Supplementary Material: Proofs of Theorems 1 and 2

Proof of Theorem 1 (Consistency and Asymptotic Normality)

Let

$$\Psi_n(\beta) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{R_i}{\pi(Y_i, X_i, V_i)}\{Y_i - X_i^\top \beta - m(V_i)\}X_i + \frac{1-R_i}{\pi(Y_i, X_i, V_i)}\{m(V_i) - X_i^\top \beta\}X_i\right],$$

and let β_0 denote the true parameter value. The AIPW estimator $\hat{\beta}$ satisfies $\Psi_n(\hat{\beta}) = 0$.

Step 1: Unbiasedness of the estimating equation. Taking conditional expectation given (Y, X, V),

$$E\left[\frac{R}{\pi(Y, X, V)} \mid Y, X, V\right] = 1, \qquad E\left[\frac{1-R}{\pi(Y, X, V)} \mid Y, X, V\right] = \frac{1-\pi(Y, X, V)}{\pi(Y, X, V)}.$$

Under either correct specification of $\pi(Y, X, V)$ or of $E(Y \mid X, V) = X^\top \beta_0 + m(V)$, straightforward algebra yields

$$E\{\Psi_n(\beta_0)\} = 0,$$

establishing consistency under the double robustness condition (A3).

Step 2: Linearization. By a first-order Taylor expansion around β_0,

$$0 = \Psi_n(\hat{\beta}) = \Psi_n(\beta_0) - A(\hat{\beta} - \beta_0) + o_p(\|\hat{\beta} - \beta_0\|),$$

where

$$A = E\left[\frac{\partial}{\partial \beta}\Psi_n(\beta_0)\right] = E\left[(X - \mu_X(V))(X - \mu_X(V))^\top\right].$$

Assumption A4 implies that A is positive definite.

Step 3: Asymptotic distribution. Define the influence function

$$\psi(Z) = (X - \mu_X(V))\left[\frac{R}{\pi(Y, X, V)}\varepsilon + \frac{1-R}{\pi(Y, X, V)}\varepsilon\right], \quad \varepsilon = Y - X^\top \beta_0 - m(V).$$

Then

$$\sqrt{n}(\hat{\beta} - \beta_0) = A^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi(Z_i) + o_p(1).$$

By the multivariate central limit theorem,

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, A^{-1}BA^{-1}), \qquad B = E\{\psi(Z)\psi(Z)^\top\}.$$

This completes the proof. $\square$

Proof of Theorem 2 (Convergence Rates)

Rate for $\hat{m}(V)$. Under Assumptions A5–A6, standard results for local linear kernel regression yield

$$\sup_{v \in \mathcal{V}} |\hat{m}(v) - m(v)| = O_p\Big(h_n^2 + (nh_n^q)^{-1/2}\Big).$$

Rate for $\hat{\beta}$. From the expansion in the proof of Theorem 1,

$$\hat{\beta} - \beta_0 = A^{-1}\Psi_n(\beta_0) + o_p(n^{-1/2}).$$

The estimating equation is orthogonal to first-order perturbations in $m(V)$, so the nonparametric estimation error enters only at higher order. Consequently,

$$\|\Psi_n(\beta_0)\| = O_p(n^{-1/2}),$$

which implies

$$\hat{\beta} - \beta_0 = O_p(n^{-1/2}).$$

Thus,

$$\hat{\beta} - \beta_0 = O_p(n^{-1/2}), \qquad \sup_{v \in \mathcal{V}} |\hat{m}(v) - m(v)| = O_p\Big(h_n^2 + (nh_n^q)^{-1/2}\Big).$$

$\square$