Appendix

Throughout the appendix, we assume that $h = h(n)$ is a sequence such that as $n \to \infty$, $h \to 0$, and $nh \to \infty$. We also assume that z is an interior point of the support of Z. We assume the following regularity conditions:

i) $\theta(\cdot)$ and $f_Z(\cdot)$ satisfy the smoothness assumptions of Fan, et. al. (1995);

ii) The estimating functions in the right hand side of naive kernel estimating equations, IPW kernel estimating equations, and AIPW kernel estimating equations are twice continuously differentiable with respect to $\boldsymbol{\alpha}$ at a target point z, and the second derivatives are uniformly bounded.

A.1 Sketch of the Proof of Theorem 1

If $\mu^{(1)}\{\widetilde{\theta}_{naive}(z)\} \neq 0$, simple calculations show that the solution of equation (7) for $\widetilde{\theta}_{naive}(z)$ is $\mu\{\widetilde{\theta}_{naive}(z)\} = E(RY|Z=z)/E(R|Z=z)$, which is equal to $cov(R,Y|Z=z)/E(R|Z=z) + \mu\{\theta(z)\}$. This gives the expression for $\widetilde{\theta}_{naive}(z)$ stated in the theorem.

Next study the expression of $\widetilde{\theta}_{IPW}(z)$. The left hand side of (8) is equal to

$$E\left[\frac{E(R|Y,Z,\boldsymbol{U})}{\widetilde{\pi}}\mu^{(1)}\{\widetilde{\theta}_{IPW}(z)\}V^{-1}\{\widetilde{\theta}_{IPW}(z);\widetilde{\boldsymbol{\zeta}}\}\left[Y-\mu\{\widetilde{\theta}_{IPW}(z)\}\right]\middle| Z=z\right]$$

by taking a double expectation given Y, Z and $\boldsymbol{U}$. If model (3) of π is correctly specified, then $\widetilde{\pi} = E(R|Z,\boldsymbol{U})$. Also under MAR, $E(R|Y,Z,\boldsymbol{U}) = E(R|Z,\boldsymbol{U})$. Therefore the above quantity equals to $E[\mu^{(1)}\{\widetilde{\theta}_{IPW}(z)\} \times V^{-1}\{\widetilde{\theta}_{IPW}(z);\widetilde{\boldsymbol{\zeta}}\}[Y-\mu\{\widetilde{\theta}_{IPW}(z)\}]|Z=z]$. If $\mu^{(1)}\{\widetilde{\theta}_{IPW}(z)\} \neq 0$, solving for $\widetilde{\theta}_{IPW}(z)$ yields $\mu\{\widetilde{\theta}_{IPW}(z)\} = E[Y|Z=z] = \mu\{\theta(z)\}$. Therefore, $\widehat{\theta}_{IPW}(z)$ is a consistent estimator of $\theta(z)$ when model (3) of π is correctly specified or π_0 is known by design.

Now study the expression of $\widetilde{\theta}_{AIPW}(z)$ from (9). Under the MAR assumption (2), the left hand side of (9) can be rewritten as

$$\begin{aligned} & E\left[\mu^{(1)}\{\widetilde{\theta}_{AIPW}(z)\}V^{-1}\{\widetilde{\theta}_{AIPW}(z);\widetilde{\boldsymbol{\zeta}}\}\left[Y-\mu\{\widetilde{\theta}_{AIPW}(z)\}\right]|Z=z\right] \\ + \; & E\left[\left(\frac{R}{\widetilde{\pi}}-1\right)\mu^{(1)}\{\widetilde{\theta}_{AIPW}(z)\}V^{-1}\{\widetilde{\theta}_{AIPW}(z);\widetilde{\boldsymbol{\zeta}}\}\left[Y-\widetilde{\delta}(Z,\boldsymbol{U})\}\right]|Z=z\right] = 0. \quad \text{(A.1)} \end{aligned}$$

If model (3) for π is correctly specified, i.e., $\tilde{\pi} = E(R|Z,\boldsymbol{U})$, or model (6) for $\delta(\cdot)$ is correctly specified, i.e., $\widetilde{\delta}(Z,\boldsymbol{U}) = E(Y|Z,\boldsymbol{U})$, one can easily see that the second term of (A.1) is 0. Hence (A.1) is equal to

$$E\left[\mu^{(1)}\{\widetilde{\theta}_{AIPW}(z)\}V^{-1}\{\widetilde{\theta}_{AIPW}(z);\widetilde{\boldsymbol{\zeta}}\}\left[Y-\mu\{\widetilde{\theta}_{AIPW}(z)\}\right]|Z=z\right] = 0.$$

It follows that if $\mu^{(1)}\{\widetilde{\theta}_{AIPW}(z)\} \neq 0$, we have $\widetilde{\theta}_{AIPW}(z) = \theta(z)$, i.e., $\widehat{\theta}_{AIPW}(z)$ is a consistent estimator of $\theta(z)$.

A.2 Proof of Theorem 2: Asymptotic Bias and Variance of the IPW Estimator

We first assume that π_0 is known by design and prove that the asymptotic distribution of $\hat{\theta}_{IPW}(z)$ is given in (10). We also assume that the variance parameter $\boldsymbol{\zeta}$ in the working variance V is known. We will then extend the results when π and $\boldsymbol{\zeta}$ are estimated. For any interior point z, reparameterize $\boldsymbol{\alpha}$ as $\left\{\theta\left(z\right), h\theta'\left(z\right)\right\}^T$ and denote by $\theta_0(z)$ the true value of $\theta(z)$, $\boldsymbol{\alpha}_0 = \{\theta_0(z), h\theta_0'(z)\}^T$ and $\hat{\boldsymbol{\alpha}}_{IPW}(z)$ the solution of the local linear IPW kernel estimating equations. A Taylor expansion of the local linear IPW kernel estimating equations gives

$$\sqrt{nh}\{\boldsymbol{\alpha}_{IPW}(z) - \boldsymbol{\alpha}_0\} = -\sqrt{nh}\,\{\boldsymbol{\Gamma}_n(\boldsymbol{\alpha}_*)\}^{-1}\,\boldsymbol{\Lambda}_n(\boldsymbol{\alpha}_0),$$

where $\boldsymbol{\alpha}_*$ is between $\hat{\boldsymbol{\alpha}}_{IPW}(z)$ and $\boldsymbol{\alpha}_0$, and

$$\boldsymbol{\Lambda}_n(\boldsymbol{\alpha}) = n^{-1}\sum_{i=1}^{n} R_i\pi_{i0}^{-1}(Z_i, \boldsymbol{U}_i)K_h(Z_i - z)\mu_i^{(1)}(z, \boldsymbol{\alpha})V_i^{-1}(z, \boldsymbol{\alpha})\boldsymbol{G}(Z_i - z)[Y_i - \mu\{\boldsymbol{G}(Z_i - z)^T\boldsymbol{\alpha}\}],$$

where $\mu_i^{(1)}(z, \boldsymbol{\alpha}) = \mu^{(1)}\{\boldsymbol{G}(Z_i - z)^T\boldsymbol{\alpha}\}$ and $V_i(z, \boldsymbol{\alpha}) = V[\mu\{\boldsymbol{G}(Z_i - z)^T\boldsymbol{\alpha}\}; \boldsymbol{\zeta}_0]$, $\boldsymbol{\Gamma}_n(\boldsymbol{\alpha}) = \partial\boldsymbol{\Lambda}_n(\boldsymbol{\alpha})/\partial\boldsymbol{\alpha}^T$.

Using the results in Appendix A.1, we have $\hat{\boldsymbol{\alpha}}_{IPW}(z) \to \boldsymbol{\alpha}_0$ in probability. Therefore, $\boldsymbol{\alpha}_* \xrightarrow{P} \boldsymbol{\alpha}_0$. Under the MAR assumption (2), simple calculations show that

$$\begin{aligned}\boldsymbol{\Gamma}_n(\boldsymbol{\alpha}_*) &= -E\left[K_h(Z - z)\left\{\mu^{(1)}(z, \boldsymbol{\alpha}_0)\right\}^2 V^{-1}(z, \boldsymbol{\alpha}_0)\boldsymbol{G}(Z - z)\boldsymbol{G}(Z - z)^T\right] + o_p(1)\\ &= -f_Z(z)\left(\mu^{(1)}\{\theta(z)\}\right)^2 V^{-1}\{\theta(z)\}\boldsymbol{D}(K) + o_p(1)\end{aligned}$$

where $\boldsymbol{D}\left(K\right)$ is a 2×2 matrix with the (j,k)th element $c_{j+k-2}(K) \times h^{(j+k-2)}$, and $c_r(K) = \int s^r K(s)ds$. It follows that

$$\sqrt{nh}\{\hat{\boldsymbol{\alpha}}_{IPW}(z) - \boldsymbol{\alpha}_0\} = \left\{f_Z(z)\left[\mu^{(1)}\{\theta(z)\}\right]^2 V^{-1}\{\theta(z)\}\boldsymbol{D}(K)\right\}^{-1}\sqrt{nh}\boldsymbol{\Lambda}_n(\boldsymbol{\alpha}_0) + o_p(1). \quad \text{(A.2)}$$

Now write $\boldsymbol{\Lambda}_n(\boldsymbol{\alpha}_0) = \boldsymbol{\Lambda}_{1n}(\boldsymbol{\alpha}_0) + \boldsymbol{\Lambda}_{2n}(\boldsymbol{\alpha}_0)$, where

$$\begin{aligned}\boldsymbol{\Lambda}_{1n}(\boldsymbol{\alpha}_0) &= n^{-1}\sum_{i=1}^{n} R_i\pi_{i0}^{-1}(Z_i, \boldsymbol{U}_i)K_h(Z_i - z)\mu_i^{(1)}(z, \boldsymbol{\alpha}_0)V_i^{-1}(z, \boldsymbol{\alpha}_0)\boldsymbol{G}(Z_i - z)[Y_i - \mu\{\theta(Z_i)\}]\\ \boldsymbol{\Lambda}_{2n}(\boldsymbol{\alpha}_0) &= n^{-1}\sum_{i=1}^{n} R_i\pi_{i0}^{-1}(Z_i, \boldsymbol{U}_i)K_h(Z_i - z)\mu_i^{(1)}(z, \boldsymbol{\alpha}_0)V_i^{-1}(z, \boldsymbol{\alpha}_0)\boldsymbol{G}(Z_i - z)[\mu\{\theta(Z_i)\} - \mu\{\boldsymbol{G}(Z_i - z)^T\boldsymbol{\alpha}_0\}].\end{aligned}$$

One can easily show that $\boldsymbol{\Lambda}_{1n}\left(\boldsymbol{\alpha}_0\right)$ is asymptotically normal with mean zero and asymptotic variance

$$\begin{aligned}var\{\boldsymbol{\Lambda}_{1n}(\boldsymbol{\alpha}_0)\} &= \frac{1}{n}E\left[K_h^2(Z - z)\left\{\mu^{(1)}(z, \boldsymbol{\alpha}_0)\right\}^2 V^{-2}(z, \boldsymbol{\alpha}_0)\left(\frac{R\left[Y - \mu\left\{\theta(Z)\right\}\right]}{\pi_0(Z, \boldsymbol{U})}\right)^2 \boldsymbol{G}(Z - z)\boldsymbol{G}(Z - z)^T\right]\\ &= \frac{1}{nh}f_Z(z)\left(\mu^{(1)}\{\theta(z)\}\right)^2 V^{-2}\{\theta(z)\}E\left[\left(\frac{R\left[Y - \mu\left\{\theta(Z)\right\}\right]}{\pi_0(Z, \boldsymbol{U})}\right)^2\middle| Z = z\right]\boldsymbol{D}(K^2) + \boldsymbol{o}(\frac{1}{nh}),\end{aligned}$$

where $\boldsymbol{D}(K^2)$ is defined similarly to $\boldsymbol{D}(K)$ with K replaced by K^2.

Now study $\boldsymbol{\Lambda}_{2n}$, which contributes to the leading bias term. One can easily show under MAR, we have

$$\begin{aligned} bias\left\{\boldsymbol{\Lambda}_{2n}(\boldsymbol{\alpha}_0)\right\} &= E\left\{K_h(Z-z)\mu^{(1)}(z,\boldsymbol{\alpha}_0)V^{-1}(z,\boldsymbol{\alpha}_0)\left[\mu\left\{\theta(Z)\right\}-\mu\left\{\boldsymbol{G}(Z-z)^T\boldsymbol{\alpha}_0\right\}\right]\boldsymbol{G}(Z-z)\right\}+o_p(1) \\ &= \frac{1}{2}\theta''(z)\left[\mu^{(1)}\{\theta(z)\}\right]^2 V^{-1}\{\theta(z)\}f_Z(z)\boldsymbol{H}(K)+o(h^2), \qquad \text{(A.3)} \end{aligned}$$

where $\boldsymbol{H}(K)$ is a 2×1 vector with the kth element $c_{k+1}(K)\times h^{(k+1)}$. Note that the asymptotic variance of $\boldsymbol{\Lambda}_{2n}$ is of order $\boldsymbol{o}(1/nh)$ and is asymptotically negligible compared to $\boldsymbol{\Lambda}_{1n}$, and the asymptotic covariance of $\boldsymbol{\Lambda}_{1n}$ and $\boldsymbol{\Lambda}_{2n}$ is $\mathbf{0}$. Applying these results to (A.2), simple calculations show that the asymptotic distribution of the IPW estimator $\hat{\theta}_{IPW}(z;\pi)$, the first element of $\hat{\boldsymbol{\alpha}}_{IPW}$, is given in (10).

We next study the distribution of $\hat{\theta}_{IPW}\{z;\pi(\hat{\boldsymbol{\tau}})\}$ when π_0 is estimated consistently at the $\sqrt{n}$-rate, i.e. $\sqrt{n}(\hat{\boldsymbol{\tau}}-\boldsymbol{\tau}_0)=O_p(1)$, where $\boldsymbol{\tau}_0$ is the true value of $\boldsymbol{\tau}$. Suppose under some regularity conditions, $\partial\hat{\theta}_{IPW}\{z;\pi(\boldsymbol{\tau})\}/\partial\boldsymbol{\tau}^T$ is bounded in the neighborhood of the $\boldsymbol{\tau}_0$, *i.e.*,

$$\partial\hat{\theta}_{IPW}\{z;\pi(\boldsymbol{\tau})\}/\partial\boldsymbol{\tau}^T|_{\boldsymbol{\tau}\in\mathcal{N}(\boldsymbol{\tau}_0)}=O_p(1),$$

where $\mathcal{N}(\boldsymbol{\tau}_0)\supset\{\boldsymbol{\tau}:||\boldsymbol{\tau}-\boldsymbol{\tau}_0||<||\hat{\boldsymbol{\tau}}-\boldsymbol{\tau}_0||\}$. We have

$$\begin{aligned} &\sqrt{nh}[\hat{\theta}_{IPW}\{z;\pi(\hat{\boldsymbol{\tau}})\}-\theta(z)] \\ &=\sqrt{nh}[\hat{\theta}_{IPW}\{z;\pi(\hat{\boldsymbol{\tau}})\}-\hat{\theta}_{IPW}\{z;\pi(\tau_0)\}]+\sqrt{nh}[\hat{\theta}_{IPW}\{z;\pi(\tau_0)\}-\theta(z)] \\ &=\sqrt{h}\left[\frac{\partial\hat{\theta}_{IPW}\{z;\pi(\boldsymbol{\tau})\}}{\partial\boldsymbol{\tau}^T}|_{\boldsymbol{\tau}^*}\right]\sqrt{n}(\hat{\boldsymbol{\tau}}-\boldsymbol{\tau}_0)+\sqrt{nh}[\hat{\theta}_{IPW}\{z;\pi(\tau_0)\}-\theta(z)] \qquad \text{(A.4)} \end{aligned}$$

for some $\boldsymbol{\tau}^*\in\{\boldsymbol{\tau}:||\boldsymbol{\tau}-\boldsymbol{\tau}_0||<||\hat{\boldsymbol{\tau}}-\boldsymbol{\tau}_0||\}$. Note $\sqrt{n}(\hat{\boldsymbol{\tau}}-\boldsymbol{\tau}_0)=O_p(1)$, $\partial\hat{\theta}_{IPW}\left\{z;\pi(\boldsymbol{\tau})\right\}/\partial\boldsymbol{\tau}^T|_{\boldsymbol{\tau}^*}=O_p(1)$, and $h\to 0$ as $n\to\infty$, the first term in (A.4) is $o_p(1)$. Therefore, the asymptotic distribution of $\hat{\theta}_{IPW}\{z;\pi(\hat{\boldsymbol{\tau}})\}$ when $\boldsymbol{\tau}$ is estimated consistently at $\sqrt{n}$-rate is the same as that of $\hat{\theta}_{IPW}(z;\pi_0)$ when π_0 is known. Similar argument shows that the asymptotic distribution of $\hat{\theta}_{IPW}\{z\}$ remains the same if $\boldsymbol{\zeta}$ is estimated at the $\sqrt{n}$-rate.

A.3 Proof of Theorem 3: Asymptotic Bias and Asymptotic Variance of AIPW estimator

Following similar arguments as those in Appendix A.2, the asymptotic results hold when the parameters $(\boldsymbol{\tau},\boldsymbol{\eta})$ in π and δ are estimated at the $\sqrt{n}$-rate, or the probability limit of $(\widehat{\boldsymbol{\tau}},\widehat{\boldsymbol{\eta}})$ is used in the AIPW kernel estimating equations (4). Denote by $(\widetilde{\boldsymbol{\tau}},\widetilde{\boldsymbol{\eta}})$ the probability limit of $(\widehat{\boldsymbol{\tau}},\widehat{\boldsymbol{\eta}})$, and let $\widetilde{\pi}(Z_i,\boldsymbol{U}_i)=\pi(Z_i,\boldsymbol{U}_i;\widetilde{\boldsymbol{\tau}})$, $\widetilde{\delta}\left(Z_i,\boldsymbol{U}_i\right)=\delta\left(Z_i,\boldsymbol{U}_i;\widetilde{\boldsymbol{\eta}}\right)$. We focus our proof on assuming that $(\widetilde{\boldsymbol{\tau}},\widetilde{\boldsymbol{\eta}})$ are known. By a linear Taylor expansion of the AIPW estimating function (4) about $\boldsymbol{\alpha}_0$, the AIPW kernel estimator satisfies

$$\sqrt{nh}\{\hat{\boldsymbol{\alpha}}_{AIPW}(z)-\boldsymbol{\alpha}_0\}=-\sqrt{nh}\left\{\boldsymbol{\Gamma}_{n,\delta}(\boldsymbol{\alpha}_*)\right\}^{-1}\boldsymbol{\Lambda}_{n,\delta}(\boldsymbol{\alpha}_0),$$

where $\boldsymbol{\alpha}_*$ is between $\hat{\boldsymbol{\alpha}}_{AIPW}(z)$ and $\boldsymbol{\alpha}_0$,

$$\begin{aligned}\boldsymbol{\Lambda}_{n,\delta}(\boldsymbol{\alpha}) &= n^{-1}\sum_{i=1}^{n}\Big\{R_i\widetilde{\pi}^{-1}(Z_i,\boldsymbol{U}_i)K_h(Z_i-z)\mu_i^{(1)}(z,\boldsymbol{\alpha})V_i^{-1}(z,\boldsymbol{\alpha})\boldsymbol{G}(Z_i-z)\left[Y_i-\mu\left\{\boldsymbol{G}(Z_i-z)^T\boldsymbol{\alpha}\right\}\right] \\ &\quad -\left\{R_i\widetilde{\pi}^{-1}(Z_i,\boldsymbol{U}_i)-1\right\}K_h(Z_i-z)\mu_i^{(1)}(z,\boldsymbol{\alpha})V_i^{-1}(z,\boldsymbol{\alpha})\boldsymbol{G}(Z_i-z)\left[\widetilde{\delta}(Z_i,\boldsymbol{U}_i)-\mu\left\{\boldsymbol{G}(Z_i-z)^T\boldsymbol{\alpha}\right\}\right]\Big\},\end{aligned}$$

and $\boldsymbol{\Gamma}_{n,\delta}(\boldsymbol{\alpha}) = \partial\boldsymbol{\Lambda}_{n,\delta}(\boldsymbol{\alpha})/\partial\boldsymbol{\alpha}^T$.

We consider the following two situations:

(1) When model (3) for the selection probability π_{i0} is correctly specified, i.e. $\widetilde{\pi}(Z_i,\boldsymbol{U}_i) = \pi_{i0}(Z_i,\boldsymbol{U}_i)$;

(2) When model (6) for $E(Y|Z,\boldsymbol{U})$ is correctly specified, i.e. $\widetilde{\delta}(Z_i,\boldsymbol{U}_i) = E(Y_i|Z_i,\boldsymbol{U}_i)$.

As shown in Appendix A.1, $\hat{\boldsymbol{\alpha}}_{AIPW}(z)$ converges to $\boldsymbol{\alpha}_0$ when either of the above conditions holds. Therefore, $\boldsymbol{\alpha}_* \xrightarrow{P} \boldsymbol{\alpha}_0$. We first show that under either of the above situations, we have

$$\boldsymbol{\Gamma}_{n,\delta}(\boldsymbol{\alpha}_*) \xrightarrow{P} -f_Z(z)\left[\mu^{(1)}\{\theta(z)\}\right]^2 V^{-1}\{\theta(z)\}\boldsymbol{D}(K). \tag{A.5}$$

First consider situation (1), i.e., when $\widetilde{\pi}(Z_i,\boldsymbol{U}_i) = \pi_{i0}(Z_i,\boldsymbol{U}_i)$. The second term of $\boldsymbol{\Lambda}_{n,\delta}(\boldsymbol{\alpha})$, i.e. the augmentation term, has mean 0 under MAR . It follows that $\boldsymbol{\Lambda}_{n,\delta}(\boldsymbol{\alpha}_*) = \boldsymbol{\Lambda}_n(\boldsymbol{\alpha}_0)+\boldsymbol{o}_p(1)$, where $\boldsymbol{\Lambda}_n$ is defined in Appendix A.2. Hence $\boldsymbol{\Gamma}_{n,\delta}(\boldsymbol{\alpha}_*) = \boldsymbol{\Gamma}_n(\boldsymbol{\alpha}_0)+\boldsymbol{o}_p(1)$. Therefore $\boldsymbol{\Gamma}_{n,\delta}(\boldsymbol{\alpha}_*)$ has the same probability limit as $\boldsymbol{\Gamma}_n(\boldsymbol{\alpha}_*)$. As shown in Appendix A.2, the probability of limit of $\boldsymbol{\Gamma}_n(\boldsymbol{\alpha}_*)$ is exactly the right hand side of $(A.5)$, and thus $(A.5)$ holds for $\boldsymbol{\Gamma}_{n,\delta}(\boldsymbol{\alpha}_*)$ as well.

Next consider situation (2), i.e., when $\widetilde{\delta}(Z_i,\boldsymbol{U}_i) = E(Y_i|Z_i,\boldsymbol{U}_i)$. Rewrite $\boldsymbol{\Lambda}_{n,\delta}(\boldsymbol{\alpha})$ as

$$\begin{aligned}\boldsymbol{\Lambda}_{n,\delta}(\boldsymbol{\alpha}) &= n^{-1}\sum_{i=1}^{n}\Big\{R_i\widetilde{\pi}^{-1}(Z_i,\boldsymbol{U}_i)K_h(Z_i-z)\mu_i^{(1)}(z,\boldsymbol{\alpha})V_i^{-1}(z,\boldsymbol{\alpha})\boldsymbol{G}(Z_i-z)\left[Y_i-\widetilde{\delta}(Z_i,\boldsymbol{U}_i)\right] \\ &\quad + K_h(Z_i-z)\mu_i^{(1)}(z,\boldsymbol{\alpha})V_i^{-1}(z,\boldsymbol{\alpha})\boldsymbol{G}(Z_i-z)\left[\widetilde{\delta}(Z_i,\boldsymbol{U}_i)-\mu\left\{\boldsymbol{G}(Z_i-z)^T\boldsymbol{\alpha}\right\}\right]\Big\}.\end{aligned}$$

One can easily see the first term on the right hand side has mean 0. It follows that

$$\boldsymbol{\Lambda}_{n,\delta}(\boldsymbol{\alpha}_*) = n^{-1}\sum_{i=1}^{n}K_h(Z_i-z)\mu_i^{(1)}(z,\boldsymbol{\alpha}_0)V_i^{-1}(z,\boldsymbol{\alpha})\boldsymbol{G}(Z_i-z)\left[E\left(Y_i|Z_i,\boldsymbol{U}_i\right)-\mu\left\{\boldsymbol{G}(Z_i-z)^T\boldsymbol{\alpha}_0\right\}\right]+\boldsymbol{o}_p(1).$$

Differentiating it with respect to $\boldsymbol{\alpha}$ shows that $\boldsymbol{\Gamma}_{n,\delta}(\boldsymbol{\alpha}_*) = \boldsymbol{\Gamma}_n(\boldsymbol{\alpha}_0)+\boldsymbol{o}_p(1)$. Therefore, (A.5) still holds in this situation.

Therefore, when either the π or δ model is correctly specified, we have

$$\sqrt{nh}\{\hat{\boldsymbol{\alpha}}_{AIPW}(z)-\boldsymbol{\alpha}_0\} = \left\{f_Z(z)\left[\mu^{(1)}\{\theta(z)\}\right]^2 V^{-1}\{\theta(z)\}\boldsymbol{D}(K)\right\}^{-1}\sqrt{nh}\boldsymbol{\Lambda}_{n,\delta}(\boldsymbol{\alpha}_0)+o_p(1). \tag{A.6}$$

Write $\boldsymbol{\Lambda}_{n,\delta}(\boldsymbol{\alpha}_0) = \boldsymbol{\Lambda}_{1n,\delta}(\boldsymbol{\alpha}_0)-\boldsymbol{\Lambda}_{2n,\delta}(\boldsymbol{\alpha}_0)+\boldsymbol{\Lambda}_{3n,\delta}(\boldsymbol{\alpha}_0)$, where

$$\boldsymbol{\Lambda}_{1n,\delta}(\boldsymbol{\alpha}_0) = n^{-1}\sum_{i=1}^{n}R_i\widetilde{\pi}^{-1}(Z_i,\boldsymbol{U}_i)K_h(Z_i-z)\mu_i^{(1)}(z,\boldsymbol{\alpha}_0)V_i^{-1}(z,\boldsymbol{\alpha}_0)\left[Y_i-\mu\{\theta(Z_i)\}\right]\boldsymbol{G}(Z_i-z),$$

$$\mathbf{\Lambda}_{2n,\delta}\left(\boldsymbol{\alpha}_0\right) = n^{-1}\sum_{i=1}^{n}\{R_i\widetilde{\pi}^{-1}(Z_i,\boldsymbol{U}_i) - 1\}K_h(Z_i - z)\mu_i^{(1)}(z,\boldsymbol{\alpha}_0)V_i^{-1}(z,\boldsymbol{\alpha}_0)\left[\widetilde{\delta}(Z_i,\boldsymbol{U}_i) - \mu\{\theta(Z_i)\}\right]\boldsymbol{G}(Z_i - z),$$

and

$$\mathbf{\Lambda}_{3n,\delta}\left(\boldsymbol{\alpha}_0\right) = n^{-1}\sum_{i=1}^{n}K_h(Z_i - z)\mu_i^{(1)}(z,\boldsymbol{\alpha}_0)V_i^{-1}(z,\boldsymbol{\alpha}_0)\left[\mu\{\theta(Z_i)\} - \mu\{\boldsymbol{G}(Z_i - z)^T\boldsymbol{\alpha}_0\}\right]\boldsymbol{G}(Z_i - z).$$

One can easily see that $\mathbf{\Lambda}_{1n,\delta}(\boldsymbol{\alpha}_0)$ and $\mathbf{\Lambda}_{2n,\delta}(\boldsymbol{\alpha}_0)$ have mean 0 when either π or δ is correctly specified. The third term $\mathbf{\Lambda}_{3n,\delta}(\boldsymbol{\alpha}_0)$ is the leading bias term. When π_i or δ_i is correctly specified, simple calculations show that $E\left[\mathbf{\Lambda}_{3n,\delta}(\boldsymbol{\alpha}_0)\right]$ is equal to (A.3). It follows that

$$bias\{\hat{\boldsymbol{\alpha}}_{AIPW}(z)\} = \frac{1}{2}h^2\theta''(z)c_2(K) + o(h^2).$$

Now study $\mathbf{\Lambda}_{1n,\delta} - \mathbf{\Lambda}_{2n,\delta}$, which contributes to the leading variance and asymptotic normality. Note that the variance of $\mathbf{\Lambda}_{3n,\delta}\left(\boldsymbol{\alpha}_0\right)$ is of order $o(1/nh)$, and hence can be ignored asymptotically. Under MAR, we have $E[R|Y,Z,\boldsymbol{U}] = E[R|Z,\boldsymbol{U}] = \pi_0(Z,\boldsymbol{U})$, the true conditional mean of $[R|Z,\boldsymbol{U}]$. It follows that when either π or δ is correctly specified, $\mathbf{\Lambda}_{1n,\delta}(\boldsymbol{\alpha}_0) - \mathbf{\Lambda}_{2n,\delta}(\boldsymbol{\alpha}_0)$ is asymptotically normal with mean 0 and variance

$$var\left\{\mathbf{\Lambda}_{1n,\delta}(\boldsymbol{\alpha}_0) - \mathbf{\Lambda}_{2n,\delta}(\boldsymbol{\alpha}_0)\right\} = \frac{1}{n}\left[var\left\{\mathbf{\Lambda}_{1,2,\delta}(\boldsymbol{\alpha}_0)\right\}\right],$$

where

$$\begin{aligned}\mathbf{\Lambda}_{1,2,\delta}(\boldsymbol{\alpha}_0) &= K_h(Z - z)\mu^{(1)}(z,\boldsymbol{\alpha}_0)V^{-1}(z,\boldsymbol{\alpha}_0)\boldsymbol{G}(Z - z)\\ &\times\left(\frac{R}{\widetilde{\pi}(Z,\boldsymbol{U})}\left[Y - \mu\{\theta(Z)\}\right] - \left\{\frac{R}{\widetilde{\pi}(Z,\boldsymbol{U})} - 1\right\}\left[\widetilde{\delta}(Z,\boldsymbol{U}) - \mu\{\theta(Z)\}\right]\right)\end{aligned}$$

Further calculations show that

$$\begin{aligned}\frac{1}{n}var\left\{\mathbf{\Lambda}_{1,2,\delta}(\boldsymbol{\alpha}_0)\right\} &= \frac{1}{n}E\Bigg[K_h^2(Z - z)\left\{\mu^{(1)}(z,\boldsymbol{\alpha}_0)\right\}^2V^{-2}(z,\boldsymbol{\alpha}_0)\boldsymbol{G}(Z - z)\boldsymbol{G}(Z - z)^T\\ &\times\left(\frac{R}{\widetilde{\pi}(Z,\boldsymbol{U})}\left[Y - \mu\{\theta(Z)\}\right] - \left\{\frac{R}{\widetilde{\pi}(Z,\boldsymbol{U})} - 1\right\}\left[\widetilde{\delta}(Z,\boldsymbol{U}) - \mu\{\theta(Z)\}\right]\right)^2\Bigg]\\ &= \frac{1}{nh}f_Z(z)\left[\mu^{(1)}\{\theta(z)\}\right]^2V^{-2}\{\theta(z)\}E\Bigg[\Bigg(\frac{R}{\widetilde{\pi}(Z,\boldsymbol{U})}\left[Y - \mu\{\theta(Z)\}\right]\\ &- \left\{\frac{R}{\widetilde{\pi}(Z,\boldsymbol{U})} - 1\right\}\left[\widetilde{\delta}(Z,\boldsymbol{U}) - \mu\{\theta(Z)\}\right]\Bigg)^2|Z = z\Bigg]\boldsymbol{D}(K^2) + \boldsymbol{o}(\frac{1}{nh})\end{aligned}$$

Applying these results to (A.6) and Theorem 3 follows.