Supplementary Material to “A semiparametric instrumental variable approach to optimal treatment regimes under endogeneity”

Yifan Cui, Eric Tchetgen Tchetgen

S1 Lower and upper bounds of $E [Y_D(L)]$

Lemma S1. Provided that $Z$ is a valid causal IV (as defined by Balke and Pearl (1997)) and outcome $Y$ is binary, we have the following lower and upper bounds of the value function,

$$E\{\omega_1(L)[\mathcal{L}(L) I\{D(L) = 1\} + \mathcal{L}_1(L)] + \omega_{-1}(L)[-U(L) I\{D(L) = -1\} + \mathcal{L}_1(L)]\} \leq E[\mathcal{L}_1(L) I\{D(L) = 1\} + \mathcal{L}_1(L) I\{D(L) = -1\}] \leq E[Y_D(L)],$$

$$E\{\omega_1(L)[U(L) I\{D(L) = 1\} + U_{-1}(L)] + \omega_{-1}(L)[-\mathcal{L}(L) I\{D(L) = -1\} + U_1(L)]\} \geq E[U_1(L) I\{D(L) = 1\} + U_{-1}(L) I\{D(L) = -1\}] \geq E[Y_D(L)],$$

where $\omega_1(l), \omega_{-1}(l) \geq 0$, $\omega_1(l) + \omega_{-1}(l) = 1$ for any $l$,

$$\mathcal{L}(l) = \max \begin{cases}
    p_{-1,-1,l} + p_{1,1,1,l} - 1 \\
    p_{-1,-1,l} + p_{1,1,1,l} - 1 \\
    p_{1,1,l} + p_{-1,-1,l} - 1 \\
    p_{-1,-1,l} + p_{1,1,-1,l} - 1 \\
    2p_{-1,-1,l} + p_{1,1,-1,l} + p_{1,-1,1,l} + p_{1,1,1,l} - 2 \\
    p_{-1,-1,l} + 2p_{1,1,-1,l} + p_{-1,-1,l} + p_{1,1,1,l} - 2 \\
    p_{1,1,l} + p_{1,1,-1,l} + 2p_{-1,-1,l} + p_{1,1,1,l} - 2 \\
    p_{-1,-1,l} + p_{1,1,-1,l} + p_{-1,-1,l} + 2p_{1,1,1,l} - 2
\end{cases},$$

1
We consider construction of bounds for the value function with a valid IV. 

\[ U(l) = \min \left\{ \begin{array}{l}
1 - p_{1,-1,-1,t} - p_{-1,1,1,t} \\
1 - p_{-1,1,-1,t} - p_{1,1,1,t} \\
1 - p_{-1,1,-1,t} - p_{1,-1,1,t} \\
1 - p_{1,1,1,t} - p_{1,-1,1,t} \\
2 - 2p_{1,1,-1,t} - p_{1,-1,1,t} - p_{1,-1,1,t} - p_{1,1,1,t} \\
2 - p_{-1,1,-1,t} - 2p_{1,1,-1,t} - p_{-1,-1,1,t} - p_{1,1,1,t} \\
2 - p_{1,-1,-1,t} - p_{1,1,-1,t} - 2p_{1,1,1,t} - p_{1,-1,1,t} \\
2 - p_{1,-1,-1,1,t} - p_{-1,1,-1,t} - p_{-1,1,1,t} - 2p_{1,-1,1,t}
\end{array} \right\}, \]

\[ L_{-1}(l) = \max \left\{ \begin{array}{l}
p_{1,-1,1,t} \\
p_{1,-1,1,t} \\
p_{1,-1,1,t} + p_{1,1,-1,t} - p_{-1,-1,1,t} - p_{1,1,1,t} \\
p_{-1,1,-1,t} + p_{1,-1,1,t} - p_{-1,-1,1,t} - p_{-1,1,1,t}
\end{array} \right\}, \]

\[ U_{-1}(l) = \min \left\{ \begin{array}{l}
1 - p_{-1,-1,1,t} \\
1 - p_{-1,-1,-1,t} \\
p_{-1,1,-1,t} + p_{1,-1,-1,t} + p_{1,-1,1,t} + p_{1,1,1,t} \\
p_{1,-1,-1,t} + p_{1,-1,1,t} + p_{-1,1,1,t} + p_{1,-1,1,t}
\end{array} \right\}, \]

\[ L_{1}(l) = \max \left\{ \begin{array}{l}
p_{1,1,1,t} \\
p_{1,1,1,t} \\
-p_{-1,-1,-1,t} - p_{-1,1,-1,t} + p_{-1,1,1,t} + p_{1,1,1,t} \\
-p_{-1,1,-1,t} - p_{1,-1,1,t} + p_{1,-1,1,t} + p_{1,1,1,t}
\end{array} \right\}, \]

\[ U_{1}(l) = \min \left\{ \begin{array}{l}
1 - p_{-1,1,1,t} \\
1 - p_{-1,1,-1,t} \\
p_{-1,1,-1,t} + p_{1,1,-1,t} + p_{1,-1,1,t} + p_{1,1,1,t} \\
p_{1,-1,1,t} + p_{1,1,-1,t} + p_{-1,-1,1,t} + p_{1,1,1,t}
\end{array} \right\}, \]

and \( p_{y,a,z,t} \) denotes \( \Pr(Y = y, A = a | Z = z, L = l) \).

Proof. We consider construction of bounds for the value function with a valid IV.
Note that

\[ E \left[ Y_D(L) | L \right] = E \left( Y_1 | L \right) I \{ D(L) = 1 \} + E \left( Y_{-1} | L \right) I \{ D(L) = -1 \}, \]

\[ E \left[ Y_D(L) | L \right] = E \left( Y_1 - Y_{-1} | L \right) I \{ D(L) = 1 \} + E \left( Y_{-1} | L \right), \]

\[ E \left[ Y_D(L) | L \right] = E \left( Y_{-1} - Y_1 | L \right) I \{ D(L) = -1 \} + E \left( Y_1 | L \right). \]

By the results from Balke and Pearl (1997), one has the following bounds,

\[
\omega_1(L)[\mathcal{L}(L) I \{ D(L) = 1 \} + \mathcal{L}_{-1}(L)] + \omega_{-1}(L)[-\mathcal{U}(L) I \{ D(L) = -1 \} + \mathcal{L}_1(L)] \\
\leq \mathcal{L}_1(L) I \{ D(L) = 1 \} + \mathcal{L}_{-1}(L) I \{ D(L) = -1 \} \leq E \left[ Y_D(L) | L \right], \tag{S1}
\]

\[
\omega_1(L)[\mathcal{U}(L) I \{ D(L) = 1 \} + \mathcal{U}_{-1}(L)] + \omega_{-1}(L)[-\mathcal{L}(L) I \{ D(L) = -1 \} + \mathcal{U}_1(L)] \\
\geq \mathcal{U}_1(L) I \{ D(L) = 1 \} + \mathcal{U}_{-1}(L) I \{ D(L) = -1 \} \geq E \left[ Y_D(L) | L \right], \tag{S2}
\]

where \( \omega_1(L), \omega_{-1}(L) \geq 0, \omega_1(L) + \omega_{-1}(L) = 1 \), \( \mathcal{L}(L) \) and \( \mathcal{U}(L) \) are lower and upper bounds for \( E (Y_1 - Y_{-1} | L) \) given by Balke and Pearl (1997), while \( \mathcal{L}_{-1}(L), \mathcal{U}_{-1}(L), \mathcal{L}_1(L), \mathcal{U}_1(L) \) are lower and upper bounds for \( E (Y_{-1} | L) \) and \( E (Y_1 | L) \) obtained by Balke and Pearl (1997). Therefore, we complete the proof by taking expectations on both sides of Equations (S1) and (S2).

Because it is not possible to directly maximize \( E[Y_D(L)] \), one may nevertheless proceed by maximizing \( E[\mathcal{L}_1(L)I \{ D(L) = 1 \} + \mathcal{L}_{-1}(L)I \{ D(L) = -1 \}] \) and its lower bounds with user-specified weights \( \omega_1(L) \) and \( \omega_{-1}(L) \). For instance, if \( A = -1 \) refers to placebo, the safest regime might be maximizing \( E[\mathcal{L}(L)I \{ D(L) = 1 \}] \), i.e., assigning only \( A = 1 \) to those for whom \( \mathcal{L}(L) > 0 \). Note that maximizing \( E[\mathcal{L}(L)I \{ D(L) = 1 \}] \) and \( E[-\mathcal{U}(L)I \{ D(L) = -1 \}] \) would recommend two conflicting treatments to patients whose \( (\mathcal{L}(L), \mathcal{U}(L)) \) covers 0, i.e., the treatment decision remains ambiguous to those patients. We caution that the interval estimate for whom \( (\mathcal{L}(L), \mathcal{U}(L)) \) covers 0 might not be further narrowed down to an optimal treatment decision given the overwhelming uncertainty inside IV bounds.
S2 Fisher consistency, excess risk bound and universal consistency of the estimated regime

In this section, we establish Fisher consistency, excess risk bound and universal consistency of the estimated treatment regime. We focus on our first estimator, however, the results hold for the second estimator.

S2.1 Preliminaries

Define the following risk

\[ R(g) \equiv E[WI\{A \neq \text{sign}(g(L))\}], \]

where \( W = AZY/(\delta(L)f(Z|L)) \). The optimal decision function associated with the optimal treatment regime \( D^* \) is defined as \( g^* \equiv \arg\min_{g \in G} R(g) \) and corresponding Bayes risk is \( R^* \equiv R(g^*) \), where \( G \) is the class of all measurable functions.

We also define the \( \phi \)-risk

\[ R_\phi(g) \equiv E[|W|\phi(\text{sign}(W)Ag(L))], \]

where \( \phi \) is the hinge loss function. The minimal \( \phi \)-risk \( R^*_\phi \equiv \inf_{g \in G} R_\phi(g) \) and \( g^*_\phi \equiv \arg\min_{g \in G} R_\phi(g) \).

S2.2 Fisher consistency and excess risk bound

Note that Theorem 2.1 of Zhou and Kosorok (2017) shows that Fisher consistency holds if and only if \( \phi'(0) \) exists and \( \phi'(0) < 0 \) provided that the loss function \( \phi \) is convex. The hinge loss function \( \phi \) satisfies this condition which essentially implies the following Fisher consistency.

**Lemma S1.** Under Assumptions 2-7, \( R^* = R(g^*_\phi) \).
The following theorem states that for any measurable decision function $g$, the excess risk under 0-1 loss is bounded by the excess $\phi$-risk.

**Lemma S2.** Under Assumptions 2-7, for any measurable decision function $g$, we have that

$$R(g) - R^* \leq R_\phi(g) - R_{\phi}^*.$$  

The proof follows from Theorem 2.2 of Zhou and Kosorok (2017). Lemma S2 implies that the loss of the value function due to the individualized treatment regime $D$ associated with the decision function $g$ can be bounded by the excess risk under the hinge loss. This excess bound also serves as an intermediate step for investigating the universal consistency of the estimated treatment regime.

### S2.3 Consistency of the estimated treatment regime

In this section, we establish the universal consistency of the estimated treatment regime with a universal kernel (e.g., Gaussian kernel). Estimation error has two potential sources. The first is from the approximation error associated with $H_K$. The second is the uncertainty in estimated weights.

Before stating the universal consistency result of the estimated treatment regime, we first introduce the concept of universal kernels (Steinwart and Christmann, 2008). A continuous kernel $K$ on a compact metric space $L$ is called universal if its associated reproducing kernel Hilbert space (RKHS) $H_K$ is dense in $C(L)$, where $C(L)$ is the space of all continuous functions on the compact metric space $L$ endowed with the usual supremum norm.

Let $K$ be a universal kernel, and $H_K$ be the associated RKHS. Suppose that $g_\phi^*$ is measurable and bounded, $|g_\phi^*| \leq M_g$, and $|\sqrt{\lambda_n}b_n| \leq M_b$ almost surely for some constants $M_g$ and $M_b$. In addition, we consider a sequence of tuning parameters $\lambda_n \to 0$ and $n\lambda_n \to \infty$ as $n \to \infty$. In order to study the excess risk bound of the $\phi$ loss, we need one additional assumption to bound the weight $W$. 


Assumption S1. The outcome $Y$ is sub-Gaussian. Furthermore, we assume that $M_1 < |\delta(L)|$, $M_2 < f(Z = 1|L) < 1 - M_2$ for some $0 < M_1 < 1$, $0 < M_2 < 1/2$ almost surely.

Then, we have the following result.

Theorem S1. Under Assumptions 2-7, S1, and further assume that

$$
\sup_{l \in \mathcal{L}} |\hat{\delta}(l) - \delta(l)| \overset{p}{\to} 0, \quad \text{and} \quad \sup_{l \in \mathcal{L}} |\hat{f}(z = 1|l) - f(z = 1|l)| \overset{p}{\to} 0,
$$

as $n \to \infty$, we have the following convergence in probability,

$$
\lim_{n \to \infty} R(g_n) = R^*,
$$

where $g_n = h_n + b_n$ is the estimated decision function from

$$
\min_{g = h + b \in \mathcal{H}_K + (1)} \frac{1}{n} \sum_{i=1}^{n} |\hat{w}_i| \phi(\text{sign}(\hat{w}_i) a_i g(l_i)) + \frac{\lambda}{2} ||h||_K^2.
$$

The proof is akin to Zhao et al. (2012); Zhou and Kosorok (2017). The rate of convergence for the estimated treatment regime might also be studied under certain regularity conditions on the distribution of the data, such as the geometric noise assumption proposed by Steinwart and Scovel (2007).

S3 A locally efficient and multiply robust estimation of value function

Consider the nonparametric model $\mathcal{M}_{np}$ which places no restriction on the observed data law. Below, we characterize the efficient influence function of the value functional $\mathcal{V}(\mathcal{D})$ in $\mathcal{M}_{np}$ and therefore characterize the semiparametric efficiency bound for the model, where functional $\mathcal{V}(\mathcal{D})$ is defined in Equation (5).
**Theorem S2.** Under Assumptions 2-6 and 8, the efficient influence function of $\mathcal{V}(\mathcal{D})$ in $\mathcal{M}_{np}$ is given by

$$
\text{EIF}_{\mathcal{V}(\mathcal{D})} = \frac{ZAYI\{A = \mathcal{D}(L)\}}{f(Z|L)\delta(L)} - \left\{ \frac{ZE[AYI\{A = \mathcal{D}(L)\}|Z, L]}{f(Z|L)\delta(L)} \right\} - \mathcal{V}(\mathcal{D}).
$$

Therefore, the semiparametric efficiency bound of $\mathcal{V}(\mathcal{D})$ in $\mathcal{M}_{np}$ equals $E[\text{EIF}^2_{\mathcal{V}(\mathcal{D})}]$.

One cannot be confident that any of the required nuisance models to evaluate the efficient influence function can be correctly specified. It is of interest to develop a multiply robust estimation approach, which is guaranteed to deliver valid inferences about $\mathcal{V}(\mathcal{D})$ provided that some but not necessarily all needed models are correct. When finite-dimensional models are used for nuisance parameters, it is likely that all of them are misspecified leading to lack of consistency. Using infinite-dimensional models can mitigate the problem, however, to achieve asymptotic linearity it is required that all the parts are consistently estimated with sufficiently fast rates (Robins et al., 2017; Chernozhukov et al., 2018).

In order to describe our proposed multiply robust approach, consider the following three semiparametric models that place restrictions on different components of the observed data likelihood while allowing the rest of the likelihood to remain unrestricted.

$\mathcal{M}_1$: models for $f(Z|L)$ and $\delta(L)$ are correct;

$\mathcal{M}_2$: models for $f(Z|L)$ and $\gamma(L) \equiv \sum_z \{ZE[AYI\{A = \mathcal{D}(L)\}|Z = z, L]/\delta(L)\}$ are correct;

$\mathcal{M}_3$: models for $\gamma(L), \gamma'(L) \equiv E[AYI\{A = \mathcal{D}(L)\}|Z = -1, L], \delta(L)$ and

$E[A|Z = -1, L]$ are correct.

Note that by Theorem S4 presented in Section S6, $\gamma(L)$ has the counterfactual
interpretation $E[Y_{D(L)}|L]$, which may help formulate appropriate parametric models for the former. For instance, in case $Y$ is binary, $\gamma(L)$ would need to be specified with an appropriate link function to ensure it falls within the unit interval $(0, 1)$.

Our proposed multiply robust estimator requires estimation of nuisance parameters $f(Z|L)$, $E(A|Z = -1, L)$, $\gamma'(L)$, $\delta(L)$ and $\gamma(L)$. One may use maximum likelihood estimation for $f(Z|L)$, $E(A|Z = -1, L)$, $\gamma'(L)$, denoted as $\hat{f}(Z|L)$, $\hat{E}(A|Z = -1, L)$ and $\hat{\gamma}'(L)$, respectively.

Because $\delta(L)$ and $\gamma(L)$ are shared across submodels of the union model, i.e., $\mathcal{M}_1 \cup \mathcal{M}_3$, $\mathcal{M}_2 \cup \mathcal{M}_3$, respectively, in order to ensure multiple robustness, one must estimate these unknown functions in their respective union model. For estimating $\delta(L)$, we propose to use doubly robust $g$-estimation (Robins, 1994, 2000),

$$\mathbb{P}_n \psi_1(L) \left[ A - \delta(L, \hat{\beta}) \frac{1+Z}{2} - \hat{E}(A|Z = -1, L) \right] \frac{Z}{\hat{f}(Z|L)} = 0,$$

and we propose the following doubly robust estimating equation to estimate $\gamma(L)$,

$$\mathbb{P}_n \psi_2(L) \left[ AY I\{A = D(L)\} - \hat{\gamma}'(L) - \frac{[A - \hat{E}(A|Z = -1, L)] \gamma(L, \hat{\eta})}{2} \right] \frac{Z}{\hat{f}(Z|L)} = 0,$$

where vector-valued functions $\psi_1(L)$ and $\psi_2(L)$ have the same dimension as $\hat{\beta}$ and $\hat{\eta}$, respectively. Thus, $\delta(L, \hat{\beta})$ is consistent and asymptotically normal in the union model $\mathcal{M}_1 \cup \mathcal{M}_3$, and $\gamma(L, \hat{\eta})$ is consistent and asymptotically normal in the union model $\mathcal{M}_2 \cup \mathcal{M}_3$. Similarly to results in Tchetgen Tchetgen et al. (2018), we have the following theorem.

**Theorem S3.** Under Assumptions 2-6, 8 and standard regularity conditions,

$$\hat{V}_{MR}(D) = \mathbb{P}_n \left[ ZAY I\{A = D(L)\} - \frac{Z \gamma'(L)}{\hat{f}(Z|L)\delta(L, \hat{\beta})} - \frac{Z[A - \hat{E}(A|Z = -1, L)] \gamma(L, \hat{\eta})}{2\hat{f}(Z|L)\delta(L, \hat{\beta})} \right]$$

is a consistent and asymptotically normal estimator of $V(D)$ under the semiparametric union model $\mathcal{M}_{\text{union}} = \mathcal{M}_1 \cup \mathcal{M}_2 \cup \mathcal{M}_3$. Furthermore, $\hat{V}_{MR}(D)$ is semiparametric
**S4 Proof of Theorem S2**

*Proof.* In order to find the efficient influence function for $V(D)$, we need to find the canonical gradient $G$ for $V(D)$ in the nonparametric model $M_{np}$, e.g., find a random variable $G$ with mean 0 and

$$\frac{\partial}{\partial t} V_t(D) \bigg|_{t=0} = E [GS(O; t)] \bigg|_{t=0},$$

where $S(O; t) = \frac{\partial \log f(O; t)}{\partial t}$, and $V_t(D)$ is the value function under a regular parametric submodel in $M_{np}$ indexed by $t$ that includes the true data generating mechanism at $t = 0$ (Van der Vaart, 1998). Note that we have

$$\frac{\partial}{\partial t} V_t(D) \bigg|_{t=0} = E \left[ \frac{ZAYI(A = D(L))}{\delta(L)f(Z|L)} S(O) \right] - E \left[ \frac{ZAYI(A = D(L))}{\delta^2(L)f^2(Z|L)} \left[ \frac{\partial}{\partial t} f_t(Z|L) \delta_t(L) + \frac{\partial}{\partial t} \delta_t(L) f_t(Z|L) \right] \right] \bigg|_{t=0}$$

$$= (I) - (II) - (III).$$
The second term

\[(II) = E \left[ \frac{ZAYI(A = D(L))}{\delta^2(L)f^2(Z|L)} \frac{\partial}{\partial t} f_t(Z|L) \delta(L) \right] \bigg|_{t=0} \]

\[= E \left[ ZAYI(A = D(L)) \frac{\delta(L)f(Z|L)}{\delta(L)f(Z|L)} S(Z|L) \right] \]

\[= E \left[ E \left[ \frac{ZAYI(A = D(L))}{\delta(L)f(Z|L)} |Z, L| \right] S(Z|L) \right] \]

\[= E \left[ \left\{ E \left[ \frac{ZAYI(A = D(L))}{\delta(L)f(Z|L)} |Z, L| \right] - E \left( \sum_z \frac{zAYI(A = D(L))}{\delta(L)} \right) |Z = z, L| \right\} S(Z, L) \right] \]

\[= E \left[ \left\{ E \left[ \frac{ZAYI(A = D(L))}{\delta(L)f(Z|L)} |Z, L| \right] - E \left( \sum_z \frac{zAYI(A = D(L))}{\delta(L)} \right) |Z = z, L| \right\} S(O) \right] \]

\[= E \left[ \left\{ \frac{ZE[AYI(A = D(L))|Z, L]}{\delta(L)f(Z|L)} - \sum_z \frac{zE[AYI(A = D(L))|Z = z, L]}{\delta(L)} \right\} S(O) \right]. \]

In order to calculate \((III)\), we need to calculate the term \(\frac{\partial}{\partial t} \delta_t(L)\). To do so, we first calculate \(\frac{\partial}{\partial t} E_t[A|Z = z, L] |_{t=0}\). Note that

\[\frac{\partial}{\partial t} E_t[A|Z = z, L] \]

\[= \frac{\partial}{\partial t} \int a f_t(a|Z = z, L) da \]

\[= \int a \frac{\partial f_t(a|Z = z, L)}{f_t(a|Z = z, L)} f_t(a|Z = z, L) da \]

\[= E[A \frac{\partial f_t(A|Z = z, L)}{f_t(A|Z = z, L)} |Z = z, L], \]
\[
\frac{\partial}{\partial t} E_t[A|Z = z, L] \bigg|_{t=0}
= E[AS(A|Z = z, L)Z = z, L]
= E[(A - E[A|Z = z, L])S(A|Z = z, L)Z = z, L]
= E[(A - E[A|Z = z, L])S(A, Z = z|L)Z = z, L].
\]

Then
\[
2 \frac{\partial}{\partial t} \delta_t(L) \bigg|_{t=0}
= E[(A - E[A|Z = 1, L])S(A, Z = 1|L)L] - E[(A - E[A|Z = -1, L])S(A, Z = -1|L)L]
= E[\frac{Z}{f(Z|L)}(A - E[A|Z, L])S(A, Z|L)L]
= E[\frac{Z}{f(Z|L)}(A - E[A|Z, L])S(A, Z|L)L] + E[\frac{Z}{f(Z|L)}(A - E[A|Z, L])S(L|L)]
= E[\frac{Z}{f(Z|L)}(A - E[A|Z, L])S(A, Z|L)L] + E[\frac{Z}{f(Z|L)}(A - E[A|Z, L])S(Y|A, Z, L|L)]
= E[\frac{Z}{f(Z|L)}(A - E[A|Z, L])S(A, Y, Z, L|L)].
\]

It follows that
\[
(III) = E \left[ \frac{ZAYI(A = D(L))}{\delta^2(L)f^2(Z|L)} \frac{\partial}{\partial t} \delta_t(L)f(Z|L) \right] \bigg|_{t=0}
= \frac{1}{2} E \left[ E\left[ \frac{ZAYI(A = D(L))}{\delta^2(L)f(Z|L)} \right] E\left[ \frac{Z}{f(Z|L)}(A - E[A|Z, L])S(\mathcal{O}|L) \right] \right]
= \frac{1}{2} E \left[ \sum_z zE[AYI(A = D(L))|Z = z, L] \frac{Z}{\delta^2(L)f(Z|L)}(A - E[A|Z, L])S(\mathcal{O}|L) \right]
= \frac{1}{2} E \left[ \sum_z zE[AYI(A = D(L))|Z = z, L] \frac{Z}{\delta^2(L)f(Z|L)}(A - E[A|Z, L])S(\mathcal{O}) \right].
\]
Thus, $\frac{\partial}{\partial t} V_t(D)|_{t=0}$ further equals to

$$E\left[\frac{ZAYI(A = D(L))}{\delta(L)f(Z|L)} S(O)\right]$$

$$- E \left[ \left\{ \frac{ZE[AYI(A = D(L))|Z,L]}{\delta(L)f(Z|L)} - \sum_z \frac{zE[AYI(A = D(L))|Z = z,L]}{\delta(L)} \right\} S(O) \right]$$

$$- \frac{1}{2} E \left[ \sum_z zE[AYI(A = D(L))|Z = z,L] \frac{Z}{\delta^2(L)} f(Z|L) (A - E[A|Z,L]) S(O) \right].$$

So the canonical gradient in $M_{np}$ is

$$\frac{ZAYI(A = D(L))}{f(Z|L)\delta(L)} - \left\{ \frac{ZE[AYI(A = D(L))|Z,L]}{f(Z|L)\delta(L)} - \sum_z \frac{zE[AYI(A = D(L))|Z = z,L]}{\delta(L)} \right\} - V(D).$$

As shown by Bickel et al. (1993); Newey (1990); Van der Vaart (1998), the canonical gradient in $M_{np}$ equals to the efficient influence function evaluated at observed data $O$, which completes our proof.

$$\square$$

S5 Proof of Theorem S3

Proof. We start from multiply robustness. Under some regularity conditions (White, 1982), the nuisance estimators $\delta(L, \hat{\beta})$, $\gamma(L, \hat{\eta})$, $\hat{\gamma}'(L)$, $\hat{f}(Z|L)$, $\hat{E}(A|Z = -1, L)$, converge in probability to $\delta(L, \beta^*)$, $\gamma(L, \eta^*)$, $\gamma''(L)$, $f^*(Z|L)$, $E^*(A|Z = -1, L)$. It suffices to show that in the union model $M_{union}$,

$$E \left[ \frac{ZAYI(A = D(L))}{f^*(Z|L)\delta(L, \beta^*)} - \left\{ \frac{Z\gamma''(L)}{f^*(Z|L)\delta(L, \beta^*)} - \gamma(L, \eta^*) \right\} - \frac{Z[A - E^*(A|Z = -1, L)]}{2f^*(Z|L)\delta(L, \beta^*)} \gamma(L, \eta^*) \right] = E[Y_{D(L)}].$$
We first note that
\[
E \left[ \frac{Z A Y I (A = D(L))}{f^*(Z|L)\delta(L, \beta^*)} \right] - \left\{ \frac{Z \gamma^*(L)}{f^*(Z|L)\delta(L, \beta^*)} - \gamma(L, \eta^*) \right\} = 0 \quad (S4)
\]

\[
= E \left[ \frac{Z A Y I (A = D(L))}{f^*(Z|L)\delta(L, \beta^*)} \right] - \left\{ \frac{Z[A - E^*(A|Z = -1, L)]}{2f^*(Z|L)\delta(L, \beta^*)} \gamma(L, \eta^*) \right\} \quad (S5)
\]

If $\mathcal{M}_1$ is correctly specified, Equation (S5) equals to
\[
E \left[ \frac{Z A Y I (A = D(L))}{f^*(Z|L)\delta(L, \beta^*)} \right] - \left\{ \frac{Z[A - E^*(A|Z = -1, L)]}{2f^*(Z|L)\delta(L, \beta^*)} \gamma(L, \eta^*) \right\}
\]

If $\mathcal{M}_2$ is correctly specified, Equation (S5) equals to
\[
E \left[ \frac{Z A Y I (A = D(L))}{f^*(Z|L)\delta(L, \beta^*)} \right] - \left\{ \frac{Z[A - E^*(A|Z = -1, L)]}{2f^*(Z|L)\delta(L, \beta^*)} \gamma(L, \eta^*) \right\}
\]

\[
= E \left[ \frac{Z A Y I (A = D(L))}{f^*(Z|L)\delta(L, \beta^*)} \right] - \left\{ \frac{Z[A - E^*(A|Z = -1, L)]}{2f^*(Z|L)\delta(L, \beta^*)} \gamma(L) \right\}
\]

\[
= E \left[ \frac{Z A Y I (A = D(L))}{f^*(Z|L)\delta(L, \beta^*)} \right] - \left\{ \frac{Z[A - E^*(A|Z = -1, L)]}{2f^*(Z|L)\delta(L, \beta^*)} \gamma(L) \right\}
\]

\[
= E[Y_D(L)].
\]
Finally, if $M_3$ is correctly specified, notice that

$$E[AYI(A = D(L))|Z, L] = \gamma(L)\delta(L)\frac{1+Z}{2} + \gamma'(L),$$

so we have Equation (S5) equals to

$$E\left[\frac{ZAYI(A = D(L))}{f^*(Z|L)\delta(L)} - \left\{ \frac{Z[\gamma(L)(1 + Z)\delta(L) + 2\gamma'(L)]}{2f^*(Z|L)\delta(L)} - \gamma(L) \right\} \right]$$

$$= E\left[\frac{ZE[AYI(A = D(L))|Z, L]}{f^*(Z|L)\delta(L)} - \frac{Z[\gamma(L)(1 + Z)\delta(L) + 2\gamma'(L)]}{2f^*(Z|L)\delta(L)} + \gamma(L) \right]$$

$$= E[Y_D(L)].$$

As shown by Robins and Rotnitzky (2011), the efficient influence function in $M_{\text{union}}$ coincides with the efficient influence function in $M_{np}$, i.e., $EIF_{V(D)}$. Thus, in order to show asymptotic normality and local efficiency, we need to derive the influence function of $\hat{V}_{MR}(D)$. Let $\eta$ be a vector including all nuisance parameters. From a standard Taylor expansion of $EIF_{V(D)}$ around $V(D)$ and $\eta$, following uniform weak law of large number (Newey and McFadden, 1994) under some regularity conditions, we have

$$\sqrt{n}(\hat{V}_{MR}(D) - V(D)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} EIF_{V(D)}(O_i) + E\left(\frac{\partial EIF_{V(D)}}{\partial \eta}\right)\sqrt{n}(\hat{\eta} - \eta) + o_p(1).$$

Following the proof of multiple robustness, we have $E(\partial EIF_{V(D)}/\partial \eta) = 0$ under the intersection model $M_{\text{int}}$. This completes our proof. \qed

S6  Theorem S4 and its proof

Theorem S4. Under Assumptions 2-6 and 8, we have that $\gamma(L) = E[Y_D(L)|L]$. 


Proof. Note that

\[
\gamma(L) = \sum_z \frac{zE[AYI(A = D(L))|Z = z, L]}{\delta(L)} \\
= \frac{E[AYI(A = D(L))|Z = 1, L]}{\delta(L)} - \frac{E[AYI(A = D(L))|Z = -1, L]}{\delta(L)} \\
= \sum_a \frac{E[aY_aI(A = D(L))I(A = a)|Z = 1, L]}{\delta(L)} \\
- \sum_a \frac{E[aY_aI(A = D(L))I(A = a)|Z = -1, L]}{\delta(L)} \\
= \sum_a \frac{E[aE[Y_a|L, U]I(D(L) = a) Pr(A = a|Z = 1, L, U)]}{\delta(L)} \\
- \sum_a \frac{E[aE[Y_a|L, U]I(D(L) = a) Pr(A = a|Z = -1, L, U)]}{\delta(L)} \\
= \frac{E[E[Y_1|L, U]I(D(L) = 1) Pr(A = 1|Z = 1, L, U)]}{\delta(L)} \\
- \frac{E[E[Y_1|L, U]I(D(L) = -1) Pr(A = -1|Z = 1, L, U)]}{\delta(L)} \\
- \frac{E[E[Y_1|L, U]I(D(L) = 1) Pr(A = 1|Z = -1, L, U)]}{\delta(L)} \\
+ \frac{E[E[Y_1|L, U]I(D(L) = -1) Pr(A = -1|Z = -1, L, U)]}{\delta(L)} \\
= E[Y_1|L]I(D(L) = 1) + E[Y_{-1}|L]I(D(L) = -1) \\
= E[Y_{D(L)}|L].
\]

\[\square\]

S7 Additional simulations

S7.1 Sensitivity analysis on the strength of the IV

In this section, we conducted the sensitivity analysis on the strength of the IV. Treatment \(A\) was generated under a logistic regression with success probabilities,

\[
Pr(A = 1|Z, L, U) = \expit\{2L^{(1)} + 3Z - 0.5U\},
\]

Pr(A = 1|Z, L, U) = \expit\{2L^{(1)} + 3Z - 0.5U\},
and

$$\Pr(A = 1|Z, L, U) = \expit\{2L^{(1)} + 2Z - 0.5U\},$$

respectively, with $Z$ a Bernoulli event with probability $1/2$, and $U$ from a bridge distribution with parameter $\phi = 1/2$. The additive associations between $A$ and $Z$ defined as $\Pr(A = 1|Z = 1) - \Pr(A = 1|Z = -1)$ are approximately equal to 0.74 and 0.54 for these two scenarios, respectively. The additive association between $A$ and $Z$ is approximately equal to 0.66 for the scenario considered in the article. Tables S1 and S3 report the mean and standard deviation of value functions evaluated at estimated optimal regimes in test samples for two scenarios, respectively. Tables S2 and S4 report the mean and standard deviation of correct classification rates in test samples for two scenarios, respectively. As can be seen from tables, higher compliance rate generally leads to a lower variance of the estimated regime in terms of both value functions and correct classification rates.

| Table S1: Simulation results: Mean $\times 10^{-2}$ (sd $\times 10^{-2}$) of value functions |
|-------------------------------------------------|-----|-----|-----|-----|
| Kernel  | OWL  | RWL  | IV-IW | IV-MR |
| 1       | Linear | 94.5 (4.7) | 96.5 (2.9) | 96.6 (3.7) | 97.6 (2.1) |
|         | Gaussian | 87.7 (7.0) | 95.0 (2.9) | 90.8 (8.2) | 94.6 (4.8) |
| 2       | Linear | 38.1 (18.9) | 40.1 (19.3) | 93.0 (6.6) | 93.8 (6.0) |
|         | Gaussian | 64.1 (9.0) | 65.0 (9.2) | 84.3 (10.4) | 87.7 (9.1) |
| 3       | Linear | 354.0 (6.1) | 358.6 (3.0) | 359.5 (2.7) | 359.6 (2.2) |
|         | Gaussian | 302.9 (33.3) | 356.4 (4.1) | 320.9 (33.0) | 357.0 (5.8) |
| 4       | Linear | 275.4 (5.4) | 275.5 (5.3) | 351.6 (8.4) | 351.7 (8.2) |
|         | Gaussian | 282.4 (14.2) | 302.4 (14.6) | 314.0 (32.9) | 337.8 (19.1) |

OWL: outcome weighted learning; RWL: residual weighted learning; IV-IW: the proposed estimator with weight $\hat{W}^{(1)}$ or $\hat{W}^{(2)}$; IV-MR: the proposed multiply robust estimator with weight $\hat{W}^{(1)}_{MR}$ or $\hat{W}^{(2)}_{MR}$. The empirical optimal value functions are 0.998, 0.995, 3.636, 3.630 for four scenarios, respectively.
Table S2: Simulation results: Mean ×10^{-2} (sd ×10^{-2}) of correct classification rates

| Kernel | OWL   | RWL   | IV-IW  | IV-MR  |
|--------|-------|-------|--------|--------|
| 1      | Linear| 86.1 (5.6) | 89.0 (4.1) | 89.3 (4.7) | 91.0 (3.3) |
|        | Gaussian| 79.2 (6.8) | 86.9 (3.9) | 83.0 (8.1) | 86.9 (5.4) |
| 2      | Linear| 43.3 (11.0) | 44.2 (10.9) | 84.9 (7.2) | 85.8 (6.7) |
|        | Gaussian| 59.4 (6.2) | 60.3 (6.6) | 76.7 (9.2) | 79.8 (8.2) |
| 3      | Linear| 86.9 (4.4) | 89.1 (4.0) | 91.3 (3.0) | 90.6 (2.9) |
|        | Gaussian| 72.4 (9.4) | 87.6 (4.3) | 77.8 (11.2) | 89.3 (4.2) |
| 4      | Linear| 37.3 (2.2) | 37.4 (2.3) | 84.5 (5.9) | 84.3 (6.1) |
|        | Gaussian| 47.6 (7.1) | 51.0 (8.3) | 71.0 (11.2) | 77.5 (9.5) |

OWL: outcome weighted learning; RWL: residual weighted learning; IV-IW: the proposed estimator with weight \( \hat{W}^{(1)} \) or \( \hat{W}^{(2)} \); IV-MR: the proposed multiply robust estimator with weight \( \hat{W}_{MR}^{(1)} \) or \( \hat{W}_{MR}^{(2)} \).

Table S3: Simulation results: Mean ×10^{-2} (sd ×10^{-2}) of value functions

| Kernel | OWL   | RWL   | IV-IW  | IV-MR  |
|--------|-------|-------|--------|--------|
| 1      | Linear| 96.8 (1.9) | 97.7 (1.4) | 93.7 (6.5) | 95.3 (4.9) |
|        | Gaussian| 87.9 (8.2) | 96.0 (2.7) | 86.3 (10.6) | 90.5 (8.0) |
| 2      | Linear| 35.0 (17.1) | 38.3 (18.7) | 88.8 (8.9) | 89.5 (9.0) |
|        | Gaussian| 59.0 (10.1) | 58.8 (11.1) | 77.6 (12.6) | 81.1 (11.4) |
| 3      | Linear| 358.0 (3.6) | 360.2 (2.1) | 356.8 (5.4) | 357.4 (4.2) |
|        | Gaussian| 288.3 (33.5) | 357.5 (4.5) | 305.7 (34.6) | 350.9 (13.0) |
| 4      | Linear| 274.6 (0.0) | 275.0 (3.6) | 343.5 (20.0) | 345.1 (16.9) |
|        | Gaussian| 279.0 (11.2) | 293.5 (13.2) | 297.9 (32.2) | 320.0 (25.1) |

OWL: outcome weighted learning; RWL: residual weighted learning; IV-IW: the proposed estimator with weight \( \hat{W}^{(1)} \) or \( \hat{W}^{(2)} \); IV-MR: the proposed multiply robust estimator with weight \( \hat{W}_{MR}^{(1)} \) or \( \hat{W}_{MR}^{(2)} \); The empirical optimal value functions are 0.998, 0.995, 3.636, 3.630 for four scenarios, respectively.
Table S4: Simulation results: Mean \(\times 10^{-2}\) (sd \(\times 10^{-2}\)) of correct classification rates

| Kernel | OWL     | RWL     | IV-IW   | IV-MR   |
|--------|---------|---------|---------|---------|
| Linear | 89.4 (3.4) | 91.1 (3.0) | 85.3 (7.0) | 87.3 (5.6) |
| Gaussian | 79.8 (8.0) | 88.4 (3.8) | 78.5 (9.4) | 82.5 (7.6) |
| Linear | 41.4 (9.7) | 43.0 (10.2) | 80.2 (8.7) | 80.9 (8.8) |
| Gaussian | 55.3 (6.4) | 55.5 (7.2) | 71.1 (10.1) | 73.8 (9.5) |
| Linear | 90.0 (3.3) | 91.0 (3.3) | 88.9 (4.1) | 88.7 (3.6) |
| Gaussian | 69.2 (10.4) | 88.9 (4.6) | 71.6 (11.8) | 85.9 (6.5) |
| Linear | 37.0 (0.0) | 37.2 (1.5) | 80.4 (9.7) | 80.8 (8.9) |
| Gaussian | 42.8 (5.4) | 45.4 (6.5) | 65.5 (10.7) | 70.7 (10.0) |

OWL: outcome weighted learning; RWL: residual weighted learning; IV-IW: the proposed estimator with weight \(\hat{W}^{(1)}\) or \(\hat{W}^{(2)}\); IV-MR: the proposed multiply robust estimator with weight \(\hat{W}_{MR}^{(1)}\) or \(\hat{W}_{MR}^{(2)}\).

S7.2 Additional simulations with different sample sizes

Tables S5-S6 and S7-S8 report the simulation results with sample sizes 250 and 1000, respectively.

Table S5: Simulation results: Mean \(\times 10^{-2}\) (sd \(\times 10^{-2}\)) of value functions (sample size \(n = 250\))

| Kernel | OWL     | RWL     | IV-IW   | IV-MR   |
|--------|---------|---------|---------|---------|
| Linear | 92.0 (6.5) | 95.5 (3.2) | 93.0 (6.2) | 94.3 (5.4) |
| Gaussian | 82.7 (8.7) | 92.8 (4.8) | 81.5 (11.4) | 87.9 (9.8) |
| Linear | 48.5 (22.7) | 51.7 (22.3) | 87.0 (9.6) | 87.4 (9.5) |
| Gaussian | 61.6 (11.1) | 62.0 (12.9) | 75.7 (12.1) | 77.4 (12.3) |
| Linear | 349.5 (11.8) | 357.4 (4.0) | 354.7 (8.4) | 355.9 (5.5) |
| Gaussian | 282.7 (35.4) | 354.0 (6.4) | 298.7 (35.7) | 346.1 (17.2) |
| Linear | 278.5 (12.7) | 280.8 (15.1) | 339.7 (21.1) | 340.3 (19.3) |
| Gaussian | 276.5 (17.7) | 299.3 (15.4) | 295.6 (33.7) | 317.8 (25.3) |

OWL: outcome weighted learning; RWL: residual weighted learning; IV-IW: the proposed estimator with weight \(\hat{W}^{(1)}\) or \(\hat{W}^{(2)}\); IV-MR: the proposed multiply robust estimator with weight \(\hat{W}_{MR}^{(1)}\) or \(\hat{W}_{MR}^{(2)}\); The empirical optimal value functions are 0.998, 0.995, 3.636, 3.630 for four scenarios, respectively.
Table S6: Simulation results: Mean $\times 10^{-2}$ (sd $\times 10^{-2}$) of correct classification rates (sample size $n = 250$)

| Kernel | OWL    | RWL    | IV-IW   | IV-MR   |
|--------|--------|--------|---------|---------|
| 1 Linear | 83.2 (7.0) | 87.4 (4.6) | 84.3 (6.7) | 86.0 (6.1) |
| Gaussian | 74.5 (8.1) | 84.4 (5.4) | 74.2 (9.8) | 80.1 (8.9) |
| 2 Linear | 50.0 (14.1) | 51.3 (13.4) | 78.3 (8.8) | 78.7 (8.7) |
| Gaussian | 58.0 (7.4) | 58.5 (8.6) | 69.4 (9.6) | 70.7 (9.8) |
| 3 Linear | 84.8 (5.3) | 88.2 (4.3) | 87.6 (5.0) | 87.6 (4.5) |
| Gaussian | 67.3 (9.8) | 85.8 (5.5) | 70.2 (11.9) | 83.3 (8.4) |
| 4 Linear | 39.0 (6.2) | 39.9 (7.2) | 78.7 (9.6) | 78.6 (9.4) |
| Gaussian | 46.6 (6.9) | 49.9 (8.6) | 65.1 (10.8) | 69.8 (10.6) |

OWL: outcome weighted learning; RWL: residual weighted learning; IV-IW: the proposed estimator with weight $\hat{W}^{(1)}$ or $\hat{W}^{(2)}$; IV-MR: the proposed multiply robust estimator with weight $\hat{W}_{MR}^{(1)}$ or $\hat{W}_{MR}^{(2)}$.

Table S7: Simulation results: Mean $\times 10^{-2}$ (sd $\times 10^{-2}$) of value functions (sample size $n = 1000$)

| Kernel | OWL    | RWL    | IV-IW   | IV-MR   |
|--------|--------|--------|---------|---------|
| 1 Linear | 97.7 (1.4) | 98.4 (0.9) | 97.6 (2.2) | 98.2 (1.1) |
| Gaussian | 91.4 (6.0) | 97.2 (1.6) | 93.4 (5.3) | 96.1 (2.6) |
| 2 Linear | 29.1 (8.6) | 31.9 (12.8) | 95.0 (5.2) | 95.7 (4.5) |
| Gaussian | 62.5 (6.6) | 62.1 (7.3) | 88.7 (8.1) | 90.8 (6.1) |
| 3 Linear | 359.4 (2.5) | 360.8 (1.5) | 360.4 (1.9) | 360.1 (1.7) |
| Gaussian | 305.2 (29.8) | 359.2 (2.8) | 331.8 (29.7) | 359.0 (3.8) |
| 4 Linear | 274.6 (0.0) | 274.7 (1.3) | 355.1 (4.2) | 354.7 (3.9) |
| Gaussian | 283.3 (10.4) | 299.0 (11.5) | 325.9 (29.2) | 345.8 (13.1) |

OWL: outcome weighted learning; RWL: residual weighted learning; IV-IW: the proposed estimator with weight $\hat{W}^{(1)}$ or $\hat{W}^{(2)}$; IV-MR: the proposed multiply robust estimator with weight $\hat{W}_{MR}^{(1)}$ or $\hat{W}_{MR}^{(2)}$. The empirical optimal value functions are 0.998, 0.995, 3.636, 3.630 for four scenarios, respectively.
Table S8: Simulation results: Mean $\times 10^{-2}$ (sd $\times 10^{-2}$) of correct classification rates (sample size $n = 1000$)

| Kernel | OWL       | RWL       | IV-IW     | IV-MR     |
|--------|-----------|-----------|-----------|-----------|
| 1 Linear | 91.0 (2.9) | 92.8 (2.4) | 90.9 (3.4) | 92.2 (2.6) |
| Gaussian | 83.4 (6.3) | 90.3 (2.8) | 85.7 (5.5) | 88.8 (3.5) |
| 2 Linear | 38.0 (4.5) | 39.4 (6.5) | 87.4 (5.9) | 88.3 (5.3) |
| Gaussian | 57.6 (4.8) | 57.5 (5.4) | 80.8 (7.4) | 82.6 (6.1) |
| 3 Linear | 91.3 (2.8) | 91.5 (2.7) | 92.2 (2.5) | 91.2 (2.4) |
| Gaussian | 73.8 (9.2) | 90.7 (3.3) | 81.2 (10.5) | 91.3 (3.0) |
| 4 Linear | 37.0 (0.0) | 37.1 (0.7) | 87.0 (3.9) | 86.5 (3.7) |
| Gaussian | 44.1 (5.4) | 47.9 (6.3) | 75.4 (10.7) | 81.5 (7.1) |

OWL: outcome weighted learning; RWL: residual weighted learning; IV-IW: the proposed estimator with weight $\hat{W}^{(1)}$ or $\hat{W}^{(2)}$; IV-MR: the proposed multiply robust estimator with weight $\hat{W}_{MR}^{(1)}$ or $\hat{W}_{MR}^{(2)}$.

References

Balke, A. and Pearl, J. (1997), “Bounds on Treatment Effects from Studies with Imperfect Compliance,” *Journal of the American Statistical Association*, 92, 1171–1176.

Bickel, P., Klaassen, C., Ritov, Y., and Wellner, J. (1993), *Efficient and Adaptive Estimation for Semiparametric Models*, Johns Hopkins series in the mathematical sciences, Springer New York.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018), “Double/debiased machine learning for treatment and structural parameters,” *The Econometrics Journal*, 21, 1–68.

Newey, W. K. (1990), “Semiparametric efficiency bounds,” *Journal of Applied Econometrics*, 5, 99–135.

Newey, W. K. and McFadden, D. (1994), “Chapter 36 Large sample estimation and hypothesis testing,” Elsevier, vol. 4 of *Handbook of Econometrics*, pp. 2111 – 2245.

Robins, J. M. (1994), “Correcting for non-compliance in randomized trials using struc-
tural nested mean models,” *Communications in Statistics - Theory and Methods*, 23, 2379–2412.

— (2000), “Robust estimation in sequentially ignorable missing data and causal inference models.” *Proceedings of the American Statistical Association Section on Bayesian Statistical Science*, 6–10.

Robins, J. M., Li, L., Mukherjee, R., Tchetgen Tchetgen, E., and van der Vaart, A. (2017), “Minimax estimation of a functional on a structured high-dimensional model,” *Ann. Statist.*, 45, 1951–1987.

Robins, J. M. and Rotnitzky, A. (2011), “Comment on “Inference for semiparametric models: Some questions and an answer” by Bickel and Kwon,” *Statistica Sinica*, 11, 920–936.

Steinwart, I. and Christmann, A. (2008), *Support Vector Machines*, Springer Publishing Company, Incorporated, 1st ed.

Steinwart, I. and Scovel, C. (2007), “Fast rates for support vector machines using Gaussian kernels,” *Ann. Statist.*, 35, 575–607.

Tchetgen Tchetgen, E. J., Michael, H., and Cui, Y. (2018), “Technical Report: Marginal Structural Models for Time-varying Endogenous Treatments: A Time-Varying Instrumental Variable Approach,” *arXiv preprint arXiv:1809.05422v1*.

Van der Vaart, A. W. (1998), *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.

White, H. (1982), “Maximum Likelihood Estimation of Misspecified Models,” *Econometrica*, 50, 1–25.

Zhao, Y., Zeng, D., Rush, A. J., and Kosorok, M. R. (2012), “Estimating individualized treatment rules using outcome weighted learning,” *Journal of the American Statistical Association*, 107, 1106–1118.
Zhou, X. and Kosorok, M. (2017), “Augmented Outcome-weighted Learning for Optimal Treatment Regimes,” *arXiv:1711.10654*.  

