\theta} - \theta_0) + o(\sqrt{n}\|\hat{\theta} - \theta_0\|) \\
=&\sqrt{n}(\hat{m}_1(y, r_y, p, r_p, \hat{\theta}) - m_1(y, r_y, p, r_p, \hat{\theta})) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \nabla_\theta m_1(y, r_y, p, r_p, \theta_0) s(D_i, Z_i, \theta_0) + o_p(1)
\end{aligned} \tag{D.4}$$

where the second equality holds because $m_1(\ell, \theta)$ is continuously differentiable in θ under Assumption C.2-(2), and the third equality is due to Assumption C.3.

Let $\hat{\mathbb{G}}_{m_1}(\theta, \ell) \equiv \sqrt{n}(\hat{m}_1(y, r_y, p, r_p, \theta) - m_1(y, r_y, p, r_p, \theta))$, $\theta \in \Theta, \ell \in \mathcal{L}$. It remains to show *that* $\sup_{\ell\in\mathcal{L}}|\hat{\mathbb{G}}_{m_1}(\hat{\theta},\ell)-\hat{\mathbb{G}}_{m_1}(\theta_0,\ell)|=o_p(1)$.

By Assumption C.2-(ii), the class of functions $\{1(p\le P(Z,\theta)\le p+r_p):\theta\in\Theta,p\in[0,1],r_p\in[0,1]\}$ *is a Vapnik-Chervonenkis (VC) class of function. Therefore, the class of functions* $\{1\{y\le Y\le y+r_y\}\times 1(p\le P(Z,\theta):\theta\in\Theta,p\in[0,1],r_p\in[0,1],r_y\in[0,1]\}$ *is also VC class. Hence, the process* $\hat{\mathbb{G}}_{m_1}$ *is stochastically equicontinuous with respect to* (θ,ℓ)*. Note* $\hat{\theta}\xrightarrow{p}\theta_0$*, then there exist* $\delta_n\downarrow 0$ *such that with probability approaching one,* $(\hat{\theta},\ell)\in B((\theta_0,\ell),\delta_n)$*, where* $B((\theta_0,\ell),\delta_n)$ *is a ball in* $\Theta\times\mathcal{L}$ *centered at* (θ_0,ℓ) *with radius* δ_n*. Therefore,*

$$\begin{aligned}
&\sup_{\ell\in\mathcal{L}}|\sqrt{n}(\hat{m}_1(y,r_y,p,r_p,\hat{\theta})-m_1(y,r_y,p,r_p,\hat{\theta}))-\sqrt{n}(\hat{m}_1(y,r_y,p,r_p,\theta_0)-m_1(y,r_y,p,r_p,\theta_0))|\\
=&\sup_{\ell\in\mathcal{L}}|\hat{\mathbb{G}}_{m_1}(\hat{\theta},\ell)-\hat{\mathbb{G}}_{m_1}(\theta_0,\ell)|\\
\le&\sup_{\theta_0\in\Theta,\ell\in\mathcal{L}}\sup_{(\theta',\ell')\in B((\theta_0,\ell),\delta_n)}|\hat{\mathbb{G}}_{m_1}(\theta',\ell')-\hat{\mathbb{G}}_{m_1}(\theta_0,\ell)|=o_p(1).
\end{aligned}\tag{D.5}$$

where the last equality is by the stochastic equicontinuity of the process $\hat{\mathbb{G}}_{m_1}$*. Combine both Equations* (D.4) *and* (D.5)*, the result then follows.* □

Lemma D.2 *Suppose Assumptions 2.1 to 2.4, C.2* and *C.3* are *satisfied, then uniform in* ℓ*,*

$$\sqrt{n}(\hat{\nu}_1(y,r_y,p_1,p_2,r_p,\hat{\theta})-\nu_1(y,r_y,p_1,p_2,r_p,\theta_0))=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_{\nu_1,i}(y,r_y,p_1,p_2,r_p,\theta_0)+o_p(1),\tag{D.6}$$

$$\sqrt{n}(\hat{\nu}_0(y,r_y,p_1,p_2,r_p,\hat{\theta})-\nu_0(y,r_y,p_1,p_2,r_p,\theta_0))=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_{\nu_0,i}(y,r_y,p_1,p_2,r_p,\theta_0)+o_p(1),\tag{D.7}$$

where

$$\begin{aligned}
&\phi_{\nu_1,i}(y,r_y,p_1,p_2,r_p,\theta_0)=w(p_1,r_p,\theta_0)\cdot\phi_{m_1,i}(y,r_y,p_2,r_p,\theta_0)+m_1(y,r_y,p_2,r_p,\theta_0)\cdot\phi_{w,i}(p_1,r_p,\theta_0)\\
&-w(p_2,r_p,\theta_0)\cdot\phi_{m_1,i}(y,r_y,p_1,r_p,\theta_0)-m_1(y,r_y,p_1,r_p,\theta_0)\cdot\phi_{w,i}(p_2,r_p,\theta_0),\\
&\phi_{\nu_0,i}(y,r_y,p_1,p_2,,r_p,\theta_0)=w(p_1,r_p,\theta_0)\cdot\phi_{m_0,i}(y,r_y,p_2,r_p,\theta_0)+m_0(y,r_y,p_2,r_p,\theta_0)\cdot\phi_{w,i}(p_1,r_p,\theta_0)\\
&-w(p_2,r_p,\theta_0)\cdot\phi_{m_0,i}(y,r_y,p_1,r_p,\theta_0)-m_0(y,r_y,p_1,r_p,\theta_0)\cdot\phi_{w,i}(p_2,r_p,\theta_0).
\end{aligned}$$

Furthermore,

$$\sqrt{n}(\widehat{\nu}_1(\cdot,\widehat{\theta})-\nu_1(\cdot,\theta_0))\Rightarrow\Phi_{\nu_1}(\cdot),\qquad\sqrt{n}(\widehat{\nu}_0(\cdot,\widehat{\theta})-\nu_0(\cdot,\theta_0))\Rightarrow\Phi_{\nu_0}(\cdot),$$

where $\Phi_{\nu_1}(\cdot)$ and $\Phi_{\nu_0}(\cdot)$ are Gaussian processes with variance-covariance kernel generated by $\phi_{\nu_1}(\cdot,\theta_0)$ and $\phi_{\nu_0}(\cdot,\theta_0)$, respectively.

Proof. *We show Equation (D.6). Equation (D.7) holds analogously. Recall*

$$\hat{\nu}_1(\ell)=\hat{m}_1(y,r_y,p_2,r_p,\hat{\theta})\cdot\hat{w}(p_1,r_p,\hat{\theta})-\hat{m}_1(y,r_y,p_1,r_p,\hat{\theta})\cdot\hat{w}(p_2,r_p,\hat{\theta})$$

To save space, for generic ℓ, we write $\hat{m}_1(\hat{\theta})\equiv\hat{m}_1(\ell,\hat{\theta})$ and $\hat{w}(\hat{\theta})\equiv\hat{w}(\ell,\hat{\theta})$. Similarly, $m_1(\theta_0)\equiv m_1(\ell,\theta_0)$ and $w(\theta_0)\equiv w(\ell,\theta_0)$. Then,

$$\begin{aligned}\hat{m}_1(\hat{\theta})\hat{w}(\hat{\theta})-m_1(\theta_0)w(\theta_0)&=(\hat{m}_1(\hat{\theta})-m_1(\theta_0)+m_1(\theta_0))(\hat{w}(\hat{\theta})-w(\theta_0)+w(\theta_0))-m_1(\theta_0)w(\theta_0)\\&=(\hat{m}_1(\hat{\theta})-m_1(\theta_0))w(\theta_0)+(\hat{w}(\hat{\theta})-w(\theta_0))m_1(\theta_0)+(\hat{m}_1(\hat{\theta})-m_1(\theta_0))(\hat{w}(\hat{\theta})-w(\theta_0))\\&=(\hat{m}_1(\hat{\theta})-m_1(\theta_0))w(\theta_0)+(\hat{w}(\hat{\theta})-w(\theta_0))m_1(\theta_0)+o_p\left(\frac{1}{\sqrt{n}}\right),\end{aligned}$$

where the last equality is because $\hat{m}_1(\hat{\theta})-m_1(\theta_0)=O_p(1/\sqrt{n})$ and $\hat{w}(\hat{\theta})-w(\theta_0)=O_p(1/\sqrt{n})$ by Lemma D.1. Then we have

$$\begin{aligned}\hat{\nu}_1(\ell)-\nu_1(\ell)=&w(p_1,r_p,\theta_0)\cdot(\hat{m}_1(y,r_y,p_2,r_p,\hat{\theta})-m_1(y,r_y,p_2,r_p,\theta_0))\\&+m_1(y,r_y,p_2,r_p,\theta_0)\cdot(\hat{w}(p_1,r_p,\hat{\theta})-w(p_1,r_p,\theta_0))\\&-w(p_2,r_p,\theta_0)\cdot(\hat{m}_1(y,r_y,p_1,r_p,\hat{\theta})-m_1(y,r_y,p_1,r_p,\theta_0))\\&-m_1(y,r_y,p_1,r_p,\theta_0)\cdot(\hat{w}(p_2,r_p,\hat{\theta})-w(p_2,r_p,\theta_0))+o_p\left(\frac{1}{\sqrt{n}}\right).\end{aligned}$$

Equation (D.6) then follows by inserting Equations (D.1) to (D.3) to the above equation.

Finally, under Assumption C.2, each element of $\nabla_\theta m_1(y,r_y,p,r_p,\theta_0)$ is Lipschitz continuous in y, r_y, p, r_p and it implies that $\{\partial m_1(y,r_y,p,r_p,\theta_0)/\partial\theta_j:(y,r_y,p,r_p)\in[0,1]^4\}$ is a VC class of functions for each j. Similarly, each element of $\nabla_\theta w(p,r_p,\theta_0)$ is Lipschitz continuous in p, r_p. It follows that $\{\phi_{m_1}(y,r_y,p,r_p,\theta_0):(y,r_y,p,r_p)\in[0,1]^4\}$, $\{\phi_{m_0}(y,r_y,p,r_p,\theta_0):(y,r_y,p,r_p)\in[0,1]^4\}$ and $\{\phi_w(p,r_p,\theta_0):(p,r_p)\in[0,1]^2\}$ are all VC classes of functions. weak convergence follows from the fact that $\{\phi_{\nu_0}(y,r_y,p_1,p_2,,r_p,\theta_0):(y,r_y,p_1,p_2,r_p)\in[0,1]^5\}$ and

$\{\phi_{\nu_0}(y, r_y, p_1, p_2, , r_p, \theta_0) : (y, r_y, p_1, p_2, r_p) \in [0,1]^5\}$ *are both VC classes of functions. Therefore, we have*

$$\sqrt{n}(\widehat{\nu}_1(\cdot, \widehat{\theta}) - \nu_1(\cdot, \theta_0)) \Rightarrow \Phi_{\nu_1}(\cdot), \quad \sqrt{n}(\widehat{\nu}_0(\cdot, \widehat{\theta}) - \nu_0(\cdot, \theta_0)) \Rightarrow \Phi_{\nu_0}(\cdot).$$

□

Lemma D.3 *Suppose Assumptions 2.1 to 2.4 and C.2 to C.4 are satisfied, then uniform in ℓ over $\mathcal{L}$,*

$$\begin{aligned} &\sqrt{n}(\hat{\nu}_1^b(y, r_y, p_1, p_2, r_p, \hat{\theta}^b) - \hat{\nu}_1(y, r_y, p_1, p_2, r_p, \hat{\theta})) \\ =&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(W_i - 1)\phi_{\nu_1,i}(y, r_y, p_1, p_2, r_p, \theta_0) + o_p(1), \quad \text{(D.8)} \\ &\sqrt{n}(\hat{\nu}_0^b(y, r_y, p_1, p_2, r_p, \hat{\theta}^b) - \hat{\nu}_0(y, r_y, p_1, p_2, r_p, \hat{\theta})) \\ =&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(W_i - 1)\phi_{\nu_0,i}(y, r_y, p_1, p_2, r_p, \theta_0) + o_p(1), \quad \text{(D.9)} \end{aligned}$$

where $\phi_{\nu_1,i}(y, r_y, p_1, p_2, r_p, \theta_0)$ and $\phi_{\nu_0,i}(y, r_y, p_1, p_2, , r_p, \theta_0)$ are the same as in Lemma D.2.

The proof to Lemma D.3 is similar to Lemma D.2 and is therefore omitted.

Lemma D.4 *Suppose Assumptions 2.1 to 2.4 and C.2 to C.4 are satisfied, then $\hat{\sigma}_d^2(\ell)$ defined in (3.8) satisfies that for $d = 0, 1$, $\sup_\ell |\hat{\sigma}_d^2(\ell) - \sigma_d^2(\ell)| = o_p(1)$.*

Proof. *Recall that for a given $\ell \in \mathcal{L}$,*

$$\hat{\sigma}_d^2(\ell) = \frac{n}{B}\sum_{b=1}^{B}\left(\hat{\nu}_d^b(\ell) - \overline{\hat{\nu}_d^b}(\ell)\right)^2, \text{ where } \overline{\hat{\nu}}_d^b(\ell) = \frac{1}{B}\sum_{b=1}^{B}\hat{\nu}_d^b(\ell).$$

It can be written as

$$\hat{\sigma}_d^2(\ell) = \frac{n}{B}\sum_{b=1}^{B}\left(\hat{\nu}_d^b(\ell) - \hat{\nu}_d(\ell)\right)^2 + 2\frac{n}{B}\sum_{b=1}^{B}\left(\hat{\nu}_d^b(\ell) - \hat{\nu}_d(\ell)\right)\left(\hat{\nu}_d(\ell) - \overline{\hat{\nu}_d^b}(\ell)\right) + \frac{n}{B}\sum_{b=1}^{B}\left(\hat{\nu}_d(\ell) - \overline{\hat{\nu}_d^b}(\ell)\right)^2 \quad \text{(D.10)}$$

We first consider the second term on the right-hand side of Equation (D.10). Let $\bar{W}_i = \frac{1}{B}\sum_{b=1}^{B} W_i^b$, Using Lemma D.3, we know that for a given $b = 1, 2, \cdots, B$, and uniformly over

$\ell \in \mathcal{L}$,

$$\hat{\nu}_d^b(\ell) - \hat{\nu}_d(\ell) = \frac{1}{n}\sum_{i=1}^{n}(W_i^b - 1)\phi_{\nu_d,i}(\ell,\theta_0) + o_p(1).$$

So it can be written as

$$\begin{aligned}
&\frac{n}{B}\sum_{b=1}^{B}\big(\hat{\nu}_d^b(\ell) - \hat{\nu}_d(\ell)\big)\big(\hat{\nu}_d(\ell) - \overline{\hat{\nu}_d^b}(\ell)\big) \\
=&\frac{1}{B}\frac{1}{n}\sum_{b=1}^{B}\big(\sum_{i=1}^{n}(W_i^b - 1)\phi_{\nu_d,i}(\ell,\theta_0)\big)\big(\sum_{i=1}^{n}(\bar{W}_i - 1)\phi_{\nu_d,i}(\ell,\theta_0)\big) + o_p(1) \\
=&\frac{1}{B}\frac{1}{n}\sum_{b=1}^{B}\sum_{i=1}^{n}(W_i^b - 1)(\bar{W}_i - 1)\phi_{\nu_d,i}^2(\ell,\theta_0) + \frac{1}{B}\frac{1}{n}\sum_{b=1}^{B}\sum_{i\neq j}^{n}(W_i^b - 1)(\bar{W}_j - 1)\phi_{\nu_d,i}(\ell,\theta_0)\phi_{\nu_d,j}(\ell,\theta_0) + o_p(1) \\
=&\frac{1}{B^2}\frac{1}{n}\sum_{b=1}^{B}\sum_{i=1}^{n}(W_i^b - 1)^2\phi_{\nu_d,i}^2(\ell,\theta_0) + \frac{1}{B^2}\frac{1}{n}\sum_{b=1}^{B}\sum_{b'\neq b}^{B}\sum_{i=1}^{n}(W_i^b - 1)(W_j^{b'} - 1)\phi_{\nu_d,i}^2(\ell,\theta_0) \\
&+\frac{1}{B}\frac{1}{n}\sum_{b=1}^{B}\sum_{i\neq j}^{n}(W_i^b - 1)(\bar{W}_j - 1)\phi_{\nu_d,i}(\ell,\theta_0)\phi_{\nu_d,j}(\ell,\theta_0) + o_p(1)
\end{aligned}$$

The first right-hand side term is of order $\frac{1}{B}$ and is negligible as $B \to \infty$. The second term on the right-hand side is negligible because $E[(W_i^b - 1)(W_i^{b'} - 1)|(Y,D,Z)] = 0$ as long as $b \neq b'$. The third term on the right-hand side is negligible because $E[(W_i^b - 1)(W_j^b - 1)|(Y,D,Z)] = 0$ as long as $i \neq j$. For similarly reasoning, the third right-hand side term of Equation (D.10) is also negligible as $B \to \infty$.

Now consider the first term on the right-hand side of Equation (D.10). Uniformly over ℓ,

$$\begin{aligned}
&\frac{n}{B}\sum_{b=1}^{B}\big(\hat{\nu}_d^b(\ell) - \hat{\nu}_d(\ell)\big)^2 = \frac{1}{B}\frac{1}{n}\sum_{b=1}^{B}\big(\sum_{i=1}^{n}(W_i^b - 1)\phi_{\nu_d,i}(\ell,\theta_0)\big)^2 + o_p(1) \\
=&\frac{1}{B}\frac{1}{n}\sum_{b=1}^{B}\sum_{i=1}^{n}(W_i^b - 1)^2\phi_{\nu_d,i}^2(\ell,\theta_0) + \frac{1}{B}\frac{1}{n}\sum_{b=1}^{B}\sum_{i=1}^{n}\sum_{j\neq i}^{n}(W_i^b - 1)(W_j^b - 1)\phi_{\nu_d,i}(\ell,\theta_0)\phi_{\nu_d,j}(\ell,\theta_0) + o_p(1).
\end{aligned}$$

Conditioning on the sample, because W_i^b are i.i.d. across b and i, has expectation and variance equal to one, we know $E[(W_i^b - 1)(W_j^b - 1)|(Y,D,Z)] = 0$ and $E[(W_i^b - 1)^2|(Y,D,Z)] = 1$. As $B \to \infty$, the right-hand side converges in probability (with respect to the distribution of $\{W^b\}_{b=1}^{B}$) to $\frac{1}{n}\sum_{i=1}^{n}\phi_{\nu_d,i}^2(\ell,\theta_0) + o_p(1)$, which in turn converges to $\sigma_d^2(\ell)$ uniformly over ℓ as $n \to \infty$. □

E The influence function with covariate case

In this subsection, we derive the influence function for estimating $\nu_d(\ell)$ in the presence of covariates. First, we estimate $\theta_0 \equiv (\theta_{0z}, \theta_{0x})$ by MLE,

$$\begin{aligned}\hat{\theta} &= \underset{\theta\in\Theta}{\operatorname{argmax}} \frac{1}{n}\sum_{i=1}^{n} \log f(Y_i, D_i, Z_i, X_i, \theta) \\ &\equiv \underset{\theta\in\Theta}{\operatorname{argmax}} \frac{1}{n}\sum_{i=1}^{n} D_i \log P(Z_i, X_i, \theta) + (1-D_i)\log(1-P(Z_i, X_i, \theta)). \end{aligned} \tag{E.1}$$

where $P(z, x, \theta)$ is parameterized and depends on (z, x) and $\theta \equiv (\theta_z', \theta_x')'$ through $z'\theta_z + x'\theta_x$. For example, $P(z, x, \theta) = \Phi(z'\theta_z + x'\theta_x)$ for Probit or $P(z, x, \theta) = \frac{exp(z'\theta_z + x'\theta_x)}{1+\exp(z'\theta_z + x'\theta_x)}$ for Logit.

As in Appendix D, we make the following assumptions.

Assumption E.1 *Assuming following conditions hold*

1. *The conditional density of (Y, X, D) given $P(Z, X, \theta_0) = p$, denoted by $f_{Y,X,D|P}(y, x, d|p)$, is Lipschitz continuous in (y, x, p) over the joint support of (Y, X, P) for $d = 0, 1$.*

2. *For all $z \in \mathcal{Z}$ and $x \in \mathcal{X}$, $P(z, x, \theta)$ is continuously differentiable in θ at θ_0 with bounded derivatives.*

Assumption E.2 *The estimator $\hat{\theta}$, $\hat{\beta}_1$, $\hat{\beta}_0$ admits an influence function of the following form,*

$$\sqrt{n}(\hat{\theta} - \theta_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} s_{\theta_0}(D_i, Z_i, X_i, \theta_0) + o_p(1), \tag{E.2}$$

$$\sqrt{n}(\hat{\beta}_1 - \beta_1) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} s_{\beta_1}(D_i, Y_i, Z_i, X_i, \beta_1) + o_p(1), \tag{E.3}$$

$$\sqrt{n}(\hat{\beta}_0 - \beta_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} s_{\beta_0}(D_i, Y_i, Z_i, X_i, \beta_0) + o_p(1), \tag{E.4}$$

where $s_{\theta_0}(\cdot)$, $s_{\beta_1}(\cdot)$ and $s_{\beta_0}(\cdot)$ are measurable, satisfying $E[s_{\theta_0}(D_i, Z_i, X_i, \theta_0)] = 0$,$E[s_{\beta_1}(D_i, Y_i, Z_i, X_i, \beta_1)] = 0$, $E[s_{\beta_0}(D_i, Y_i, Z_i, X_i, \beta_0)] = 0$, $E[\sup_\theta \|s_{\theta_0}(D_i, Z_i, \theta)\|^{2+\delta}] < \infty$, $E[\sup_\beta \|s_{\beta_1}(D_i, Y_i, Z_i, X_i, \beta)\|^{2+\delta}] < \infty$, and $E[\sup_\beta \|s_{\beta_0}(D_i, Y_i, Z_i, X_i, \beta)\|^{2+\delta}] < \infty$ for some $\delta > 0$.

Note that under similar conditions as in Section 4 of Hsu, Liao and Lin (2022, Econometric Reviews), (E.3) and (E.4) would hold. We define the following quantities for generic

$(y, r_y, p, r_p, b, \theta)$:

$$m_1(y, r_y, p, r_p, b, \theta) = \mathbb{E}[D1(y \leq Y - X'b \leq y + r_y)1(p \leq P(Z, X, \theta) \leq p + r_p)],$$
$$m_0(y, r_y, p, r_p, b, \theta) = \mathbb{E}[(D - 1)1(y \leq Y - X'b \leq y + r_y)1(p \leq P(Z, X, \theta) \leq p + r_p)],$$
$$w(p, r_p, \theta) = \mathbb{E}[1(p \leq P(Z, X, \theta) \leq p + r_p)].$$

Let $f_P(p)$ denote the density function of $P(Z, X, \theta_0) \equiv \mathbb{P}(D = 1|X, Z; \theta_0)$. Following the calculation in Hsu and Lieli (2021), we can analogously obtain the derivatives with respect to θ, evaluating at the true parameter values $(\beta_1, \beta_0, \theta_0)$ as

$$\begin{aligned}
&\nabla_\theta m_1(y, r_y, p, r_p, \beta_1, \theta_0)\\
=&\mathbb{E}[D1(y \leq Y - X'\beta_1 \leq y + r_y)|P(Z, X, \theta_0) = p] \cdot f_P(p)\mathbb{E}[\nabla_\theta P(Z, X, \theta_0)|P(Z, X, \theta_0) = p]\\
&- \mathbb{E}[D1(y \leq Y - X'\beta_1 \leq y + r_y)|P(Z, X, \theta_0) = p + r_p] \cdot f_P(p + r_p)\mathbb{E}[\nabla_\theta P(Z, X, \theta_0)|P(Z, X, \theta_0) = p + r_p],\\
&\nabla_\theta m_0(y, r_y, p, r_p, \beta_0, \theta_0)\\
=&\mathbb{E}[(D - 1)1(y \leq Y - X'\beta_0 \leq y + r_y)|P(Z, X, \theta_0) = p] \cdot f_P(p)\mathbb{E}[\nabla_\theta P(Z, X, \theta_0)|P(Z, X, \theta_0) = p]\\
&- \mathbb{E}[(D - 1)1(y \leq Y - X'\beta_0 \leq y + r_y)|P(Z, X, \theta_0) = p + r_p] \cdot f_P(p + r_p)\mathbb{E}[\nabla_\theta P(Z, X, \theta_0)|P(Z, X, \theta_0) = p + r_p],\\
&\nabla_\theta w(p, r_p, \theta_0)\\
=&f_P(p)\mathbb{E}[\nabla_\theta P(Z, X, \theta_0)|P(Z, X, \theta_0) = p] - f_P(p + r_p)\mathbb{E}[\nabla_\theta P(Z, X, \theta_0)|P(Z, X, \theta_0) = p + r_p].
\end{aligned}$$

In addition, let $f_{u_d|zxd}(y|z, x, d)$ denote the conditional density of U_d conditional on $(Z, X, D) = (z, x, d)$, then the derivatives with respect to β, evaluating at the true parameter values $(\beta_1, \beta_0, \theta_0)$ are

$$\begin{aligned}
&\nabla_\beta m_1(y, r_y, p, r_p, \beta_1, \theta_0)\\
=&\mathbb{E}[P(Z, X, \theta_0)(f_{u_1|zxd}(y + r_y|Z, X, 1) - f_{u_1|zxd}(y|Z, X, 1)X \cdot 1(p \leq P(Z, X, \theta) \leq p + r_p)]],\\
&\nabla_\beta m_0(y, r_y, p, r_p, \beta_0, \theta_0)\\
=&\mathbb{E}[(1 - P(Z, X, \theta_0))(f_{u_0|zxd}(y + r_y|Z, X, 0) - f_{u_0|zxd}(y|Z, X, 0)X \cdot 1(p \leq P(Z, X, \theta) \leq p + r_p)]].
\end{aligned}$$

Let the estimators for $m_1(y, r_y, p, r_p, \beta, \theta)$, $m_0(y, r_y, p, r_p, \beta, \theta)$ and $w(p, r_p, \theta)$ be

$$\hat{m}_1(y, r_y, p, r_p, \beta, \theta) = \frac{1}{n}\sum_{i=1}^{n} m_{1,i}(y, r_y, p, r_p, \beta, \theta),$$
$$\hat{m}_0(y, r_y, p, r_p, \beta, \theta) = \frac{1}{n}\sum_{i=1}^{n} m_{0,i}(y, r_y, p, r_p, \beta, \theta),$$
$$\hat{w}(p, r_p, \theta) = \frac{1}{n}\sum_{i=1}^{n} w_i(p, r_p, \theta).$$

where

$$m_{1,i}(y, r_y, p, r_p, \beta, \theta) = D_i 1(y \leq Y_i - X_i\beta \leq y + r_y)1(p \leq P(Z_i, X_i, \theta) \leq p + r_p),$$
$$m_{0,i}(y, r_y, p, r_p, \beta, \theta) = (1 - D_i)1(y \leq Y_i - X_i\beta \leq y + r_y)1(p \leq P(Z_i, X_i, \theta) \leq p + r_p),$$
$$w_i(p, r_p, \theta) = 1(p \leq P(Z_i, X_i, \theta) \leq p + r_p),$$

and

$$\begin{aligned}
&\sqrt{n}(\hat{m}_1(y, r_y, p, r_p, \hat{\beta}_1, \hat{\theta}) - m_1(y, r_y, p, r_p, \beta_1, \theta_0)) \\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n} m_{1,i}(y, r_y, p, r_p, \beta_1, \theta_0) - m_1(y, r_y, p, r_p, \beta_1, \theta_0) + \nabla_\theta m_1(y, r_y, p, r_p, \beta_1, \theta_0) \cdot s(D_i, Z_i, X_i, \theta_0) \\
&\quad + \nabla_\beta m_1(y, r_y, p, r_p, \beta_1, \theta_0) \cdot s_{\beta_1}(D_i, Y_i, Z_i, X_i, \beta_1) + o_p(1) \\
\equiv&\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \phi_{m_1,i}(y, r_y, p, r_p, \beta_1, \theta_0) + o_p(1),
\end{aligned}$$

$$\begin{aligned}
&\sqrt{n}(\hat{m}_0(y, r_y, p, r_p, \hat{\beta}_0, \hat{\theta}) - m_0(y, r_y, p, r_p, \beta_0, \theta_0)) \\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n} m_{0,i}(y, r_y, p, r_p, \beta_0, \theta_0) - m_0(y, r_y, p, r_p, \beta_0, \theta_0) + \nabla_\theta m_0(y, r_y, p, r_p, \beta_0, \theta_0) \cdot s(D_i, Z_i, X_i, \theta_0) \\
&\quad + \nabla_\beta m_0(y, r_y, p, r_p, \beta_0, \theta_0) \cdot s_{\beta_0}(D_i, Y_i, Z_i, X_i, \beta_0) + o_p(1) \\
\equiv&\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \phi_{m_0,i}(y, r_y, p, r_p, \theta_0) + o_p(1),
\end{aligned}$$

$$\begin{aligned}
&\sqrt{n}(\hat{w}(p, r_p, \hat{\theta}) - w(p, r_p, \theta_0)) \\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_i(p, r_p, \theta_0) - w(p, r_p, \theta_0) + \nabla_\theta w(p, r_p, \theta_0) \cdot s(D_i, Z_i, X_i, \theta_0) + o_p(1) \\
\equiv&\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \phi_{w,i}(p, r_p, \theta_0) + o_p(1).
\end{aligned}$$

By Assumption E.1, all elements of $\nabla_\theta m_1(y, r_y, p, r_p, \beta_1, \theta_0)$, $\nabla_\beta m_1(y, r_y, p, r_p, \beta_1, \theta_0)$, $\nabla_\theta m_0(y, r_y, p, r_p, \beta_0, \theta_0)$, and $\nabla_\beta m_0(y, r_y, p, r_p, \beta_0, \theta_0)$, are Lipschitz continuous in y, r_y, p, r_p, and each element of $\nabla_\theta w(p, r_p, \theta_0)$ is Lipschitz continuous in p, r_p. It follows that $\{\phi_{m_1}(y, r_y, p, r_p, \beta_1\theta_0) : (y, r_y, p, r_p) \in [0,1]^4\}$, $\{\phi_{m_0}(y, r_y, p, r_p, \beta_0, \theta_0) : (y, r_y, p, r_p) \in [0,1]^4\}$ and $\{\phi_w(p, r_p, \theta_0) : (p, r_p) \in [0,1]^2\}$ are all VC classes of functions. Finally, let

$$\begin{aligned}
&\nu_1(y, r_y, p_1, p_2, r_p, \beta_1, \theta_0) = m_1(y, r_y, p_2, r_p, \beta_1, \theta_0) \cdot w(p_1, r_p, \theta_0) - m_1(y, r_y, p_1, r_p, \beta_1, \theta_0) \cdot w(p_2, r_p, \theta_0), \\
&\nu_0(y, r_y, p_1, p_2, r_p, \beta_1, \beta_0, \theta_0) = m_0(y, r_y, p_2, r_p, \beta_0, \theta_0) \cdot w(p_1, r_p, \theta_0) - m_0(y, r_y, p_1, r_p, \beta_0, \theta_0) \cdot w(p_2, r_p, \theta_0), \\
&\hat{\nu}_1(y, r_y, p_1, p_2, r_p, \hat{\beta}_1, \hat{\theta}) = \hat{m}_1(y, r_y, p_2, r_p, \hat{\beta}_1, \hat{\theta}) \cdot \hat{w}(p_1, r_p, \hat{\theta}) - \hat{m}_1(y, r_y, p_1, r_p, \hat{\beta}_1, \hat{\theta}) \cdot \hat{w}(p_2, r_p, \hat{\theta}), \\
&\hat{\nu}_0(y, r_y, p_1, p_2, r_p, \hat{\beta}_0, \hat{\theta}) = \hat{m}_0(y, r_y, p_2, r_p, \hat{\beta}_0, \hat{\theta}) \cdot \hat{w}(p_1, r_p, \hat{\theta}) - \hat{m}_0(y, r_y, p_1, r_p, \hat{\beta}_0, \hat{\theta}) \cdot \hat{w}(p_2, r_p, \hat{\theta}).
\end{aligned}$$

Lemma E.1 *Suppose Assumptions 2.1 to 2.4, 3.3, E.1 and E.2 are satisfied, then,*

$$\sqrt{n}(\hat{\nu}_1(y, r_y, p_1, p_2, r_p, \hat{\beta}_1, \hat{\theta}) - \nu_1(y, r_y, p_1, p_2, r_p, \beta_1, \theta_0)) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \phi_{\nu_1,i}(y, r_y, p_1, p_2, r_p, \beta_1, \theta_0) + o_p(1), \tag{E.5}$$

$$\sqrt{n}(\hat{\nu}_0(y, r_y, p_1, p_2, r_p, \hat{\beta}_0, \hat{\theta}) - \nu_0(y, r_y, p_1, p_2, r_p, \beta_0, \theta_0)) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \phi_{\nu_0,i}(y, r_y, p_1, p_2, r_p, \beta_0, \theta_0) + o_p(1), \tag{E.6}$$

where

$$
\begin{aligned}
&\phi_{\nu_1,i}(y, r_y, p_1, p_2, r_p, \beta_1, \theta_0)\\
=&w(p_1, r_p, \theta_0) \cdot \phi_{m_1,i}(y, r_y, p_2, r_p, \beta_1, \theta_0) + m_1(y, r_y, p_2, r_p, \beta_1, \theta_0) \cdot \phi_{w,i}(p_1, r_p, \theta_0)\\
&- w(p_2, r_p, \theta_0) \cdot \phi_{m_1,i}(y, r_y, p_1, r_p, \beta_1, \theta_0) + m_1(y, r_y, p_1, r_p, \beta_1, \theta_0) \cdot \phi_{w,i}(p_2, r_p, \theta_0),\\
&\phi_{\nu_0,i}(y, r_y, p_1, p_2, r_p, \beta_0, \theta_0)\\
=&w(p_1, r_p, \theta_0) \cdot \phi_{m_0,i}(y, r_y, p_2, r_p, \beta_0, \theta_0) + m_0(y, r_y, p_2, r_p, \beta_0, \theta_0) \cdot \phi_{w,i}(p_1, r_p, \theta_0)\\
&- w(p_2, r_p, \theta_0) \cdot \phi_{m_0,i}(y, r_y, p_1, r_p, \beta_0, \theta_0) + m_0(y, r_y, p_1, r_p, \beta_0, \theta_0) \cdot \phi_{w,i}(p_2, r_p, \theta_0).
\end{aligned}
$$

The proofs are similar to those in Appendix D, so we omit the details.