Efficient and Robust Estimation of the Generalized LATE Model

Haitian Xie
Guanghua School of Management
Peking University
xht@gsm.pku.edu.cn

September 2, 2023

SUPPLEMENTARY MATERIAL

Appendix A contains the proofs for theorems and propositions stated in the main text. Appendix B contains simulation results. Appendix C studies parameters implicitly defined by overidentifying moment conditions.

A Technical Proofs

In this section, we prove the theorems and propositions stated in the main text. We assume that Assumptions 1 - 3 hold throughout this section.

A.1 Proofs for the identification results

Lemma 1. *$S \perp Z \mid X$ and $t \in \mathcal{T}$, $Y_t \perp T \mid S, X$.*

Proof of Lemma 1. The first statement follows from the definition of S and the fact that Z is independent of the vector $(T_{z_1}, \cdots, T_{z_{N_Z}})$ conditioning on X. For the second statement, T is entirely determined by (S, Z, X). Hence, given S and X, T is independent of Y_t since Z is independent of $(Y_{t_1}, \cdots, Y_{t_{N_T}})$ conditional on X. $\square$

Lemma 2. *For each $t \in \mathcal{T}$ and $k = 1, \cdots, N_Z$, the following identification results hold.*

(i) $\mathbb{P}(S \in \Sigma_{t,k} \mid X) = b_{t,k} P_t(X)$ *a.s.*

(ii) $\mathbb{E}\left[Y_t \mid S \in \Sigma_{t,k}, X\right] = (b_{t,k} Q_t(X))/(b_{t,k} P_t(X))$ *a.s.*

Proof of Lemma 2. This is Theorem T-6 in Heckman and Pinto (2018a). The conditioning is explicitly presented. □

Proof of Theorem 2.1. The first statement follows from applying the law of iterated expectation to Lemma 2(i). For the second statement, we can apply Bayes' rule to Lemma 2 and obtain that

$$\begin{aligned}
\mathbb{E}\left[Y_t \mid S \in \Sigma_{t,k}\right] &= \int \mathbb{E}\left[Y_t \mid S \in \Sigma_{t,k}, X = x\right] f_{X \mid S \in \Sigma_{t,k}}(x) dx \\
&= \int \mathbb{E}\left[Y_t \mid S \in \Sigma_{t,k}, X = x\right] \frac{\mathbb{P}(S \in \Sigma_{t,k} \mid X = x)}{\mathbb{P}(S \in \Sigma_{t,k})} f_X(x) dx \\
&= \mathbb{E}\left[b_{t,k} Q_t(X)\right] / p_{t,k},
\end{aligned}$$

where $f_{X \mid S \in \Sigma_{t,k}}$ denotes the conditional density function of X given type $S \in \Sigma_{t,k}$. □

Proof of Theorem 2.2. By Lemma L-16 of Heckman and Pinto (2018b), we know that under the unordered monotonicity assumption, $B_t[\cdot, i] = B_t[\cdot, i']$ for all $s_i, s_{i'} \in \Sigma_{t,k}$. Thus, the set $\mathcal{Z}_{t,k}$ always exists. For the first statement, we have

$$\begin{aligned}
\mathbb{P}\left(T = t, S \in \Sigma_{t,k}\right) &= \mathbb{P}\left(Z \in \mathcal{Z}_{t,k}, S \in \Sigma_{t,k}\right) \\
&= \mathbb{E}\left[\mathbb{P}\left(Z \in \mathcal{Z}_{t,k}, S \in \Sigma_{t,k} \mid X\right)\right] \\
&= \mathbb{E}\left[\mathbb{P}\left(Z \in \mathcal{Z}_{t,k} \mid X\right) \mathbb{P}\left(S \in \Sigma_{t,k} \mid X\right)\right] \\
&= \mathbb{E}\left[b_{t,k} P_t(X) \pi_{t,k}(X)\right],
\end{aligned}$$

where the second equality follows from the law of iterated expectations and the third equality follows from the fact that $Z \perp S \mid X$ (Lemma 1). For the second statement, notice that

$$\begin{aligned}
\mathbb{P}(T = t, S \in \Sigma_{t,k} \mid X = x) &= \mathbb{P}(T = t \mid S \in \Sigma_{t,k}, X = x) \mathbb{P}(S \in \Sigma_{t,k} \mid X = x) \\
&= \mathbb{P}(Z \in \mathcal{Z}_{t,k} \mid X) \mathbb{P}(S \in \Sigma_{t,k} \mid X = x) \\
&= \pi_{t,k}(X) b_{t,k} P_t(X).
\end{aligned}$$

By Lemma 1, we know that

$$\mathbb{E}\left[Y_t \mid T=t, S \in \Sigma_{t,k}, X=x\right] = \mathbb{E}\left[Y_t \mid S \in \Sigma_{t,k}, X=x\right].$$

Therefore, we can apply Bayes' rule and obtain that

$$\begin{aligned}
&\mathbb{E}\left[Y_t \mid T=t, S \in \Sigma_{t,k}\right] \\
=&\int \mathbb{E}\left[Y_t \mid T=t, S \in \Sigma_{t,k}, X=x\right] f_{X \mid T=t, S \in \Sigma_{t,k}}(x) dx \\
=&\int \mathbb{E}\left[Y_t \mid S \in \Sigma_{t,k}, X=x\right] \frac{P(T=t, S \in \Sigma_{t,k} \mid X=x)}{P(T=t, S \in \Sigma_{t,k})} f_X(x) dx \\
=&\int \frac{b_{t,k} Q_t(X)}{b_{t,k} P_t(X)} \times \frac{\pi_{t,k}(X) b_{t,k} P_t(X)}{q_{t,k}} f_X(x) dx \\
=&\mathbb{E}\left[b_{t,k} Q_t(X) \pi_{t,k}(X)\right] / q_{t,k}.
\end{aligned}$$

$\square$

A.2 Semiparametric efficiency calculations

We follow the method developed by Newey (1990). The likelihood function of the GLATE model can be specified as

$$\mathcal{L}(Y, T, Z, X) = f_X(X) \prod_{z \in \mathcal{Z}} \left(f_z(Y, T \mid X) \pi_z(X)\right)^{\mathbf{1}\{Z=z\}},$$

where $f_z(\cdot, \cdot \mid X)$ denotes the conditional density of Y, T given $Z=z$ and X. In a regular parametric submodel, where the true underlying probability measure P is indexed by θ^o, we use the following notations to represent the score functions:

$$\begin{aligned}
s_z(Y, Z \mid X; \theta) &= \frac{\partial}{\partial \theta} \log\left(f_z(Y, T \mid X; \theta)\right), \\
s_\pi(Z \mid X; \theta) &= \sum_{z \in \mathcal{Z}} \mathbf{1}\{Z=z\} \frac{\partial}{\partial \theta} \log\left(\pi_z(X; \theta)\right), \\
s_X(X; \theta) &= \frac{\partial}{\partial \theta} \log\left(f_X(X; \theta)\right).
\end{aligned}$$

The score in a regular parametric submodel is

$$s_{\theta^o}(Y, T, Z, X) = \sum_{z \in \mathcal{Z}} \mathbf{1}\{Z=z\} s_z\left(Y, T \mid X; \theta^o\right) + s_\pi(Z \mid X; \theta^o) + s_X(X; \theta^o).$$

Hence, the tangent space of the model is

$$\mathscr{S} = \Big\{ s \in L_0^2 : s(Y,T,Z,X) = \sum_{z \in \mathcal{Z}} \mathbf{1}\{Z=z\} s_z \big(Y,T \mid X\big) + s_\pi(Z \mid X) + s_X(X)$$
$$\text{for some } s_z, s_\pi, s_X \text{ such that } \int s_z(y,t \mid X) f_z(y,t \mid X) dy dt = 0, \forall z;$$
$$\sum_{z \in \mathcal{Z}} s_\pi(z \mid X) \pi_z(X) = 0, \text{ and } \int s_X(x) f_X(x) dx = 0 \Big\},$$

where L_0^2 is a subspace of L^2 that contains the mean zero functions.

Proof of Theorem 3.1. We only prove statements (i) and (ii) since (iii) and (iv) are easier cases that can be proved along the way. We start with the first statement. The path-wise differentiability of the parameter $\beta_{t,k}$ can be verified in the following way: in any parametric submodel, we have

$$\begin{aligned}
\frac{\partial}{\partial \theta} \beta_{t,k}(\theta)\Big|_{\theta=\theta^o} &= \frac{\partial}{\partial \theta} (b_{t,k} \mathbb{E}_\theta \left[Q_t(X) \right] / p_{t,k})\big|_{\theta=\theta^o} \\
&= \frac{1}{p_{t,k}} \Big((\partial b_{t,k} \mathbb{E}_\theta \left[Q_t(X) \right] / \partial \theta)|_{\theta=\theta^o} - (b_{t,k} \mathbb{E}_\theta \left[Q_t(X) \right] / p_{t,k}) (\partial p_{t,k} / \partial \theta)|_{\theta=\theta^o} \Big) \\
&= \frac{1}{p_{t,k}} b_{t,k} \left(\frac{\partial}{\partial \theta} \mathbb{E}_\theta \left[Q_t(X) \right] |_{\theta=\theta^o} - \frac{\partial}{\partial \theta} \mathbb{E}_\theta \left[P_t(X) \right] |_{\theta=\theta^o} \beta_{t,k} \right),
\end{aligned}$$

where $\frac{\partial}{\partial \theta} \mathbb{E}_\theta \left[Q_t(X) \right] |_{\theta=\theta^o}$ and $\frac{\partial}{\partial \theta} \mathbb{E}_\theta \left[P_t(X) \right] |_{\theta=\theta^o}$ are $N_Z \times 1$ random vectors whose typical element can be represented respectively by

$$\begin{aligned}
&\int y \mathbf{1}\{\tau = t\} s_z(y, \tau \mid x; \theta^o) f_z(y, \tau \mid x; \theta^o) f_X(x; \theta^o) dy d\tau dx \\
&+ \int y \mathbf{1}\{\tau = t\} s_X(x; \theta^o) f_z(y, \tau \mid x; \theta^o) f_X(x; \theta^o) dy d\tau dx
\end{aligned}$$

and

$$\begin{aligned}
&\int \mathbf{1}\{\tau = t\} s_z(y, \tau \mid x; \theta^o) f_z(y, \tau \mid x; \theta^o) f_X(x; \theta^o) dy d\tau dx \\
&+ \int \mathbf{1}\{\tau = t\} s_X(x; \theta^o) f_z(y, \tau \mid x; \theta^o) f_X(x; \theta^o) dy d\tau dx,
\end{aligned}$$

respectively, for $z \in \mathcal{Z}$. The EIF is characterized by the condition that

$$\frac{\partial}{\partial \theta} \beta_{t,k}(\theta)\Big|_{\theta=\theta^o} = \mathbb{E} \left[\psi_{\beta_{t,k}} s_{\theta^o} \right], \text{ and } \psi_{\beta_{t,k}} \in \mathscr{S}.$$

The expression of $\psi_{\beta_{t,k}}$ given in Equation (1) meets the above requirements. In particular, the correspondence between terms in the EIF and path-wise derivative appears exactly as in Lemma 1 of Hong and Nekipelov (2010b).

For the second statement, the path-wise derivative of $\gamma_{t,k}$ can be computed similarly.

$$\begin{aligned}\frac{\partial}{\partial\theta}\gamma_{t,k}(\theta)\Big|_{\theta=\theta^o} &= \frac{1}{q_{t,k}}b_{t,k}\frac{\partial}{\partial\theta}\mathbb{E}_\theta\left[Q_t(X)\pi_{t,k}(X)\right]\Big|_{\theta=\theta^o} \\ &\quad - \frac{\gamma_{t,k}}{q_{t,k}}b_{t,k}\frac{\partial}{\partial\theta}\mathbb{E}_\theta\left[P_t(X)\pi_{t,k}(X)\right]\Big|_{\theta=\theta^o},\end{aligned}$$

where $\frac{\partial}{\partial\theta}\mathbb{E}_\theta[Q_t(X)\pi_{W_{t,k}}(X)]|_{\theta=\theta^o}$ and $\frac{\partial}{\partial\theta}\mathbb{E}_\theta[P_t(X)\pi_{W_{t,k}}(X)]|_{\theta=\theta^o}$ are $N_Z\times 1$ random vectors whose typical element can be represented by

$$\begin{aligned}&\int y\mathbf{1}\{\tau=t\}s_z(y,\tau\mid x;\theta^o)\pi_{W_{t,k}}(x;\theta^o)f_z(y,\tau\mid x;\theta^o)f_X(x;\theta^o)dyd\tau dx \\ &+\int y\mathbf{1}\{\tau=t\}s_X(x;\theta^o)\pi_{W_{t,k}}(x;\theta^o)f_z(y,\tau\mid x;\theta^o)f_X(x;\theta^o)dyd\tau dx \\ &+\int y\mathbf{1}\{\tau=t\}\left(\frac{\partial}{\partial\theta}\pi_{t,k}(X;\theta)\big|_{\theta=\theta^o}\right)f_z(y,\tau\mid x;\theta^o)f_X(x;\theta^o)dyd\tau dx,\end{aligned}$$

and

$$\begin{aligned}&\int \mathbf{1}\{\tau=t\}s_z(y,\tau\mid x;\theta^o)\pi_{W_{t,k}}(x;\theta^o)f_z(y,\tau\mid x;\theta^o)f_X(x;\theta^o)dyd\tau dx \\ &+\int \mathbf{1}\{\tau=t\}s_X(x;\theta^o)\pi_{W_{t,k}}(x;\theta^o)f_z(y,\tau\mid x;\theta^o)f_X(x;\theta^o)dyd\tau dx \\ &+\int \mathbf{1}\{\tau=t\}\left(\frac{\partial}{\partial\theta}\pi_{t,k}(X;\theta)\big|_{\theta=\theta^o}\right)f_z(y,\tau\mid x;\theta^o)f_X(x;\theta^o)dyd\tau dx,\end{aligned}$$

respectively, for $z\in\mathcal{Z}$. The main difference appears when dealing with the last terms in the above two expressions, which can be matched with terms in the efficient influence function of the following two forms

$$\begin{aligned}&\mathbb{E}\left[Y\mathbf{1}\{T=t\}\mid Z=z,X\right]\left(\mathbf{1}\{Z\in\mathcal{Z}_{t,k}\}-\pi_{t,k}(X)\right),\ \text{and} \\ &\mathbb{E}\left[\mathbf{1}\{T=t\}\mid Z=z,X\right]\left(\mathbf{1}\{Z\in\mathcal{Z}_{t,k}\}-\pi_{t,k}(X)\right).\end{aligned}$$

Take the latter one as an example. Notice that

$$\mathbf{1}\{Z\in\mathcal{Z}_{t,k}\}-\pi_{t,k}(X)=\sum_{z\in\mathcal{Z}_{t,k}}\left(\mathbf{1}\{Z=z\}-\pi_z(X)\right),$$

and

$$\big(\mathbf{1}\{Z=z\}-\pi_z(X)\big)\, s_\pi(Z\mid X;\theta^o)=\frac{\mathbf{1}\{Z=z\}}{\pi_z(X)}\frac{\partial}{\partial\theta}\pi_z(X;\theta)\big|_{\theta=\theta^o}-\pi_z(X)s_\pi(Z\mid X;\theta^o).$$

By the law of iterated expectation, we have

$$\begin{aligned}
&\mathbb{E}\Big[\mathbb{E}\left[\mathbf{1}\{T=t\}\mid Z=z,X\right]\big(\mathbf{1}\{Z=z\}-\pi_z(X)\big)\, s_\pi(Z\mid X;\theta^o)\Big]\\
=&\mathbb{E}\left[\mathbb{E}\left[\mathbf{1}\{T=t\}\mid Z=z,X\right]\mathbb{E}\left[\mathbf{1}\{Z=z\}/\pi_z(X)\mid X\right]\frac{\partial}{\partial\theta}\pi_z(X;\theta)\big|_{\theta=\theta^o}\right]\\
-&\mathbb{E}\Big[\mathbb{E}\left[\mathbf{1}\{T=t\}\mid Z=z,X\right]\pi_z(X)\mathbb{E}\left[s_\pi(Z\mid X;\theta^o)\mid X\right]\Big]\\
=&\mathbb{E}\left[\mathbb{E}\left[\mathbf{1}\{T=t\}\mid Z=z,X\right]\frac{\partial}{\partial\theta}\pi_z(X;\theta)\big|_{\theta=\theta^o}\right]\\
=&\int\mathbf{1}\{\tau=t\}\left(\frac{\partial}{\partial\theta}\pi_z(X;\theta)\big|_{\theta=\theta^o}\right)f_z(y,\tau\mid x;\theta^o)f_X(x;\theta^o)dyd\tau dx.
\end{aligned}$$

□

Proof of Proposition 3.2. This proof is based on Section 4 in Newey (1994). We focus on the case of $\beta_{t,k}$. The other cases are similar. To ease notation, let $h_t=\big(h_{Y,t,Z},h_{t,Z},\pi\big)'$. The estimator $\hat\beta_{t,k}$ is defined by the moment condition

$$\mathbb{E}[M\left(X,\beta_{t,k},h_t\right)]=0,$$

where $M\left(X,\beta_{t,k},h_t\right)$ equals

$$b_{t,k}\left(\frac{h_{Y,t,z_1}(X)}{\pi_{z_1}(X)},\dots,\frac{h_{Y,t,z_{N_Z}}(X)}{\pi_{z_{N_Z}}(X)}\right)'-\beta_{t,k}b_{t,k}\left(\frac{h_{t,z_1}(X)}{\pi_{z_1}(X)},\dots,\frac{h_{t,z_{N_Z}}(X)}{\pi_{z_{N_Z}}(X)}\right)'.$$

We then compute the derivatives of M with respect to the parameters:

$$\begin{aligned}
\mathbb{E}\left[\partial M/\partial\beta_{t,k}\right]&=-b_{t,k}\mathbb{E}\left[P_t(X)\right]=-p^o_{t,k}\\
\partial M/\partial h_{Y,t,z_i}|_{h_t=h_t^o}&=b_{t,k}[i]/\pi^o_{z_i}(X)=\delta_{Y,t,z_i}(X)\\
\partial/\partial h_{t,z_i}M|_{h_t=h_t^o}&=-(\beta_{t,k}b_{t,k}[i])/\pi^o_{z_i}(X)=\delta_{t,z_i}(X)\\
\partial M/\partial\pi_{z_i}|_{h_t=h_t^o}&=-(b_{t,k}[i]Q^o_{t,z_i}(X))/\pi^o_{z_i}(X)+(\beta_{t,k}b_{t,k}[i]P^o_{t,z_i}(X))/\pi^o_{z_i}(X)=\delta_{\pi,z_i}(X),
\end{aligned}$$

where $b_{t,k}[i]$ denotes the ith element of the vector $b_{t,k}$. Define

$$\begin{aligned}\alpha\left(Y,T,Z,X\right) = &\sum_{z\in\mathcal{Z}} \delta_{Y,t,z}(X)\left(\mathbf{1}\{Z=z\}Y\mathbf{1}\{T=t\} - h^o_{Y,t,z}(X)\right)\\ &+\sum_{z\in\mathcal{Z}} \delta_{t,z}(X)\left(\mathbf{1}\{Z=z\}\mathbf{1}\{T=t\} - h^o_{t,z}(X)\right)\\ &+\sum_{z\in\mathcal{Z}} \delta_{\pi,z}(X)\left(\mathbf{1}\{Z=z\} - \pi^o_z(X)\right).\end{aligned}$$

We have

$$\begin{aligned}\alpha\left(Y,T,Z,X\right) = &\, b_{t,k}\zeta(Z,X,\pi^o)\left(\iota(Y\mathbf{1}\{T=t\}) - Q^o_t(X)\right)\\ &- \beta^o_{t,k} b_{t,k}\zeta(Z,X,\pi^o)\left(\iota\mathbf{1}\{T=t\} - P^o_t(X)\right).\end{aligned}$$

Then Newey's (1994) Proposition 4 suggests that the influence function of the estimator $\hat{\beta}_{t,k}$ is $(M+\alpha)/p_{t,k}$ which is equal to the EIF $\psi^{\beta_{t,k}}$.

□

Proof of Theorem 3.3. Based on Proposition 3.2, we only need to show that the estimators are asymptotically linear. By the delta-method argument, if we have two estimators that are asymptotically linear, then their ratio is also asymptotically linear. Therefore, we only need to show that $\hat{p}_{t,k}$ is asymptotically linear, and then the other estimators can be dealt with in the same way. Because $\hat{p}_{t,k}$ is a linear combination of $\frac{1}{n}\sum_{i=1}^n \hat{P}_{t,z}(X_i), z \in \mathcal{Z}$, we only need to work with the latter expression. After adding and subtracting the true first step functions, we obtain that

$$\begin{aligned}\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n \hat{P}_{t,z}(X_i) - \mathbb{E}[P_{t,z}(X)]\right) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n (\hat{P}_{t,z}(X_i) - \mathbb{E}[P_{t,z}(X)])\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \left(\frac{\hat{h}_{t,z}(X_i)}{\hat{\pi}_z(X_i)} - \frac{h_{t,z}(X_i)}{\pi_z(X_i)}\right)\\ &+ \frac{1}{\sqrt{n}}\sum_{i=1}^n \left(\frac{h_{t,z}(X_i)}{\pi_z(X_i)} - \mathbb{E}\left[\frac{h_{t,z}(X_i)}{\pi_z(X_i)}\right]\right).\end{aligned}$$

The second term is already in the form of an influence function, and we only need to focus on the first term. Since π_z is bounded away from zero, we can apply Taylor expansion to the ratio function and obtain that

$$\frac{\hat{h}_{t,z}(X_i)}{\hat{\pi}_z(X_i)} - \frac{h_{t,z}(X_i)}{\pi_z(X_i)} = \frac{1}{\pi_z(X_i)}(\hat{h}_{t,z}(X_i) - h_{t,z}(X_i)) - \frac{h_{t,z}(X_i)}{\pi_z(X_i)^2}(\hat{\pi}_z(X_i) - \pi_z(X_i)) \\ + O_p\left(\|\hat{h}_{t,z} - h_{t,z}\|_\infty^2 \vee \|\hat{\pi}_z - \pi_z\|_\infty^2\right).$$

The above expansion is uniform for $1 \leq i \leq n$ because the remainder term is small with respect to the sup norm. By the standard uniform convergence rate of local polynomial regressions (see, e.g., Masry, 1996, Theorem 6), we know that

$$\|\hat{h}_{t,z} - h_{t,z}\|_\infty^2 \vee \|\hat{\pi}_z - \pi_z\|_\infty^2 = o_p\left(\frac{\log n}{na^{d_X}} + a^{2\lambda}\right) = o_p(n^{-1/2}),$$

where the second equality follows from our assumption on the bandwidth. The remaining task is to show that the following two terms are asymptotically linear:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{1}{\pi_z(X_i)}(\hat{h}_{t,z}(X_i) - h_{t,z}(X_i)), \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{h_{t,z}(X_i)}{\pi_z(X_i)^2}(\hat{\pi}_z(X_i) - \pi_z(X_i)).$$

We focus on the first term because the second term can be analyzed analogously. Notice that $\hat{h}_{t,z}/\pi_z$ and $h_{t,z}/\pi_z$ are smooth functions. By Theorem 2.5.1 and Theorem 2.7.1 in van der Vaart and Wellner (1996), we know that smooth functions form a Donsker class and hence have the following stochastic equicontinuity result:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{1}{\pi_z(X_i)}(\hat{h}_{t,z}(X_i) - h_{t,z}(X_i)) = \sqrt{n}\int \frac{1}{\pi_z(x)}(\hat{h}_{t,z}(x) - h_{t,z}(x))f_X(x)dx + o_p(1).$$

That is, we can replace the empirical measure with the true probability of X, and the error is negligible. Lastly, the by standard Bahadur representation of local polynomial estimators (e.g., Kong et al., 2010, Theorem 1), we know that $\hat{h}_{t,z}(x) - h_{t,z}(x)$ is first-order equivalent to a function of X_i multiplied by $\frac{1}{na^d}\sum_{i=1}^n \varepsilon_i \tilde{K}((x - X_i)/a)$, where $\varepsilon_i = \mathbf{1}\{Z_i = z\}\mathbf{1}\{T_i = t\} - \mathbb{E}[\mathbf{1}\{Z_i = z\}\mathbf{1}\{T_i = t\}|X_i]$ and $\tilde{K}$ is a kernel function that depends on the original kernel K and the order of the local polynomial estimator. By using a change of variables $u = (x - X_i)/a$, we obtain that the integral of the above expression is equal to

$$\int \frac{1}{na^d}\sum_{i=1}^n \varepsilon_i \tilde{K}((x - X_i)/a)f_X(x)/\pi_z(x)dx = \frac{1}{n}\sum_{i=1}^n \int \varepsilon_i f_X(X_i + au)/\pi_z(X_i + au)\tilde{K}(u)du \\ \sim \frac{1}{n}\sum_{i=1}^n \varepsilon_i f_X(X_i)/\pi_z(X_i)\int \tilde{K}(u)du.$$

This verifies that the term $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{\pi_z(X_i)}(\hat{h}_{t,z}(X_i)-h_{t,z}(X_i))$ has an asymptotically linear representation. Similarly, we can also verify that the term $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{h_{t,z}(X_i)}{\pi_z(X_i)^2}(\hat{\pi}_z(X_i)-\pi_z(X_i))$ is asymptotically linear. This proves the result of the theorem. □

A.3 Proofs for the robustness results

Proof of Proposition 4.1. We prove the case for $\psi^{p_{t,k}}$, the other cases can be dealt with analogously. First assume $\pi = \pi^o$, then

$$\mathbb{E}\left[\mathbf{1}\{Z=z\}/\pi_z^o(X)\mid X\right]=1,$$

which implies that $\mathbb{E}\left[\zeta(Z,X,\pi^o)\mid X\right]$ is almost surly equal to the identity matrix $\mathbf{I}$. By the law of total expectations, we have

$$\mathbb{E}\left[\mathbf{1}\{T=t\}\mathbf{1}\{Z=z\}/\pi_z^o(X)\mid X\right]=\mathbb{E}\left[\mathbf{1}\{T=t\}\mid Z=z,X\right]=P_{t,z}^o(X),$$

which implies that $\mathbb{E}\left[\zeta(Z,X,\pi^o)\iota\mathbf{1}\{T=t\}\right]=\mathbb{E}\left[P_t^o(X)\right]$. Therefore,

$$\begin{aligned}
&b_{t,k}\mathbb{E}[\zeta(Z,X,\pi^o)\left(\iota(\mathbf{1}\{T=t\})-P_t(X)\right)+P_t(X)]\\
=&b_{t,k}\mathbb{E}\left[\zeta(Z,X,\pi^o)\iota\mathbf{1}\{T=t\}\right]+b_{t,k}\mathbb{E}\left[(\mathbf{I}-\zeta(Z,X,\pi^o))P_t(X)\right]=b_{t,k}\mathbb{E}\left[P_t^o(X)\right]=p_{t,k}^o.
\end{aligned}$$

Now suppose that $P_t = P_t^o$. Then by the law of total expectation, we have

$$\begin{aligned}
&\mathbb{E}[\mathbf{1}\{Z=z\}(\mathbf{1}\{T=t\}-P_{t,z}^o(X))\mid X]\\
=&\pi_z(X)\mathbb{E}[\mathbb{E}[\mathbf{1}\{T=t\}\mid Z=z,X]-P_{t,z}^o(X)\mid X]=0.
\end{aligned}$$

This implies that $\mathbb{E}[\zeta(Z,X,\pi)(\iota(\mathbf{1}\{T=t\})-P_t^o(X))]=0$. Hence,

$$b_{t,k}\mathbb{E}\left[\zeta(Z,X,\pi)\left(\iota(\mathbf{1}\{T=t\})-P_t^o(X)\right)+P_t^o(X)\right]=b_{t,k}\mathbb{E}\left[P_t^o(X)\right]=p_{t,k}^o.$$

This proves the proposition. □

Proof of Proposition 4.2. Since $b_{t,k}$ is a finite vector, it suffices to verify the Neyman orthogonality condition for ψ_z, which is defined by

$$\begin{aligned}
&\psi_z(Y,T,Z,X,\beta_{t,k},Q_t,P_t,\pi_z)\\
=&\big((\mathbf{1}\{Z=z\}/\pi_z(X))\left(\mathbf{1}\{T=t\}-P_{t,z}(X)\right)+P_{t,z}(X)\big)\beta_{t,k}\\
&-(\mathbf{1}\{Z=z\}/\pi_z(X))\left(Y\mathbf{1}\{T=t\}-Q_{t,z}(X)\right)-Q_{t,z}(X).
\end{aligned}$$

We want to show that

$$\frac{d}{dr}\mathbb{E}\left[\psi_z(Y,T,Z,X,\beta_{t,k},Q_t^r,P_t^r,\pi_z^r)\right]\Big|_{r=0}=0,$$

where $Q_t^r = Q_t^o + r(Q_t - Q_t^o)$, $P_t^r = P_t^o + r(P_t - P_t^o)$, and $\pi_z^r = \pi_z^o + r(\pi_z - \pi_z^o)$. In fact,

$$\begin{aligned}
&\frac{d}{dr}\mathbb{E}\left[\psi_z(Y,T,Z,X,\beta_{t,k},Q_t^r,P_t^r,\pi_z^r)\right]\big|_{r=0}\\
=&\mathbb{E}\Big[\frac{-\mathbf{1}\{Z=z\}}{(\pi_z^r(X))^2}\Big(\mathbf{1}\{T=t\}-P_{t,z}^r(X)\Big)\left(\pi_z(X)-\pi_z^o(X)\right)\beta_{t,k}\\
&+\left(P_{t,z}(X)-P_{t,z}^o(X)-\frac{\mathbf{1}\{Z=z\}}{\pi_z^r(X)}\Big(P_{t,z}(X)-P_{t,z}^o(X)\Big)\right)\beta_{t,k}\\
&+\frac{\mathbf{1}\{Z=z\}}{(\pi_z^r(X))^2}\Big(Y\mathbf{1}\{T=t\}-Q_{t,z}^r(X)\Big)\left(\pi_z(X)-\pi_z^o(X)\right)\\
&-\left(Q_{t,z}(X)-Q_t^o(X)\right)+\frac{\mathbf{1}\{Z=z\}}{\pi_z^r(X)}\Big(Q_{t,z}(X)-Q_{t,z}^o(X)\Big)\Big]\Big|_{r=0}\\
=&\mathbb{E}\Big[\frac{-\mathbf{1}\{Z=z\}}{(\pi_z^o(X))^2}\Big(\mathbf{1}\{T=t\}-P_{t,z}^o(X)\Big)\left(\pi_z(X)-\pi_z^o(X)\right)\beta_{t,k}\\
&+\left(P_{t,z}(X)-P_{t,z}^o(X)-\frac{\mathbf{1}\{Z=z\}}{\pi_z^o(X)}\Big(P_{t,z}(X)-P_{t,z}^o(X)\Big)\right)\beta_{t,k}\\
&+\frac{\mathbf{1}\{Z=z\}}{(\pi_z^o(X))^2}\Big(Y\mathbf{1}\{T=t\}-Q_{t,z}^o(X)\Big)\left(\pi_z(X)-\pi_z^o(X)\right)\\
&-\left(Q_{t,z}(X)-Q_{t,z}^o(X)\right)+\frac{\mathbf{1}\{Z=z\}}{\pi_z^o(X)}\Big(Q_{t,z}(X)-Q_{t,z}^o(X)\Big)\Big],
\end{aligned}$$

which equals zero because of the following three identities:

$$\begin{aligned}
\mathbb{E}[\mathbf{1}\{Z=z\}/\pi_z^o(X)\mid X]&=1,\\
\mathbb{E}[\mathbf{1}\{Z=z\}/\pi_z^o(X)(\mathbf{1}\{T=t\}-P_{t,z}^o(X))\mid X]&=0,\\
\mathbb{E}[\mathbf{1}\{Z=z\}/\pi_z^o(X)(Y\mathbf{1}\{T=t\}-Q_{t,z}^o(X))\mid X]&=0.
\end{aligned}$$

□

Proof of Theorem 4.3. The asserted claims follow from Theorem 3.1, Theorem 3.2, and Corollary 3.2 of Chernozhukov et al. (2018) (henceforth referred to as the DML paper). We want to verify their Assumption 3.1 and 3.2. Adopting the notation from the DML paper, we let

$$\psi^a(T,Z,X,P_t,\pi)=-b_{t,k}\Big(\zeta(Z,X,\pi)\left(\iota\mathbf{1}\{T=t\}-P_t(X)\right)+P_t(X)\Big)$$

and

$$\psi^b(Y,T,Z,X,Q_t,\pi) = b_{t,k}\Big(\zeta(Z,X,\pi)\big(\iota(Y\mathbf{1}\{T=t\}) - Q_t(X)\big) + Q_t(X)\Big)$$

so that the linearity of the moment condition (with respect to $\beta_{t,k}$) is verified by the fact that $\psi = \psi^a\beta_{t,k} + \psi^b$. Define[1]

$$\epsilon_n = \max_{z\in\mathcal{Z}}\big(\|\hat{Q}_{t,z} - Q^o_{t,z}\|_2 \vee \|\hat{P}_{t,z} - P^o_t\|_2 \vee \|\hat{\pi}_z - \pi^o_z\|_2\big).$$

By assumption on the convergence rates of the nonparametric estimators, we have $\epsilon_n = o(n^{-1/4})$. Define $C_\epsilon = C_{\epsilon,1}\vee C_{\epsilon,2}\vee C_{\epsilon,3}\vee C_{\epsilon,4}$, where $C_{\epsilon,1}, C_{\epsilon,2}, C_{\epsilon,3}$, and $C_{\epsilon,4}$ are positive constant that only depends on C and ϵ and are specified later in the proof. Let δ_n be a sequence of positive constants approaching zero and satisfies that $\delta_n \geq C_\epsilon\big(\epsilon_n^2\sqrt{n}\vee n^{-1/4}\vee n^{-(1-2/q)}\big)$. Such construction is possible since $\sqrt{n}\epsilon_n^2 = o(1)$. We set the nuisance realization set N_n (denoted by $\mathcal{T}_N$ in the DML paper) to be the set of all vector functions $(Q_t, P_t, \pi_z : z\in\mathcal{Z})$ consisting of square-integrable functions $Q_{t,z}, P_{t,z}$, and π_z such that for all $z\in\mathcal{Z}$:

$$\begin{aligned}
&\big\|Q_{t,z}\big\|_q \leq C, P_{t,z}\in[0,1], \pi_z\in[\epsilon,1], z\in\mathcal{Z},\\
&\|Q_{t,z} - Q^o_{t,z}\|_q \vee \|P_{t,z} - P^o_{t,z}\|_q \vee \|\pi_z - \pi^o_z\|_q \leq \epsilon_n,\\
&\|\pi_z - \pi^o_z\|_2 \times \big(\|Q_{t,z} - Q^o_{t,z}\|_2 + \|P_{t,z} - P^o_{t,z}\|_2\big) \leq \epsilon_n^2.
\end{aligned}$$

Consider Assumption 3.1 in the DML paper. Assumption 3.1(d), the Neyman orthogonality condition, is verified by Proposition 4.2, where the validity of the differentiation under the integral operation is verified later in the proof. Assumption 3.1(e), the identification condition, is verified by the condition that $p^o_{t,k}\in[\epsilon,1]$. The remaining conditions of Assumption 3.1 in the DML paper are trivially verified.

Next, we consider Assumption 3.2 in the DML paper. Note that Assumption 3.2(a) holds by the construction of N_n and ϵ_n and our assumptions on the nuisance estimates. Assumption 3.2(d) is verified by our assumption that the semiparametric efficiency bound of $\beta_{t,k}$ is above ϵ. The remaining task is to verify Assumption 3.2(b) and 3.2(c) in the DML paper. To do that, we choose n sufficiently large and let $(Q_{t,z}, P_{t,z}, \pi_z : z\in\mathcal{Z})$ be an arbitrary element of the nuisance realization set N_n. We keep the above notations

[1] For simplicity, we drop the superscript l in the nonparametric estimators.

throughout the remaining part of the proof. Define

$$\psi_z^a(T, Z, X, P_t, \pi_z) = \frac{\mathbf{1}\{Z = z\}}{\pi_z(X)}(\mathbf{1}\{T = t\} - P_{t,z}(X)) + P_{t,z}(X)$$

and

$$\psi_z^b(Y, T, Z, X, Q_t, \pi_z) = \frac{\mathbf{1}\{Z = z\}}{\pi_z(X)}(Y\mathbf{1}\{T = t\} - Q_{t,z}(X)) + Q_{t,z}(X).$$

Since ψ^a is a linear combination of $\psi_z^a, z \in \mathcal{Z}$ and ψ^b is a linear combination of $\psi_z^b, z \in \mathcal{Z}$, we only need $\left\|\psi_z^a(T, Z, X, P_t, \pi_z)\right\|_q$ and $\left\|\psi_z^b(Y, T, Z, X, Q_t, \pi_z)\right\|_q$ to be uniformly bounded (i.e., the bounds do not depend on n) for $z \in \mathcal{Z}$ in order to verify Assumption 3.2(b) in the DML paper. In fact,

$$\begin{aligned}\left\|\psi_z^b(Y, T, Z, X, P_t, \pi_z)\right\|_q \leq& \left\|\mathbf{1}\{Z = z\}/\pi_z(X)\left|Y\mathbf{1}\{T = t\} - Q_{t,z}(X)\right|\right\|_q + \left\|Q_{t,z}(X)\right\|_q \\ \leq& \frac{1}{\epsilon}\left(\left\|Y\mathbf{1}\{T = t\}\right\|_q + \left\|Q_{t,z}(X)\right\|_q\right) + \left\|Q_{t,z}(X)\right\|_q \leq 2C/\epsilon + C,\end{aligned}$$

where we have used the assumption that $\pi_z \geq \epsilon$, $\left\|Y\mathbf{1}\{T = t\}\right\|_q \leq C$, and $\left\|Q_t(X)\right\|_q \leq C$. Similarly, we have

$$\begin{aligned}\left\|\psi_z^a(T, Z, X, P_t, \pi_z)\right\|_q \leq& \left\|\mathbf{1}\{Z = z\}/\pi_z(X)\left|\mathbf{1}\{T = t\} - P_{t,z}(X)\right|\right\|_q + \left\|P_{t,z}(X)\right\|_q \\ \leq& \frac{1}{\epsilon}\left(1 + \left\|P_{t,z}(X)\right\|_q\right) + \left\|P_{t,z}(X)\right\|_q \leq 2/\epsilon + 1,\end{aligned}$$

where we have used the assumption that $\pi_z \geq \epsilon$ and $P_t \in [0, 1]$. Thus, Assumption 3.2(b) in the DML paper is verified.

To verify Assumption 3.2(c) in the DML paper, we again only need to verify the corresponding conditions for ψ_z^a and ψ_z^b, respectively. For ψ_z^a, we have

$$\begin{aligned}&\left\|\psi_z^a(T, Z, X, P_t, \pi_z) - \psi_z^a(T, Z, X, P_t^o, \pi_z^o)\right\|_2 \\ \leq& \left\|\frac{\pi_z(X) - \pi_z^o(X)}{\pi_z(X)\pi_z^o(X)}\right\|_2 + \left\|\frac{P_{t,z}(X)}{\pi_z(X)} - \frac{P_{t,z}^o(X)}{\pi_z^o(X)}\right\|_2 + \left\|P_{t,z}(X) - P_{t,z}^o(X)\right\|_2 \\ \leq& \frac{1}{\epsilon^2}\left\|\pi_z(X) - \pi_z^o(X)\right\|_2 + \frac{1}{\epsilon^2}\left\|(P_{t,z}(X) - P_{t,z}^o(X))\pi_z^o(X) + P_{t,z}^o(X)(\pi_z^o(X) - \pi_z(X))\right\|_2 \\ &+ \left\|P_{t,z}(X) - P_{t,z}^o(X)\right\|_2 \\ \leq& \frac{2}{\epsilon^2}\left\|\pi_z(X) - \pi_z^o(X)\right\|_2 + \left(1/\epsilon^2 + 1\right)\left\|P_{t,z}(X) - P_{t,z}^o(X)\right\|_2 \leq C_{\epsilon,1}\varepsilon_n \leq \delta_n,\end{aligned}$$

where the second to last inequality follows from the fact that $P^o_{t,z}, \pi^o_z \in [0,1]$. For ψ^b_z, we have

$$
\begin{aligned}
&\left\|\psi^b_z(Y,T,Z,X,Q_t,\pi_z) - \psi^b_z(Y,T,Z,X,Q^o_t,\pi^o_z)\right\|_2 \\
\leq & \frac{1}{\epsilon^2}\left\|\pi^o_z(X)(Y\mathbf{1}\{T=t\} - Q_{t,z}(X)) - \pi_z(X)(Y\mathbf{1}\{T=t\} - Q^o_{t,z}(X))\right\|_2 \\
& + \left\|Q_{t,z}(X) - Q^o_{t,z}(X)\right\|_2 \\
= & \frac{1}{\epsilon^2}\left\|(Y\mathbf{1}\{T=t\} - Q^o_{t,z}(X))(\pi^o_z(X) - \pi_z(X)) + \pi^o_z(X)(Q^o_{t,z}(X) - Q_{t,z}(X))\right\|_2 \\
& + \left\|Q_{t,z}(X) - Q^o_{t,z}(X)\right\|_2 \\
\leq & \frac{1}{\epsilon^2}\left\|(Y\mathbf{1}\{T=t\} - Q^o_{t,z}(X))(\pi^o_z(X) - \pi_z(X))\right\|_2 + \left\|\pi^o_z(X)(Q^o_{t,z}(X) - Q_{t,z}(X))\right\|_2 \\
& + \left\|Q_{t,z}(X) - Q^o_{t,z}(X)\right\|_2 \\
\leq & \frac{C}{\epsilon^2}\left\|\pi^o_z(X) - \pi_z(X)\right\|_2 + \left(\frac{1}{\epsilon^2} + 1\right)\left\|Q^o_{t,z}(X) - Q_{t,z}(X)\right\|_2 \leq C_{\epsilon,2}\varepsilon_n \leq \delta_n,
\end{aligned}
$$

where the last inequality follows from our assumption that $|Y\mathbf{1}\{T=t\} - Q^o_t(X)| \leq C$ and the fact that $\pi^o_z \in [\epsilon, 1]$. Combining the above two inequality results, we can verify the first two conditions of Assumption 3.2(c) in the DML paper.

For the last condition of Assumption 3.2(c) in the DML paper, which bounds the second-order Gateaux derivative, we again consider ψ^a_z and ψ^b_z separately. For $r \in [0,1)$, recall that $Q^r_{t,z} = Q^o_{t,z} + r(Q_{t,z} - Q^o_{t,z})$, $P^r_{t,z} = P^o_{t,z} + r(P_{t,z} - P^o_{t,z})$, and $\pi^r_z = \pi^o_z + r(\pi_z - \pi^o_z)$. Clearly, $P^r_{t,z}, \pi^r_z \in [0,1]$. With differentiation under the integral, we have

$$
\begin{aligned}
&\frac{\partial^2}{\partial r^2}\mathbb{E}\left[\psi^a_z(T,Z,X,P^r_t,\pi^r_z)\right] \\
= & \frac{\partial}{\partial r}\mathbb{E}\Big[\frac{-\mathbf{1}\{Z=z\}}{(\pi^r_z(X))^2}\Big(\mathbf{1}\{T=t\} - P^r_{t,z}(X)\Big)\big(\pi_z(X) - \pi^o_z(X)\big) \\
& + P_{t,z}(X) - P^o_{t,z}(X) - \frac{\mathbf{1}\{Z=z\}}{\pi^r_z(X)}\Big(P_{t,z}(X) - P^o_{t,z}(X)\Big)\Big] \\
= & \mathbb{E}\Big[\frac{2\times\mathbf{1}\{Z=z\}}{(\pi^r_z(X))^3}(\pi_z(X) - \pi^o_z(X))^2(\mathbf{1}\{T=t\} - P^r_{t,z}(X))\Big] \\
& + \mathbb{E}\Big[\frac{\mathbf{1}\{Z=z\}}{(\pi^r_z(X))^2}(\pi_z(X) - \pi^o_z(X))(P_{t,z}(X) - P^o_{t,z})\Big] \\
& + \mathbb{E}\Big[\frac{\mathbf{1}\{Z=z\}}{(\pi^r_z(X))^2}(\pi_z(X) - \pi^o_z(X))(\mathbf{1}\{T=t\} - P^r_{t,z}(X))(P_{t,z}(X) - P^o_{t,z})\Big] \\
& - \mathbb{E}\Big[\frac{\mathbf{1}\{Z=z\}}{\pi^r_z(X)}(\mathbf{1}\{T=t\} - P^r_{t,z}(X))(P_{t,z}(X) - P^o_{t,z})^2\Big].
\end{aligned}
$$

Using the fact that $|\mathbf{1}\{T=t\} - P_t^r(X)| \leq 1$ and $\pi_z^r \geq \epsilon$, we can bound the above derivative by

$$\Big|\frac{\partial^2}{\partial r^2}\mathbb{E}\left[\psi_z^a(T,Z,X,P_t^r,\pi_z^r)\right]\Big| \leq C_\epsilon\big(\big\|\pi_z(X)-\pi_z^o(X)\big\|_2^2 + \big\|P_{t,z}(X)-P_{t,z}^o(X)\big\|_2^2\big)$$
$$+ C_\epsilon\big\|\pi_z(X)-\pi_z^o(X)\big\|_2 \times \|P_{t,z}(X)-P_{t,z}^o(X)\|_2$$
$$\leq C_{\epsilon,3}\varepsilon_n^2 \leq \delta_n/\sqrt{n}.$$

By bounding the first and second derivative uniformly with respect to r, we know that the differentiation under the integral operation is valid. Therefore, the Neyman orthogonality condition is verified. Analogously, we can show that

$$\begin{aligned}
&\frac{\partial^2}{\partial r^2}\mathbb{E}\left[\psi_z^b(Y,T,Z,X,Q_t^r,\pi_z^r)\right]\\
=&\mathbb{E}\Big[\frac{2\times\mathbf{1}\{Z=z\}}{(\pi_z^r(X))^3}(\pi_z(X)-\pi_z^o(X))^2(Y\mathbf{1}\{T=t\}-Q_{t,z}^r(X))\Big]\\
&+\mathbb{E}\Big[\frac{\mathbf{1}\{Z=z\}}{(\pi_z^r(X))^2}(\pi_z(X)-\pi_z^o(X))(Q_{t,z}(X)-Q_{t,z}^o)\Big]\\
&-\mathbb{E}\Big[\frac{\mathbf{1}\{Z=z\}}{(\pi_z^r(X))^2}(\pi_z(X)-\pi_z^o(X))(Y\mathbf{1}\{T=t\}-Q_{t,z}^r(X))(Q_{t,z}(X)-Q_{t,z}^o)\Big]\\
&-\mathbb{E}\Big[\frac{\mathbf{1}\{Z=z\}}{\pi_z^r(X)}(Y\mathbf{1}\{T=t\}-Q_{t,z}^r(X))(Q_{t,z}(X)-Q_{t,z}^o)^2\Big].
\end{aligned}$$

Under the assumption $|Y\mathbf{1}\{T=t\}-Q_{t,z}^o(X)| \leq C$, we have

$$|Y\mathbf{1}\{T=t\}-Q_{t,z}^r(X)| \leq |Y\mathbf{1}\{T=t\}-Q_{t,z}^o(X)| + r|Q_{t,z}(X)-Q_{t,z}^o| \leq C+1,$$

for all $r\in[0,1]$ and n large enough. Then we can bound the above derivative by

$$\Big|\frac{\partial^2}{\partial r^2}\mathbb{E}\left[\psi_z^b(Y,T,Z,X,Q_t^r,\pi_z^r)\right]\Big| \leq C_\epsilon\big(\big\|\pi_z(X)-\pi_z^o(X)\big\|_2^2 + \big\|Q_{t,z}(X)-Q_{t,z}^o(X)\big\|_2^2\big)$$
$$+ C_\epsilon\big\|\pi_z(X)-\pi_z^o(X)\big\|_2 \times \Big\|Q_{t,z}(X)-Q_{t,z}^o(X)\Big\|_2$$
$$\leq C_{\epsilon,4}\varepsilon_n^2 \leq \delta_n/\sqrt{n}.$$

Therefore, we have verified the last condition of Assumption 3.2(c) in the DML paper.

Lastly, we need to verify the condition on δ_n in Theorem 3.1 and 3.2 in the DML paper, that is, $\delta_n \geq n^{-[(1-2/q)\wedge(1/2)]}$. This directly follows from the construction of δ_n. $\square$

A.4 Proofs for weak IV inference results

Proof of Theorem 5.1. We first prove part (i). Consider applying the DML method to the moment condition (5) to estimate the parameter $\upsilon - \beta_0 p$ and obtain the standard error. We want to show the convergence in distribution of

$$\check{\sigma}_\psi^{-1}\sqrt{n}\left[(\check{\upsilon} - \beta_0\check{p}) - (\upsilon - \beta_0 p)\right] = \check{\rho} - \sqrt{n}(\upsilon - \beta_0 p)/\check{\sigma}_\psi \tag{A.1}$$

to the standard normal distribution uniformly over the DGPs in $\mathcal{P}^{\mathrm{WI}}(c_0, c_1)$. To do that, we need to verify Assumptions 3.1 and 3.2 in the DML paper regarding the above moment condition. Assumptions 3.1(a)-(c) hold trivially. Assumption 3.1(d), the Neyman orthogonality condition, is verified by Proposition 4.2. That is, the Gateaux derivatives with respect to the nuisance parameters are zero regardless of the value of β. Assumption 3.1(e), the identification condition, is verified since the Jacobian of the parameter in the moment condition is 1. Assumption 3.2 in the DML paper can be verified in the same way as in the proof of Theorem 4.3. For brevity, we do not repeat the verification here.

For DGPs in $\mathcal{P}^{\mathrm{WI}}_{\beta_0}(c_0, c_1)$, (A.1) is equal to $\check{\rho}$. Therefore, the uniform convergence in distribution of $|\check{\rho}|$ is established in the null space, and the size of the test is uniformly controlled accordingly. For DGPs in $\mathcal{P}^{\mathrm{WI}}_{\beta}(c_0, c_1)$, where $\beta > \beta_0$, we have

$$\begin{aligned}\check{\rho} &= \left(\check{\rho} - \sqrt{n}(\upsilon - \beta_0 p)/\check{\sigma}_\psi\right) + \sqrt{n}(\upsilon - \beta_0 p)/\check{\sigma}_\psi \\ &= \left(\check{\rho} - \sqrt{n}(\upsilon - \beta_0 p)/\check{\sigma}_\psi\right) + \sqrt{n}(\beta - \beta_0) p/\check{\sigma}_\psi.\end{aligned}$$

The first term on the RHS of the last equality converges in distribution to $N(0,1)$. In contrast, the second term diverges to infinity since $\check{\sigma}_\psi$ converges in probability to $\sigma_\psi \geq \sqrt{c_0}$ by Theorem 3.2 in the DML paper. Therefore, the probability of $|\check{\rho}|$ exceeding any finite number converges to 1. The case where $\beta < \beta_0$ is essentially the same.

To prove part (ii) of the theorem, notice that $(\beta - \beta_0)p \leq 0$ for any DGP in the null space $\bigcup_{\beta \leq \beta_0} \mathcal{P}^{\mathrm{WI}}_{\beta}(c_0, c_1)$, which implies that $\check{\rho} \leq \check{\rho} - \sqrt{n}(\upsilon - \beta_0 p)/\check{\sigma}_\psi$. Therefore,

$$\sup_P \mathbb{P}_P\left(\check{\rho} > \mathcal{N}_{1-\alpha}\right) \leq \sup_P \mathbb{P}_P\left(\check{\rho} - \sqrt{n}(\upsilon - \beta_0 p)/\check{\sigma}_\psi > \mathcal{N}_{1-\alpha}\right) \to \alpha,$$

where the supremum is taken over $P \in \bigcup_{\beta \leq \beta_0} \mathcal{P}^{\mathrm{WI}}_{\beta}(c_0, c_1)$. Consistency can be derived in the same way as part (i). $\square$

B Simulation Study

To evaluate the performance of our weak-identification-robust inference approach, we designed an experiment wherein the variables Y, Z, and S exhibit a nonlinear relationship with the regressor X. Addressing both the "curse of dimensionality" and weak identification simultaneously can be a challenging task; therefore, we focus on the performance of the weak-identification-robust inference procedure in a low-dimensional setting with $d_X = 3$. This low-dimensional setup is typical in the simulation studies found in the literature. For example, Cattaneo (2010)'s simulation has two covariates and Hong and Nekipelov (2010a) has one covariate.

The data generating process is adapted from (Hong and Nekipelov, 2010a) and modified to more effectively address the multiple treatment and weak identification scenarios. The setup is identical to that in the empirical study, featuring three treatment levels and two instrument levels. We initially draw n i.i.d. samples of mutually independent $X = (X_1, X_2, X_3)$ from the uniform distribution on $[-0.5, 0.5]$. Subsequently, the instrument Z is generated according to the Bernoulli distribution with parameter $(X_1 + X_2 + X_3)/3 + 0.5$. The type S is generated according to the following distribution:

$$
\begin{aligned}
&\mathbb{P}(S = s_1|X) = 0.2 + X_1/10,\\
&\mathbb{P}(S = s_2|X) = 0.2 + X_2/10,\\
&\mathbb{P}(S = s_3|X) = 0.2 + X_1X_2/10,\\
&\mathbb{P}(S = s_4|X) = c|X_1 + X_2|/\sqrt{n},\\
&\mathbb{P}(S = s_5|X) = 0.4 - (X_1 + X_2 + X_1X_2)/10 - |X_1 + X_2|/\sqrt{n}.
\end{aligned}
$$

In this way, the type S is dependent on the covariates X, but independent with the instrument Z given X. The weak identification issue is modeled by the drifting sequence $\mathbb{P}(S = s_4) = O(1/\sqrt{n})$, with the concentration parameter c. For simplicity, we denote $S_j = \mathbf{1}\{S = s_j\}, j = 1, \cdots, 5$. The treatment is determined by the instrument and the type: $T_1 = S_1 + S_4(1 - Z)$, $T_2 = S_2 + S_5(1 - Z)$, and $T_3 = S_3 + (S_4 + S_5)Z$. The potential outcomes are generated based on Poisson distributions. We first generate six Poisson variables:

$$\xi_j \sim \text{Poisson}(\exp(X_1 + X_2 + X_3) + j), \tilde{\xi}_j \sim \text{Poisson}(j), j = 1, 2, 3.$$

Then potential outcomes are generated by

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix} + S_1 \begin{pmatrix} \tilde{\xi}_1 \\ \tilde{\xi}_1 \\ \tilde{\xi}_1 \end{pmatrix} + S_2 \begin{pmatrix} \tilde{\xi}_2 \\ \tilde{\xi}_2 \\ \tilde{\xi}_2 \end{pmatrix} + S_3 \begin{pmatrix} \tilde{\xi}_3 \\ \tilde{\xi}_3 \\ \tilde{\xi}_3 \end{pmatrix}.$$

The potential outcomes structure is similar to the simulation study presented in Hong and Nekipelov (2010a). For both *no*-compliers and *nm*-compliers, the potential outcomes are independent. Conversely, for *no*-never-takers, *nm*-never-takers, and always-takers, the potential outcomes are correlated. Our primary interest lies in obtaining the confidence region for $\beta_{no,1}$, the LASF for the vanishing subpopulation s_4. Under this framework, $\beta_{no,1} = (\exp(0.5) - \exp(-0.5))^3$.

Table 1 summarizes the simulation results from 1000 iterations. We explore different values for the concentration parameter $c = 1, 2, 3$ in the data-generating process. In constructing the test statistic $\check{\rho}$, we select the number of cross-fitting folds as $L = 2, 5$. The nonparametric estimators for nuisance parameters are generated via local linear regressions using the Epanechnikov kernel function. We evaluate three bandwidth choices, $0.7, 0.8, 0.9$, and examine sample sizes of $n = 250, 500, 1000$.

The simulation results indicate that, generally, the confidence region exhibits satisfactory coverage probabilities. It should be noted that the performance does not improve with increasing sample size due to the construction of $\mathbb{P}(S = s_4)$, which implies that the weak identification issue escalates as the sample size grows. As the concentration parameter c represents the severity of weak identification, it is observed that in many instances, coverage probability decreases as c increases. The choice of the number of folds for cross-fitting and bandwidth does not significantly affect the performance.[2]

[2]We have also experimented with other bandwidth values. For values below 0.7, there are insufficient observations within the bandwidth window, resulting in uninvertible matrices in the nonparametric regression. For values moderately above 0.9, the results are comparable.

Table 1: Simulation study for weak identification.

L	bandwidth	c	$n = 250$			$n = 500$			$n = 1000$		
			nominal coverage			nominal coverage			nominal coverage		
			.900	.950	.990	.900	.950	.990	.900	.950	.990
		1	.896	.949	.992	.889	.947	.989	.890	.944	.995
2	0.7	2	.915	.959	.992	.893	.942	.985	.892	.950	.989
		3	.907	.954	.988	.892	.940	.986	.883	.937	.984
		1	.912	.947	.991	.894	.950	.986	.899	.949	.995
2	0.8	2	.917	.960	.992	.893	.948	.987	.889	.951	.989
		3	.916	.961	.990	.881	.937	.995	.869	.938	.987
		1	.911	.947	.996	.891	.948	.987	.903	.948	.994
2	0.9	2	.908	.956	.993	.896	.946	.985	.885	.950	.991
		3	.910	.961	.993	.882	.929	.994	.877	.937	.985
		1	.916	.963	.994	.906	.950	.987	.906	.957	.993
5	0.7	2	.913	.960	.992	.901	.952	.991	.887	.949	.993
		3	.911	.964	.993	.891	.946	.994	.873	.935	.988
		1	.918	.963	.994	.905	.946	.988	.908	.955	.994
5	0.8	2	.910	.958	.993	.897	.944	.990	.886	.950	.993
		3	.920	.969	.991	.892	.945	.987	.878	.936	.991
		1	.910	.960	.995	.912	.951	.991	.906	.953	.995
5	0.9	2	.908	.953	.993	.901	.943	.991	.884	.951	.993
		3	.916	.962	.990	.889	.944	.987	.874	.939	.991

Simulated coverage probabilities of the confidence region for the LASF $\beta_{no,1}$. The confidence region is obtained by inverting the null-restricted test. The DGP models weak identification by having $\mathbb{P}(S = s_4)$ proportional to $c/\sqrt{n}$. Number of replications is 1000.

C Implicitly Defined Parameters

This section studies general parameters defined implicitly through moment conditions. We allow the moment conditions to be non-smooth, which is the case when the parameter of interest is the quantile. We also allow the moment conditions to be overidentifying, which could be the result of imposing the underlying economic theory on multiple levels of treatment and instrument.

To facilitate the exposition, we define a random variable $Y_{t,k}^*$ such that the marginal distribution of $Y_{t,k}^*$ is equal to the conditional distribution of Y_t given $S \in \Sigma_{t,k}$. The joint distribution of the $Y_{t,k}^*$'s is irrelevant and hence left unspecified. For convenience, we use a single index $j \in J$ rather than (t,k) for labeling. That is, we collect the $Y_{t,k}^*$'s into the vector $Y^* = (Y_1^*, \cdots, Y_J^*)$. Let t_j be the treatment level associated with Y_j^*. The quantities p_j and b_j are analogously defined.[3]

Let the parameter of interest be η, which lies in the parameter space $\Lambda \subset \mathbb{R}^{d_\eta}$, $d_\eta \leq J$. The true value of the parameter η_0 satisfies the moment condition

$$\mathbb{E}\left[m(Y^*, \eta^o)\right] = 0,$$

where $m : \mathcal{Y}^J \times \mathbb{R}^{d_\eta} \to \mathbb{R}^J$ is a vector of functions:

$$m(Y^*, \eta) = \left(m_1(Y_1^*, \eta), \cdots, m_J(Y_J^*, \eta)\right)'$$

Since the vector η appears in each m_j, restrictions are allowed both within and across different subpopulations. Another interesting feature of this specification is that the moment conditions are defined for the random variables that are not observed. But their marginal distributions can be identified similar to Theorem 2.1.

Let $\bar{m} = (\bar{m}_1', \cdots, \bar{m}_J')'$, where

$$\bar{m}_j(X, \eta) = \Big(\bar{m}_{j,z_1}(X, \eta), \cdots, \bar{m}_{j,z_{N_Z}}(X, \eta)\Big)'$$

and

$$\bar{m}_{j,z}(X, \eta) = \mathbb{E}\left[m_j(Y, \eta)\mathbf{1}\{T = t_j\} \mid Z = z, X\right].$$

[3]We can further extend the vector Y^* to include variables whose marginal distributions are the same as the conditional distributions of Y_t given $T = t, S \in \Sigma_{t,k}$. Efficient estimation in this more general case is similar and hence omitted for brevity.

The functions $\bar{m}_{j,z}$ are identified from the data. Similar to Theorem 2.1, we can show that the parameter η is identified by the moment conditions:

$$b_j \mathbb{E}\left[\bar{m}_j(X,\eta)\right] = 0, 1 \le j \le J \iff \eta = \eta^o.$$

The following theorem gives the SPEB for the estimation of η.

Theorem C.1. *Assume the following conditions hold.*

(i) $\mathbb{E}\left[m(Y^*,\eta)^2\right] < \infty, \eta \in \Lambda$.

(ii) For each j and z, $m_{j,t_j,z}$ is continuously differentiable in its second argument. Let Γ be the $J \times d_\eta$ matrix whose jth row is $b_j \frac{d}{d\eta}\mathbb{E}\left[\bar{m}_j(X,\eta)\right]\big|'_{\eta=\eta^o}$, and assume Γ has full column rank.

Then for the estimation of η, the EIF is

$$-\left(\Gamma' V^{-1}\Gamma\right)^{-1}\Gamma' V^{-1}\psi^\eta(Y,T,Z,X,\eta^o,\pi^o,\bar{m}^o), \tag{C.1}$$

where

$$V = \mathbb{E}\left[\psi^\eta(Y,T,Z,X,\eta,\pi,\bar{m})\psi^\eta(Y,T,Z,X,\eta,\pi,\bar{m})'\right]$$

and $\psi^\eta(Y,T,Z,X,\eta,\pi,\bar{m})$ is a $J\times 1$ random vector whose jth element is

$$b_j\left(\zeta(Z,X,\pi)\left(\iota(m_j(Y,\eta)\mathbf{1}\{T=t_j\}) - \bar{m}_j(X,\eta)\right) + \bar{m}_j(X,\eta)\right) \tag{C.2}$$

In particular, the semiparametric efficiency bound is $\left(\Gamma' V^{-1}\Gamma\right)^{-1}$.

Proof of Theorem C.1. The proof is based on the approach described in section 3.6 of Hong and Nekipelov (2010a) and the proof of Theorem 1 in Cattaneo (2010). We use a constant $d_\eta \times d_m$ matrix A to transform the overidentified vector of moments into an exactly identified system of equations $A\left(b_j\mathbb{E}\left[\bar{m}_j(X,\eta)\right]\right)_{j=1}^J = 0$, find the A-dependent EIF for the exactly-identified parameter, and choose the optimal A. In a parametric submodel, the implicit function theorem gives that

$$\frac{\partial}{\partial\theta}\eta\big|_{\theta=\theta^o} = -\left(A\Gamma\right)^{-1} A\frac{\partial}{\partial\theta}\left(b_j\mathbb{E}_\theta\left[\bar{m}_j(X,\eta^o)\right]\right)_{j=1}^J\big|_{\theta=\theta^o},$$

where $\frac{\partial}{\partial\theta}\mathbb{E}_{\theta}\left[\bar{m}_j(X,\eta^o)\right]\big|_{\theta=\theta^o}$ is an $N_Z \times 1$ random vector whose typical element can be represented by

$$\int m_j(y,\eta^o)\mathbf{1}\{\tau = t_j\} s_z(y,\tau \mid x;\theta^o) f_z(y,\tau \mid x;\theta^o) f_X(x;\theta^o) dy d\tau dx$$
$$+ \int m_j(y,\eta^o)\mathbf{1}\{\tau = t_j\} s_X(x;\theta^o) f_z(y,\tau \mid x;\theta^o) f_X(x;\theta^o) dy d\tau dx,$$

for $z \in \mathcal{Z}$. So the EIF for this exactly-identified parameter is

$$\psi^A(Y,T,Z,X,\eta^o,\pi^o,\bar{m}^o) = -\left(A\Gamma\right)^{-1} A\Psi^\eta(Y,T,Z,X,\eta^o,\pi^o,\bar{m}^o),$$

where ψ^η is defined by Equation (C.2). It is straightforward to verify that ψ^A satisfies $\frac{\partial}{\partial\theta}\eta\big|_{\theta=\theta^o} = \mathbb{E}\left[\psi^A s'_{\theta^o}\right]$, and $\psi^A \in \mathscr{S}$. The optimal A is chosen by minimizing the sandwich matrix $\mathbb{E}\left[\psi^A(\psi^A)'\right] = \left(A\Gamma\right)^{-1} A\mathbb{E}\left[\psi^\eta(\psi^\eta)'\right] A' \left(\Gamma' A'\right)^{-1}$. Thus, the EIF for the over-identified parameter is obtained when $A = \Gamma' V^{-1}$. Plugging this expression into ψ^A, we obtain Equation (C.1). □

Note that, for example, $m_j(Y_j^*,\eta) = Y_j^* - \eta$, then $\eta = \beta_j$, and the efficiency bound shown above reduces to the one computed in Theorem 3.1. If $T = Z$, that is, the treatment satisfies the unconfounded, then the Theorem C.1 reduces to Theorem 1 in Cattaneo (2010).

For estimation, we use the EIFs to generate moment conditions and propose a three-step semiparametric GMM procedure. The criterion function is

$$\Psi_n^\eta(\eta,\pi,m) = \frac{1}{n}\sum_{i=1}^n \psi^\eta(Y_i,T_i,Z_i,X_i,\eta,\pi,\bar{m}). \tag{C.3}$$

Its probability limit is denoted as

$$\Psi^\eta(\eta,\pi,m_Z) = \mathbb{E}\left[\psi^\eta(Y,T,Z,X,\eta,\pi,\bar{m})\right], \tag{C.4}$$

where the expectation is taken with respect to the true parameters $(\pi^o,\bar{m}^o)$. The implementation procedure is as follows. Assume that we have nonparametric estimators $\hat{\pi}$ and $\hat{m}$ that consistently estimate π^o and $\bar{m}^o$, respectively. We first find a consistent GMM estimator $\tilde{\eta}$ using the identity matrix as the weighting matrix, that is,

$$\left\|\Psi_n^\eta(\tilde{\eta},\hat{\pi},\hat{m})\right\|_2 \leq \inf_{\eta\in\Lambda}\left\|\Psi_n^\eta(\eta,\hat{\pi},\hat{m})\right\|_2 + o_p(1). \tag{C.5}$$

Next, we use this estimate to form a consistent estimator $\hat{V}$ of the covariance matrix V, where

$$\hat{V} = \frac{1}{n}\sum_{i=1}^{n}\psi^{\eta}(Y_i, T_i, Z_i, X_i, \tilde{\eta}, \hat{\pi}, \hat{m})\psi^{\eta}(Y_i, T_i, Z_i, X_i, \tilde{\eta}, \hat{\pi}, \hat{m})'.$$

Then we let $\hat{\eta}$ be the optimally-weighted GMM estimator:

$$\begin{aligned}&\Psi_n^{\eta}(\hat{\eta}, \hat{\pi}, \hat{m}_Z)V_n(\tilde{\eta}, \hat{\pi}, \hat{m}_Z)^{-1}\Psi_n^{\eta}(\hat{\eta}, \hat{\pi}, \hat{m}_Z)' \\ &\leq \inf_{\eta\in\Lambda}\Psi_n^{\eta}(\eta, \hat{\pi}, \hat{m}_Z)V_n(\tilde{\eta}, \hat{\pi}, \hat{m}_Z)^{-1}\Psi_n^{\eta}(\eta, \hat{\pi}, \hat{m}_Z)' + o_p\big(n^{-1/2}\big).\end{aligned}$$

To conduct inference, we estimate Γ using the estimator $\hat{\Gamma}$ whose elements are defined as

$$\hat{\Gamma}_{jl} = \frac{1}{n}\sum_{i=1}^{n} b_j \frac{\partial}{\partial\eta}\hat{m}_j(X_i, \eta)\Big|_{\eta=\hat{\eta}},$$

where we have implicitly assumed that the estimator $\hat{m}_j$ is differentiable in its second argument.

In the following theorem, we derive the asymptotic properties of the GMM estimators. The main theoretical difficulty is that the random criterion function $\Psi_n(\cdot, \hat{\pi}, \hat{m})$ could potentially be discontinuous because we allow $m(Y^*, \cdot)$ to be discontinuous. We use the theory developed in Chen et al. (2003) to overcome this problem.[4] Let Π_z be the function class that contains π_z^o. Let $\mathcal{M}_{j,z}$ be the function class that contains $\bar{m}_{j,z}^o$.

Theorem C.2. *Let the assumptions in Theorem C.1 hold. Assume the following conditions hold.*

(i) The parameter space Λ is compact. The true parameter η^o is in the interior of Λ.

(ii) For any j, z and $\bar{m}_{j,z} \in \mathcal{M}_{j,z}$, there exists $C > 0$ such that for $\delta > 0$ sufficiently small,

$$\sup_{|\eta'-\eta|\leq\delta}\mathbb{E}\big|\bar{m}_{j,z}(X, \eta') - \bar{m}_{j,z}(X, \eta)\big|^2 \leq C\delta^2.$$

[4] Cattaneo (2010) instead uses the theory from Pakes and Pollard (1989). However, the general theory of Chen et al. (2003) is more straightforward to apply in this case since they explicitly assume the presence of infinite-dimensional nuisance parameters, which can depend on the parameters to be estimated.

(iii) *Donsker properties:*

$$\int_0^\infty \log N(\varepsilon, \Pi_z, \|\cdot\|_\infty) d\varepsilon, \int_0^\infty \log N(\varepsilon, \mathcal{M}_{j,z}, \|\cdot\|_\infty) d\varepsilon < \infty,$$

where $N(\varepsilon, \mathcal{F}, \|\cdot\|)$ *denotes the covering number of the space* $(\mathcal{F}, \|\cdot\|)$.

(iv) *Convergence rates of the nonparametric estimators:*

$$\|\hat{\pi}_z - \pi_z^o\|_\infty, \|\hat{m}_{j,z} - \bar{m}_{j,z}^o\|_\infty = o_p(n^{-1/4}).$$

(v) *The function* $\sup_{\eta\in\Lambda}\left|\frac{\partial}{\partial\eta}\bar{m}_j^o(\cdot,\eta)\right|$ *is integrable. The estimator* $\frac{\partial}{\partial\eta}\hat{m}_j$ *is consistent uniformly in its second argument, that is,*

$$\left\|\frac{\partial}{\partial\eta}\hat{m}_j(x,\eta) - \frac{\partial}{\partial\eta}\bar{m}_j^o(x,\eta)\right\|_\infty = o_p(1), \forall x.$$

Then $\tilde{\eta} = \eta^o + o_p(1)$, $\hat{V} = V + o_p(1)$, $\hat{\Gamma} = \Gamma + o_p(1)$, *and*

$$\sqrt{n}\left(\hat{\eta} - \eta^o\right) \implies N\left(\mathbf{0}, (\Gamma' V^{-1}\Gamma)^{-1}\right),$$

where $\mathbf{0}$ *denotes a vector of zeros.*

The following lemma is helpful for proving Theorem C.2.

Lemma 3. *Under the assumptions of Theorem C.1, the class*

$$\mathcal{F} = \left\{\psi^\eta(Y,T,Z,X,\eta,\pi,\bar{m}) : \pi \in \Pi_z, \bar{m}_{j,z} \in \mathcal{M}_{j,z}, 1 \le j \le J, z \in \mathcal{Z}\right\}$$

is Donsker with a finite integrable envelope. The following stochastic equicontinuity condition hold: for any positive sequence $\delta_n = o(1)$,

$$\sup\left\{\Psi_n^\eta(\eta,\pi,\bar{m}) - \Psi^\eta(\eta,\pi,\bar{m}) - \Psi_n^\eta(\eta^o,\pi^o,m_Z^o) : \|\eta - \eta^o\|_2 \vee \|\pi - \pi^o\|_\infty \vee \|\bar{m} - \bar{m}^o\|_\infty \le \delta_n\right\} = o_p\left(n^{-1/2}\right),$$

where the supremum is taken over $\eta \in \Lambda$, $\pi_z \in \Pi_z$, *and* $\bar{m}_{j,z} \in \mathcal{M}_{j,z}$.

Proof of Lemma 3. We first verify that the moment condition ψ^η satisfies Condition (3.2) of Theorem 3 in Chen et al. (2003) (hereafter CLK). In fact, when $\|\bar{m}'_{j,z} - \bar{m}_{j,z}\|_\infty \vee \|\eta' - \eta\|_\infty \le$

δ, the triangle inequality gives that

$$\begin{aligned}
&\mathbb{E}\Big|\bar{m}'_{j,z}(X,\eta') - \bar{m}_{j,z}(X,\eta)\Big|^2 \\
\leq &2\mathbb{E}\Big|\bar{m}'_{j,z}(X,\eta') - \bar{m}'_{j,z}(X,\eta)\Big|^2 + 2\mathbb{E}\Big|\bar{m}'_{j,z}(X,\eta) - \bar{m}_{j,z}(X,\eta)\Big|^2 \\
\leq &\mathit{const} \times \delta^2,
\end{aligned}$$

where we use the notation *const* to denote a generic constant that may have different values at each appearance. The last inequality follows from the assumption (ii). Similarly, we can verify that the remaining terms in ψ^η also satisfy the same condition. Therefore, ψ^η is locally uniformly L_2-continuous, that is,

$$\begin{aligned}
&\mathbb{E}\big[\sup\big\{\big|\psi^\eta(Y,T,Z,X,\eta',\pi',\bar{m}') - \psi^\eta(Y,T,Z,X,\eta,\pi,\bar{m})\big| : \\
&\qquad \big\|\eta'-\eta\big\| \vee \big\|\pi'-\pi\big\|_\infty \vee \big\|\bar{m}'-\bar{m}\big\|_\infty \leq \delta\big\}\big] \leq \mathit{const.} \times \delta^2.
\end{aligned}$$

Following the same steps as in the proof of Theorem 3 in CLK (p. 1607), we can show that the bracketing number of $\mathcal{F}$ is bounded by

$$\begin{aligned}
&N_{[]}\big(\varepsilon, \mathcal{F}, \|\cdot\|_{L_2}\big) \\
\leq &N(\varepsilon/\mathit{const}, \Lambda, \|\cdot\|) \times \prod_z N(\varepsilon/\mathit{const}, \Pi_z, \|\cdot\|) \times \prod_{j,z} N(\varepsilon/\mathit{const}, \mathcal{M}_{j,z}, \|\cdot\|).
\end{aligned}$$

Therefore, the bracketing entropy of class $\mathcal{F}$ is bounded by

$$\begin{aligned}
&\log N_{[]}\big(\varepsilon, \mathcal{F}, \|\cdot\|_{L_2}\big) \\
\leq &\mathit{const} \times \Big(\log N(\varepsilon/\mathit{const}, \Lambda, \|\cdot\|) \vee \max_z \log N(\varepsilon/\mathit{const}, \Pi_z, \|\cdot\|) \\
\vee &\max_{j,z} \log N(\varepsilon/\mathit{const}, \mathcal{M}_{j,z}, \|\cdot\|)\Big).
\end{aligned}$$

Under the assumption that Λ is compact and

$$\int_0^\infty \log N(\varepsilon, \Pi_z, \|\cdot\|) d\varepsilon, \int_0^\infty \log N(\varepsilon, \mathcal{M}_{j,z}, \|\cdot\|) d\varepsilon < \infty, \forall j, z,$$

we have that

$$\int_0^\infty \log N_{[]}\big(\varepsilon, \mathcal{F}, \|\cdot\|_{L_2}\big) d\varepsilon < \infty.$$

This implies that $\mathcal{F}$ is Donsker with a finite integrable envelope. Lastly, as stated in Lemma 1 of CLK, the asserted stochastic equicontinuity condition is implied by the fact that $\mathcal{F}$ is Donsker and ψ^η is L_2-continuous. $\square$

Proof of Theorem C.2. We follow the large sample theory in CLK and set $\theta = \eta$, $h = (\pi, \bar{m})$, $M(\theta, h) = \Psi^\eta(\eta, \pi, \bar{m})$, and $M_n(\theta, h) = \Psi^\eta_n(\eta, \pi, \bar{m})$.

We first use Theorem 1 in CLK to show the consistency of $\tilde{\eta}$. Condition (1.2) in CLK is satisfied because Λ is compact, and $\Psi^\eta(\eta, \pi^o, \bar{m}^o)$ has a unique zero and is continuous by our second condition in Theorem C.1. As for Condition (1.3) of CLK, we can easily see from the expression of Ψ that it is continuous with respect to $\bar{m}_{j,z}$ and π_z (since π_z is bounded away from zero), and the uniformity in η follows from the fact that $\mathbb{E}\left[m(Y^*, \eta)\right]$ is bounded as a function of η. Condition (1.4) of CLK is satisfied by the assumption of Theorem C.2. The uniform stochastic equicontinuity condition (1.5) of CLK is implied by Lemma 3. Therefore, $\tilde{\eta} = \eta^o + o_p(1)$.

We use Corollary 1 (which is based on Theorem 2) in CLK to show the consistency of $\hat{V}$ and the asymptotic normality of $\hat{\eta}$. Condition (2.2) in CLK is verified by the assumptions of Theorem C.1. Similar to the proof of Proposition 4.2, we can show that the moment condition Ψ^η, based on the EIF, satisfies the Neyman orthogonality condition for the nuisance parameters π and m_Z. In fact, for any j and z, we let $\pi^r_z = \pi^o_z(X) + r(\pi_z(X) - \pi^o_z(X))$ and $\bar{m}^r_{j,z}(X,\eta) = \bar{m}^o_{j,z}(X,\eta) + r\big(\bar{m}_{j,z}(X,\eta) - \bar{m}^o_{j,z}(X,\eta)\big)$. Then we have

$$
\begin{aligned}
&\frac{d}{dr}\mathbb{E}\left[\frac{\mathbf{1}\{Z=z\}}{\pi^r_z(X)}\Big(m_j(Y,\eta)\mathbf{1}\{T=t_j\} - \bar{m}^r_{j,z}(X,\eta)\Big) + \bar{m}^r_{j,z}(X,\eta)\right]\Bigg|_{r=0} \\
&= \mathbb{E}\Bigg[-\frac{\mathbf{1}\{Z=z\}}{\big(\pi^o_z(X)\big)^2}\big(\pi_z(X) - \pi^o_z(X)\big)\Big(m_j(Y,\eta)\mathbf{1}\{T=t_j\} - \bar{m}^o_{j,z}(X,\eta)\Big) \\
&\quad + \Big(\bar{m}^o_{j,z}(X,\eta) - \bar{m}_{j,z}(X,\eta)\Big)\left(\frac{\mathbf{1}\{Z=z\}}{\pi^o_z(X)} - 1\right)\Bigg] = 0,
\end{aligned}
$$

where we have applied the law of iterated expectations and used the fact that

$$
\mathbb{E}\left[\frac{\mathbf{1}\{Z=z\}}{\pi^o_z(X)}\Big(m_j(Y,\eta)\mathbf{1}\{T=t_j\} - \bar{m}^o_{j,z}(X,\eta)\Big)\Big| X\right] = 0.
$$

Thus, the path-wise derivative of Ψ^η with respect to $h = (\pi, \bar{m})$ is zero in any direction. Hence, Condition (2.3) of CLK is verified. Condition (2.4) in CLK directly follows from our assumptions of Theorem C.2. The stochastic equicontinuity condition (condition (2.6) in CLK) follows from Lemma 3. Lastly, condition (2.6) in CLK is verified using the central limit theorem since the path-wise derivative is zero. Due to the presence of $\hat{V}$, we also need the uniform convergence condition in Corollary 1 of CLK, which can be verified by using Lemma 3 and an application of Theorem 2.10.14 of van der Vaart and Wellner (1996).

Lastly, to show the consistency of $\hat{\Gamma}$, we only need to show that

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\eta}\hat{m}_{j,t_j,z}(X_i,\hat{\eta}) \xrightarrow{p} \mathbb{E}\left[\frac{\partial}{\partial\eta}\hat{m}_{j,z}(X,\eta^o)\right] = \frac{\partial}{\partial\eta}\mathbb{E}\left[\hat{m}_{j,z}(X,\eta^o)\right],$$

where the inequality follows from the differentiation under integral operation which holds under the last assumption of the theorem. The convergence in probability follows from the uniform convergence of $\frac{\partial}{\partial\eta}\hat{m}_{j,z}$ and the consistency of $\hat{\eta}$. Therefore, the desired convergence results follow. $\square$

References

Cattaneo, M. D. (2010). Efficient semiparametric estimation of multi-valued treatment effects under ignorability. *Journal of Econometrics 155*(2), 138–154.

Chen, X., O. Linton, and I. Van Keilegom (2003). Estimation of semiparametric models when the criterion function is not smooth. *Econometrica 71*(5), 1591–1608.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal 21*(1), C1–C68.

Heckman, J. J. and R. Pinto (2018a). Unordered monotonicity. *Econometrica 86*(1), 1–35.

Heckman, J. J. and R. Pinto (2018b). Web appendix for unordered monotonicity. *Econometrica 86*(1), 1–35.

Hong, H. and D. Nekipelov (2010a). Semiparametric efficiency in nonlinear late models. *Quantitative Economics 1*(2), 279–304.

Hong, H. and D. Nekipelov (2010b). Supplement to "semiparametric efficiency in nonlinear late models". *Quantitative Economics 1*(2), 279–304.

Kong, E., O. Linton, and Y. Xia (2010). Uniform bahadur representation for local polynomial estimates of m-regression and its application to the additive model. *Econometric Theory 26*(5), 1529–1564.

Masry, E. (1996). Multivariate local polynomial regression for time series:uniform strong consistency and rates. *Journal of Time Series Analysis 17*(6), 571–599.

Newey, W. K. (1990). Semiparametric efficiency bounds. *Journal of applied econometrics 5*(2), 99–135.

Newey, W. K. (1994). The asymptotic variance of semiparametric estimators. *Econometrica 62*(6), 1349–1382.

Pakes, A. and D. Pollard (1989). Simulation and the asymptotics of optimization estimators. *Econometrica 57*(5), 1027–1057.

van der Vaart, A. W. and J. A. Wellner (1996). *Weak convergence and empirical processes.* Springer: New York.