Online Appendix to "Covariate Adjustment in Stratified Experiments"

Max Cytrynbaum

A Appendix

A.1 Experiments with Noncompliance

In this section, we extend our main results to the case of experiments with imperfect compliance. The theorems in this section are simple corollaries of our main results. For completeness, full proofs are provided in Section A.9.

Previously, Ansel et al. (2018) studied covariate adjustment in experiments with noncompliance and iid or coarsely stratified treatment assignment. Bai et al. (2023a) study matched pairs experiments with noncompliance. See also Jiang et al. (2023) and Ren (2023) for nonlinear adjustment in coarsely stratified experiments and completely randomized experiments with noncompliance, respectively.

Let $z \in \{0,1\}$ denote a binary instrument. Let $D(z)$ be the potential treatments and $Y(d,z) = Y(d)$ the potential outcomes, satisfying exclusion. Define the intention-to-treat (ITT) potential outcomes $W_i(z) = Y_i(D_i(z))$, so that $Y_i = Z_i W_i(1) + (1-Z_i)W_i(0)$ and $D_i = Z_i D_i(1) + (1-Z_i)D_i(0)$. Impose monotonicity $D(1) \geq D(0)$ and positive compliance $\tau_D = P(D(1) > D(0)) > 0$. Define the ITT effect $\tau_W = E[W(1) - W(0)]$. Under these assumptions, the parameter $\tau_L \equiv \tau_W/\tau_D = E[Y(1) - Y(0)|D(1) > D(0)]$ is the local average treatment effect (LATE) (Imbens and Angrist (1994)). To estimate τ_L, we consider adjusted Wald estimators of the form

$$\widehat{\tau}_{adj} = \frac{\bar{W}_1 - \bar{W}_0 - \widehat{\gamma}_W'(\bar{h}_1 - \bar{h}_0)c_p}{\bar{D}_1 - \bar{D}_0 - \widehat{\gamma}_D'(\bar{h}_1 - \bar{h}_0)c_p} \tag{A.1}$$

To analyze $\widehat{\tau}_{adj}$, we require that Assumption 3.1 holds for both potential outcomes $W(z)$ and $D(z)$ and covariates $h(X)$, and also impose Assumption 3.14. Suppose the adjustment coefficients $(\widehat{\gamma}_W, \widehat{\gamma}_D) = (\gamma_W, \gamma_D) + o_p(1)$. Our first result is a consequence of Theorem 3.4. To state the result, we define the modified potential outcomes $Q(z) = W(z) - \tau_L D(z)$ for $z \in \{0,1\}$ and modified adjustment coefficient $\gamma_Q = \gamma_W - \tau_L \gamma_D$.

Theorem A.1. *If* $Z_{1:n} \sim \text{Loc}(\psi, p)$ *then* $\sqrt{n}(\widehat{\tau}_{adj} - \tau_L) \Rightarrow \mathcal{N}(0, V(\gamma_Q)/\tau_D^2)$ *with*

$$V(\gamma_Q) = \text{Var}(c_Q) + E\big[\text{Var}(b_Q - \gamma_Q' h|\psi)\big] + E\left[\frac{\sigma_{1Q}^2(X)}{p} + \frac{\sigma_{0Q}^2(X)}{1-p}\right].$$

The terms $c_Q(X) = E[Q(1) - Q(0)|X]$, similarly for b_Q and σ_{zQ}^2, substituting the potential outcomes $Q(z)$ for $Y(d)$ in each formula.

Optimal Adjustment. Let $\widehat{\gamma}_Q = \widehat{\gamma}_W - \tau_L \widehat{\gamma}_D$ and define the adjustment scheme $\widehat{\tau}_{adj}$ to be efficient if $\widehat{\gamma}_Q \xrightarrow{p} \gamma_Q^* \in \text{argmin}_\gamma V(\gamma)$. We construct efficient adjusted Wald estimators using the generic efficient estimators of Section 3.4. Let $\widehat{\theta}_k^W$ and $\widehat{\theta}_k^D$ for $k \in \{PL, GO, TM\}$ be any of the generic efficient estimators of Section, plugging in outcomes W or D in place of Y. For example, $\widehat{\theta}_{PL}^W$ is the coefficient on Z_i in the regression $W_i \sim (1, \check{h}_i) + Z_i(1, \check{h}_i)$ and $\widehat{\theta}_{PL}^D$ the coefficient on Z_i in $D_i \sim (1, \check{h}_i) + Z_i(1, \check{h}_i)$. Define the LATE estimators $\widehat{\tau}_L^k = \widehat{\theta}_k^W / \widehat{\theta}_k^D$ for $k \in \{PL, GO, TM\}$. Our next theorem is a consequence of the efficiency results in Section 3.4.

Theorem A.2. *Suppose $Z_{1:n} \sim \text{Loc}(\psi, p)$. For each $k \in \{PL, GO, TM\}$, the estimator $\widehat{\tau}_L^k$ is efficient with $\sqrt{n}(\widehat{\tau}_L^k - \tau_L) \Rightarrow \mathcal{N}(0, V^*)$ for $V^* = \min_\gamma V(\gamma)$.*

Finally, we provide asymptotically exact inference on τ_L using the adjusted estimators $\widehat{\tau}_L^k$ above. Define the augmented outcomes $Q_i^a = W_i - \widehat{\tau}_L^k D_i - h_i'(\widehat{\gamma}_W - \widehat{\tau}_L^k \widehat{\gamma}_D)$. Let $\widehat{v}_1^q$, $\widehat{v}_0^q$, and $\widehat{v}_{10}^q$ be the variance estimators in Equation 4.3, plugging in Q_i^a in place of Y_i^a. Define the variance estimator

$$\widehat{V} = \frac{1}{(\widehat{\theta}_k^D)^2} \left[\text{Var}_n \left(\frac{(D_i - p) Q_i^a}{p - p^2} \right) - \widehat{v}_1^q - \widehat{v}_0^q - 2\widehat{v}_{10}^q \right] \tag{A.2}$$

Theorem A.3. *Suppose $Z_{1:n} \sim \text{Loc}(\psi, p)$. Then $\widehat{V} = V^* + o_p(1)$.*

Theorems A.1 and A.3 show that the confidence interval $\widehat{C} = [\widehat{\tau}_L^k \pm \widehat{V}^{1/2} c_{1-\alpha/2}/\sqrt{n}]$ with $c_\alpha = \Phi^{-1}(\alpha)$ is asymptotically exact in the sense that $P(\tau_L \in \widehat{C}) = 1 - \alpha + o(1)$.

A.2 Varying Propensities

In this section, we extend our results to fine stratification with varying propensities $p(\psi)$. To that end, let $p(\psi) \in \{a_l/k_l : l \in L\}$ with $|L| < \infty$ a finite index set. Cytrynbaum (2023) extends Definition 2.1 to non-constant $p(\psi)$ by the following double stratification procedure:

(1) Partition the units $\{1, \dots, n\}$ into propensity strata $S_l \equiv \{i : p(X_i) = a_l/k_l\}$.

(2) In each propensity stratum S_l, draw samples $(D_i)_{i \in S_l} \sim \text{Loc}(\psi, a_l/k_l)$.

To implement this, we run the algorithm of Cytrynbaum (2023) to match units into groups of k_l separately in each propensity stratum S_l, drawing treatment assignments $(D_i)_{i \in g} \sim \text{CR}(a_l/k_l)$ independently for each $g \in \mathcal{G}_l$. Define $\widehat{\theta}_{adj}(\gamma)$ to be the AIPW estimator of Section 3.2, with linear models $f_d(X_i) = \gamma_d' h(X_i)$ for $d \in \{0, 1\}$, so that

$$\widehat{\theta}_{adj}(\gamma) = (\gamma_1 - \gamma_0)' E_n[h_i] + E_n \left[\frac{D_i(Y_i - \gamma_1' h_i)}{p(\psi_i)} \right] - E_n \left[\frac{(1 - D_i)(Y_i - \gamma_0' h_i)}{1 - p(\psi_i)} \right].$$

Define $\gamma = (\gamma_0, \gamma_1)$ and weighted covariates $h_i^p = \left(h_i\sqrt{\frac{p_i}{1-p_i}}, h_i\sqrt{\frac{1-p_i}{p_i}}\right)$. Under assumption 3.1, Theorem 3.4 may be extended to show that if $\widehat{\gamma} \xrightarrow{p} \gamma$ and $D_{1:n} \sim \text{Loc}(\psi, p(\psi))$ then $\sqrt{n}(\widehat{\theta}_{adj}(\widehat{\gamma}) - \text{ATE}) \Rightarrow \mathcal{N}(0, V(\gamma))$ with variance

$$V(\gamma) = \text{Var}(c(X)) + E\left[\text{Var}\left(b - \gamma' h^p | \psi\right)\right] + E\left[\frac{\sigma_1^2(X)}{p(\psi)} + \frac{\sigma_0^2(X)}{1-p(\psi)}\right].$$

The optimal adjustment coefficient is $\gamma^* = E[\text{Var}(h_i^p|\psi_i)]^{-1}E[\text{Cov}(h_i^p, b_i|\psi_i)]$ if the condition $E[\text{Var}(h_i^p|\psi_i)] \succ 0$ is satisfied. Let k_i denote the size of the group that unit i belongs to. Extending the work in Section 3.4, the estimator

$$\widehat{\gamma} = E_n\left[\check{h}_i^p(\check{h}_i^p)'\frac{k_i}{k_i-1}\right]^{-1} E_n\left[\check{h}_i^p Y_i^{TM}\frac{k_i}{k_i-1}\right]$$

with weighted outcomes $Y_i^{TM} = D_iY_i(1-p_i)^{1/2}p_i^{-3/2} + (1-D_i)Y_ip_i^{1/2}(1-p_i)^{-3/2}$ has $\widehat{\gamma} = \gamma^* + o_p(1)$. Then the estimator $\widehat{\theta}_{adj}(\widehat{\gamma})$ is efficient in the sense of achieving the minimal variance $\min_\gamma V(\gamma)$.

A.3 Non-Interacted Regression Adjustment

For completeness, before continuing we describe the asymptotic behavior of the commonly used non-interacted regression estimator under stratified designs. Let $\widehat{\theta}_N$ be the coefficient on D_i in $Y \sim 1 + D + h$.

Theorem A.4. *Suppose Assumptions 3.1 and 3.14 hold. The estimator has representation* $\widehat{\theta}_N = \widehat{\theta} - \widehat{\gamma}_N'(\bar{h}_1 - \bar{h}_0) + O_p(n^{-1})$. *If* $D_{1:n} \sim \text{Loc}(\psi, p)$ *then* $\sqrt{n}(\widehat{\theta}_N - \text{ATE}) \Rightarrow \mathcal{N}(0, V)$ *with variance*

$$V = \text{Var}(c(X)) + E[\text{Var}(b - \gamma_N' h|\psi)] + E\left[\frac{\sigma_1^2(X)}{p} + \frac{\sigma_0^2(X)}{1-p}\right].$$

The coefficient $\gamma_N = \text{argmin}_{\gamma \in \mathbb{R}^{d_h}} \text{Var}(f - \gamma' h)$ *for target function*

$$f(x) = m_1(x)\sqrt{\frac{p}{1-p}} + m_0(x)\sqrt{\frac{1-p}{p}}$$

with $f(x) \neq b(x)$ *in general. The fixed effects estimator is efficient if either* $p = 1/2$ *or* $\text{Cov}(h, Y(1) - Y(0)) = 0$.

Theorem A.4 shows that $\widehat{\theta}_N$ is generally inefficient since it uses the wrong objective function. In particular, the target function $f(x) \neq b(x)$ unless $p = 1/2$. Also, the limiting coefficient γ_N minimizes marginal instead of conditional variance. The results in Section 4 show how to construct asymptotically exact confidence intervals for the ATE using $\widehat{\theta}_N$.

A.4 Nonlinear Adjustment

Alternately, we may consider general nonlinear covariate adjustment strategies. Let $\widehat{h}(x)$ be a function estimated in some class $\mathcal{H}$ and consider the adjusted estimator

$$\widehat{\theta}_{adj}(\widehat{h}) = E_n\left[\frac{(Y_i - \widehat{h}(X_i))(D_i - p_i)}{p_i - p_i^2}\right].$$

For example, the usual AIPW estimator in Section 3.2 can be shown to take this form. Linear adjustment corresponds to the parametric family $\mathcal{H} = \{h(x)'\gamma : \gamma \in \mathbb{R}^{d_h}\}$. Similar to Bai et al. (2024), suppose that for some function $h(X) \in L_2$ the equicontinuity condition holds

$$\sqrt{n}E_n\left[\frac{(\widehat{h} - h)(X_i)(D_i - p_i)}{p_i - p_i^2}\right] = o_p(1).$$

Theorem 3.4 can be extended to show that if $D_{1:n} \sim \text{Loc}(\psi, p(\psi))$ then $\sqrt{n}(\widehat{\theta}_{adj}(\widehat{h}) - \text{ATE}) \Rightarrow \mathcal{N}(0, V(h))$ with asymptotic variance

$$V(h) = \text{Var}(c(X)) + E\left[\text{Var}\left(b - h/c_p(\psi)|\psi\right)\right] + E\left[\frac{\sigma_1^2(X)}{p(\psi)} + \frac{\sigma_0^2(X)}{1 - p(\psi)}\right]$$

for $c_p(\psi) = \sqrt{p(\psi) - p(\psi)^2}$. One natural extension of the current work would be to solve a general version of the optimal adjustment problem over a nonlinear or general nonparametric function class $\mathcal{H}$.

$$\min_{h \in \mathcal{H}} E\left[\text{Var}\left(b - h/c_p(\psi)|\psi\right)\right] \tag{A.3}$$

This requires new technical tools, the development of which we leave to future work.

A.5 Proofs for Section 3.1

Proof of Theorem 3.4. First, note that since $E[|h|_2^2] < \infty$ we may apply Lemma A.2 of Cytrynbaum (2023) to show that

$$\widehat{\gamma}'(\bar{h}_1 - \bar{h}_0)c_p = \widehat{\gamma}'E_n\left[\frac{(D_i - p)}{\sqrt{p - p^2}}h_i\right] = \gamma'E_n\left[\frac{(D_i - p)}{\sqrt{p - p^2}}h_i\right] + (\widehat{\gamma} - \gamma)'E_n\left[\frac{(D_i - p)}{\sqrt{p - p^2}}h_i\right]$$
$$= \gamma'E_n\left[\frac{(D_i - p)}{\sqrt{p - p^2}}h_i\right] + o_p(n^{-1/2}) = \gamma'(\bar{h}_1 - \bar{h}_0)c_p + o_p(n^{-1/2}).$$

Define auxiliary potential outcomes $Z(d) = Y(d) - c_p\gamma'h(X)$ for $d \in \{0, 1\}$ with $Z_i = Z(D_i)$. Summarizing, we have shown that $\widehat{\theta}_{adj} = \bar{Z}_1 - \bar{Z}_0 + o_p(n^{-1/2})$. Observe that $E[Z(d)^2] \lesssim E[Y(d)^2] + c_p^2|\gamma|_2^2E[|h(X)|_2^2] < \infty$. Then we may apply the general version of Theorem 3.11 in Cytrynbaum (2023) (Equation 3.6). Setting $q = 1$ and

$\psi_1 = \psi_2$ and applying the theorem to the auxiliary potential outcomes $Z(d)$, we have $\sqrt{n}(\widehat{\theta}_{adj} - \text{ATE}) \Rightarrow \mathcal{N}(0, V)$

$$V = \text{Var}(c_Z(X)) + E[\text{Var}(b_Z(X;p)|\psi)] + E\left[\frac{\sigma^2_{1,Z}(X)}{p} + \frac{\sigma^2_{0,Z}(X)}{1-p}\right].$$

Calculating, we have $c_Z(X) = E[Z(1) - Z(0)|X] = c(X)$ and

$$b_Z(X) = E[Z(1)|X]\left(\frac{1-p}{p}\right)^{1/2} + E[Z(0)|X]\left(\frac{p}{1-p}\right)^{1/2} = b(X;p) - \gamma' h(X).$$

Finally, $\sigma^2_{d,Z}(X) = \text{Var}(Z(d)|X) = \text{Var}(Y(d)|X) = \sigma^2_d(X)$. Then the variance V above is

$$V = \text{Var}(c(X)) + E[\text{Var}(b - \gamma' h|\psi)] + E\left[\frac{\sigma^2_1(X)}{p} + \frac{\sigma^2_0(X)}{1-p}\right]$$

as claimed. □

Proof of Theorem 3.2. Define $W_i = (1, \tilde{h}_i)$. First consider the regression $Y_i \sim D_i W_i + (1 - D_i)W_i$, with coefficients $(\widehat{\gamma}_1, \widehat{\gamma}_0)$. By Frisch-Waugh and orthogonality of regressors, $\widehat{\gamma}_1$ is numerically equivalent to the regression coefficient $Y_i \sim D_i W_i$ and similarly for $\widehat{\gamma}_0$. Then consider $Y_i = D_i W_i' \widehat{\gamma}_1 + e_i$ with $E_n[e_i(D_i W_i)] = 0$. Then $D_i Y_i = D_i W_i' \widehat{\gamma}_1 + D_i e_i$ and $E_n[D_i e_i(D_i W_i)] = E_n[e_i(D_i W_i)] = 0$. Then $\widehat{\gamma}_1$ can be identified with the regression coefficient of $Y_i \sim W_i$ in the set $\{i : D_i = 1\}$. Let $\widehat{\gamma}_1 = (\widehat{c}_1, \widehat{\alpha}_1)$. By the usual OLS formula $\widehat{c}_1 = E_n[Y_i|D_i = 1] - \widehat{\alpha}_1' E_n[\tilde{h}_i|D_i = 1]$ and $\widehat{\alpha}_1 = \text{Var}_n(\tilde{h}_i|D_i = 1)^{-1}\,\text{Cov}_n(\tilde{h}_i, Y_i|D_i = 1)$. Similar formulas hold for $D_i = 0$ by symmetry. Next, note that for $m = d_h + 1$ the original regressors can be written as a linear transformation

$$\begin{pmatrix} D_i W_i \\ W_i \end{pmatrix} = \begin{pmatrix} I_m & 0 \\ I_m & I_m \end{pmatrix} \begin{pmatrix} D_i W_i \\ (1 - D_i) W_i \end{pmatrix}.$$

Then the OLS coefficients for the original regression $Y_i \sim D_i W_i + W_i$ are given by the change of variables formula

$$\left(\begin{pmatrix} I_k & 0 \\ I_k & I_k \end{pmatrix}'\right)^{-1} \begin{pmatrix} \widehat{\gamma}_1 \\ \widehat{\gamma}_0 \end{pmatrix} = \begin{pmatrix} I_k & -I_k \\ 0 & I_k \end{pmatrix} \begin{pmatrix} \widehat{\gamma}_1 \\ \widehat{\gamma}_0 \end{pmatrix} = \begin{pmatrix} \widehat{\gamma}_1 - \widehat{\gamma}_0 \\ \widehat{\gamma}_0 \end{pmatrix}.$$

In particular, the coefficient on D_i in the original regression is

$$
\begin{aligned}
\widehat{\theta}_L = \widehat{c}_1 - \widehat{c}_0 &= E_n[Y_i - \widehat{\alpha}_1'\tilde{h}_i|D_i = 1] - E_n[Y_i - \widehat{\alpha}_0'\tilde{h}_i|D_i = 0] \\
&= \widehat{\theta} - E_n\left[\frac{\widehat{\alpha}_1'\tilde{h}_i D_i}{p}\right] + E_n\left[\frac{\widehat{\alpha}_0'\tilde{h}_i(1-D_i)}{1-p}\right] \\
&= \widehat{\theta} - E_n\left[\frac{\widehat{\alpha}_1' h_i (D_i - p)}{p}\right] - E_n\left[\frac{\widehat{\alpha}_0' h_i(D_i - p)}{1-p}\right] \\
&= \widehat{\theta} - (\widehat{\alpha}_1(1-p) + \widehat{\alpha}_0 p)' E_n\left[\frac{h_i(D_i - p)}{p(1-p)}\right] \\
&= \widehat{\theta} - \left(\widehat{\alpha}_1\sqrt{\frac{1-p}{p}} + \widehat{\alpha}_0\sqrt{\frac{p}{1-p}}\right)' (\bar{h}_1 - \bar{h}_0)c_p.
\end{aligned}
$$

The second equality since $E_n[D_i] = p$ identically. The third equality by expanding $D_i = D_i - p + p$ and using $E_n[\tilde{h}_i] = 0$ and $E_n[(D_i - p)E_n[h_i]] = 0$. The fourth equality is algebra and collecting terms. The fifth equality since $\bar{h}_1 - \bar{h}_0 = E_n[h_i(D_i - p)/p(1-p)]$ again using $E_n[D_i] = p$ and $c_p = \sqrt{p(1-p)}$ by definition.

Next, consider the coefficient $\widehat{\alpha}_1 = \mathrm{Var}_n(\tilde{h}_i|D_i = 1)^{-1}\,\mathrm{Cov}_n(\tilde{h}_i, Y_i|D_i = 1)$. We have $\mathrm{Var}_n(\tilde{h}_i|D_i = 1) = p^{-1}E_n[D_i\tilde{h}_i\tilde{h}_i'] - p^{-2}E_n[D_i\tilde{h}_i]E_n[D_i\tilde{h}_i']$. Let $1 \le t, t' \le d_h$. Then we may compute $E_n[D_i\tilde{h}_{it}\tilde{h}_{it'}] = E_n[(D_i - p)\tilde{h}_{it}\tilde{h}_{it'}] + pE_n[\tilde{h}_{it}\tilde{h}_{it'}]$. Expanding the first term

$$
\begin{aligned}
E_n[(D_i - p)\tilde{h}_{it}\tilde{h}_{it'}] = E_n[(D_i - p)h_{it}h_{it'}] &- E_n[h_{it}]E_n[(D_i - p)h_{it'}] - E_n[h_{it'}]E_n[(D_i - p)h_{it}] \\
&+ E_n[h_{it'}]E_n[h_{it}]E_n[D_i - p] = o_p(1).
\end{aligned}
$$

The final equality follows since $E_n[(D_i - p)h_{it}h_{it'}] = o_p(1)$ by applying Lemma A.2 of Cytrynbaum (2023), using that $E[|h_{it}\tilde{h}_{it'}|] \le E[|h_i|_2^2] < \infty$, and similarly for the other terms. By WLLN, we also have $E_n[\tilde{h}_{it}\tilde{h}_{it'}] \xrightarrow{p} \mathrm{Var}(h)$. Then by continuous mapping $\mathrm{Var}_n(\tilde{h}_i|D_i = 1)^{-1} = \mathrm{Var}(h)^{-1} + o_p(1)$. Similar reasoning shows $\mathrm{Cov}_n(\tilde{h}_i, Y_i|D_i = 1) = \mathrm{Cov}(h_i, Y_i(1)) + o_p(1)$.

Then we have shown $\widehat{\alpha}_1 = \mathrm{Var}(h)^{-1}\,\mathrm{Cov}(h, Y(1)) + o_p(1) = \mathrm{Var}(h)^{-1}\,\mathrm{Cov}(h, m_1) + o_p(1)$. By symmetry, we also have $\widehat{\alpha}_0 = \mathrm{Var}(h)^{-1}\,\mathrm{Cov}(h, m_0) + o_p(1)$. Putting this all together, we have $\widehat{\alpha}_1\sqrt{\frac{1-p}{p}} + \widehat{\alpha}_0\sqrt{\frac{p}{1-p}} = \mathrm{Var}(h)^{-1}\,\mathrm{Cov}(h, b) + o_p(1) = \gamma_L + o_p(1)$. Then by Theorem 3.4, $\sqrt{n}(\widehat{\theta}_L - \mathrm{ATE}) \Rightarrow \mathcal{N}(0, V)$ with

$$
V = V(\gamma_L) = \mathrm{Var}(c(X)) + E\Big[\mathrm{Var}(b - \gamma_L' h|\psi)\Big] + E\left[\frac{\sigma_1^2(X)}{p} + \frac{\sigma_0^2(X)}{1-p}\right]
$$

as claimed. The claimed representation follows from the change of variables formula above, since $\widehat{\alpha}_1 = \widehat{a}_1 + \widehat{a}_0$ and $\widehat{\alpha}_0 = \widehat{a}_0$. This finishes the proof. □

Proof of Theorem A.4. We have $Y_i = \widehat{c} + \widehat{\theta}_N D_i + \widehat{\gamma}_N' h_i + e_i$ with $E_n[e_i(1, D_i, h_i)] = 0$. By applying Frisch-Waugh twice, we have $\tilde{Y}_i = \widehat{\theta}_N(D_i - p) + \widehat{\gamma}_N'\tilde{h}_i + e_i$ and $\widehat{\theta}_N =$

$E_n[(\check{D}_i)^2]^{-1}E_n[\check{D}_iY_i]$ with partialled treatment $\check{D}_i = (D_i - p) - (E_n[\tilde{h}_i\tilde{h}_i']^{-1}E_n[\tilde{h}_i(D_i - p)])'\tilde{h}_i$. Squaring this expression gives

$$\begin{aligned}(\check{D}_i)^2 &= (D_i - p)^2 - 2(D_i - p)(E_n[\tilde{h}_i\tilde{h}_i']^{-1}E_n[\tilde{h}_i(D_i - p)])'\tilde{h}_i \\ &\quad + ((E_n[\tilde{h}_i\tilde{h}_i']^{-1}E_n[\tilde{h}_i(D_i - p)])'\tilde{h}_i)^2 \equiv \eta_{i1} + \eta_{i2} + \eta_{i3}.\end{aligned}$$

Using $E_n[\tilde{h}_i(D_i - p)] = O_p(n^{-1/2})$ by Lemma A.2 of Cytrynbaum (2023) and $E_n[\tilde{h}_i\tilde{h}_i'] \xrightarrow{p} \mathrm{Var}(h) \succ 0$, we see that $E_n[\eta_{i2}] = O_p(n^{-1})$ and $E_n[\eta_{i3}] = O_p(n^{-1})$. Then we have $E_n[(\check{D}_i)^2] = E_n[(D_i - p)^2] + O_p(n^{-1}) = p - p^2 + O_p(n^{-1})$. Then apparently $\widehat{\theta}_N = (p - p^2)^{-1}E_n[\check{D}_iY_i] + O_p(n^{-1})$. Now note that

$$\begin{aligned}E_n[\check{D}_iY_i] &= E_n[(D_i - p)Y_i] - E_n[(E_n[\tilde{h}_i\tilde{h}_i']^{-1}E_n[\tilde{h}_i(D_i - p)])'\tilde{h}_iY_i] \\ &= E_n[(D_i - p)Y_i] - E_n[(D_i - p)\tilde{h}_i]'(E_n[\tilde{h}_i\tilde{h}_i']^{-1}E_n[\tilde{h}_iY_i]).\end{aligned}$$

By using Frisch-Waugh to partial out $D_i - p$ from the original regression, we have $\widehat{\gamma}_N = E_n[\bar{h}_i\bar{h}_i']^{-1}E_n[\bar{h}_iY_i]$ with $\bar{h}_i = \tilde{h}_i - (E_n[(D_i - p)^2]^{-1}E_n[\tilde{h}_i(D_i - p)])(D_i - p)$. Then using $E_n[\tilde{h}_i(D_i - p)] = O_p(n^{-1/2})$ again, we have $E_n[\bar{h}_i\bar{h}_i{}'] = E_n[\tilde{h}_i\tilde{h}_i'] + O_p(n^{-1})$. Similarly, $E_n[\bar{h}_iY_i] = E_n[\tilde{h}_iY_i] - \widehat{\theta}E_n[\tilde{h}_i(D_i - p)] = E_n[\tilde{h}_iY_i] + O_p(n^{-1/2})$. Then the coefficient $\widehat{\gamma}_N = E_n[\tilde{h}_i\tilde{h}_i']^{-1}E_n[\tilde{h}_iY_i] + O_p(n^{-1/2})$. Then we have shown that

$$\begin{aligned}\widehat{\theta}_N &= \widehat{\theta} - E_n\left[\frac{(D_i - p)\tilde{h}_i}{\sqrt{p - p^2}}\right]'(E_n[\tilde{h}_i\tilde{h}_i']^{-1}E_n[\tilde{h}_iY_i])(p - p^2)^{-1/2} + O_p(n^{-1}) \\ &= \widehat{\theta} - E_n\left[\frac{(D_i - p)h_i}{\sqrt{p - p^2}}\right]'\widehat{\gamma}_N(p - p^2)^{-1/2} + O_p(n^{-1}) \\ &= \widehat{\theta} - (\widehat{\gamma}_N/c_p)'(\bar{h}_1 - \bar{h}_0)c_p + O_p(n^{-1}).\end{aligned}$$

The second line uses that $E_n[(D_i - p)c] = 0$ for any constant. This shows the claimed representation. We have $E_n[\tilde{h}_i\tilde{h}_i'] = \mathrm{Var}(h) + o_p(1)$. Note also that $E_n[\tilde{h}_iY_i(1)D_i] = p\,\mathrm{Cov}(h, Y(1)) + o_p(1)$ and $E_n[\tilde{h}_iY_i(0)(1 - D_i)] = (1 - p)\,\mathrm{Cov}(h, Y(0)) + o_p(1)$. Putting this together, we have shown that

$$\begin{aligned}\widehat{\gamma}_N/c_p &= \mathrm{Var}(h)^{-1}\,\mathrm{Cov}\left(h, m_1\sqrt{\frac{p}{1-p}} + m_0\sqrt{\frac{1-p}{p}}\right) + o_p(1) \\ &= \operatorname*{argmin}_\gamma \mathrm{Var}(f - \gamma'h) + o_p(1) = \gamma_N + o_p(1).\end{aligned}$$

Then the first claim follows from Theorem 3.4. For the efficiency claims, (a) if $p = 1/2$ and $\psi = 1$, then $f = b$ and $\gamma_N = \operatorname{argmin}_\gamma \mathrm{Var}(f - \gamma'h) = \operatorname{argmin}_\gamma E[\mathrm{Var}(b - \gamma'h|\psi)]$. For (c), if $\psi = 1$ and $\text{Cov}(h, m_1 - m_0) = 0$, then we have

$$\text{Cov}(h, f) - \text{Cov}(h, b) = \text{Cov}\left(h, (m_1 - m_0)\frac{2p-1}{\sqrt{p(1-p)}}\right) = 0.$$

By expanding the variance, we have $\text{argmin}_\gamma \text{Var}(f - \gamma' h) = \text{argmin}_\gamma \text{Var}(b - \gamma' h)$. If (b) holds, then $m_1 - m_0 = 0$ and the same conclusion follows. This finishes the proof. □

Proof of Theorem 3.7. For any $\gamma \in \mathbb{R}^{d_h}$, we have $\text{argmin}_{g \in L_2(\psi)} E[(Y(d) - g(\psi) - \gamma' h)^2] = E[Y(d) - \gamma' h | \psi]$ by standard arguments. Then the coefficients

$$\gamma_d = \underset{\gamma \in \mathbb{R}^{d_h}}{\text{argmin}}\, E[(Y(d) - \gamma' h - E[Y(d) - \gamma' h|\psi])^2] = \underset{\gamma \in \mathbb{R}^{d_h}}{\text{argmin}}\, E[\text{Var}(Y(d) - \gamma' h|\psi)]$$

and $g_d(\psi) = E[Y(d) - \gamma_d' h|\psi]$. Define $f_d(x) = g_d(\psi) + \gamma_d' h$. Then the AIPW estimator

$$\begin{aligned}
\widehat{\theta}_{AIPW} &= E_n[f_1(X_i) - f_0(X_i)] + E_n\left[\frac{D_i(Y_i - f_1(X_i))}{p}\right] - E_n\left[\frac{(1-D_i)(Y_i - f_0(X_i))}{1-p}\right] \\
&= \widehat{\theta} - E_n\left[f_1(X_i)\frac{(D_i - p)}{p}\right] - E_n\left[f_0(X_i)\frac{(D_i - p)}{1-p}\right] \\
&= \widehat{\theta} - E_n\left[(D_i - p)\left(\frac{f_1(X_i)}{p} + \frac{f_0(X_i)}{1-p}\right)\right] \\
&= E_n\left[\frac{D_i - p}{p - p^2}\left(Y_i - (1-p)f_1(X_i) - p f_0(X_i)\right)\right].
\end{aligned}$$

Let $F(x) = (1+p)f_1(x) + pf_0(x)$. Then by vanilla CLT we have $\sqrt{n}(\widehat{\theta}_{AIPW} - \text{ATE}) \Rightarrow \mathcal{N}(0, V)$ with $V = \text{Var}\left(\frac{D_i - p}{p-p^2}(Y_i - F(X_i))\right) \equiv \text{Var}(W_i)$ with $W_i = \frac{D_i - p}{p-p^2}(Y_i - F(X_i)) - \text{ATE}$. By fundamental expansion of the IPW estimator from Cytrynbaum (2023)

$$\begin{aligned}
W_i = \frac{D_i - p}{p - p^2}(Y_i - F(X_i)) - \text{ATE} &= \left[\frac{D_i \epsilon_i^1}{p} - \frac{(1-D_i)\epsilon_i^0}{1-p}\right] \\
&+ [c(X_i) - \text{ATE}] + \left[\frac{D_i - p}{\sqrt{p - p^2}}\left((m_1 - f_1)\sqrt{\frac{1-p}{p}} + (m_0 - f_0)\sqrt{\frac{p}{1-p}}\right)\right].
\end{aligned}$$

By the law of total variance and tower law

$$\begin{aligned}
\text{Var}(W) &= \text{Var}(E[W|X]) + E[\text{Var}(W|X)] \\
&= \text{Var}(E[W|X]) + E[\text{Var}(E[W|X, D]|X)] + E[\text{Var}(W|X, D)].
\end{aligned}$$

From the expansion above, $\mathrm{Var}(E[W|X]) = \mathrm{Var}(c(X) - \mathrm{ATE}) = \mathrm{Var}(c(X))$. Next

$$E[W|X,D] = [c(X_i) - \mathrm{ATE}] + \left[\frac{D_i - p}{\sqrt{p-p^2}}\left((m_1 - f_1)\sqrt{\frac{1-p}{p}} + (m_0 - f_0)\sqrt{\frac{p}{1-p}}\right)\right]$$

$$E[\mathrm{Var}(E[W|X,D]|X)] = E\left[\left((m_1 - f_1)\sqrt{\frac{1-p}{p}} + (m_0 - f_0)\sqrt{\frac{p}{1-p}}\right)^2\right]$$

Using the definition of $f_d(x)$ gives

$$\begin{aligned}
&E\left[\left((m_1 - \gamma_1' h - E[m_1 - \gamma_1' h|\psi])\sqrt{\frac{1-p}{p}} + (m_0 - \gamma_0' h - E[Y(0) - \gamma_0' h|\psi])\sqrt{\frac{p}{1-p}}\right)^2\right] \\
&= E\left[\mathrm{Var}\left((m_1 - \gamma_1' h)\sqrt{\frac{1-p}{p}} + (m_0 - \gamma_0' h)\sqrt{\frac{p}{1-p}}\middle|\psi\right)\right] \\
&= E\left[\mathrm{Var}\left(b - \left(\gamma_1\sqrt{\frac{1-p}{p}} + \gamma_0\sqrt{\frac{p}{1-p}}\right)' h\middle|\psi\right)\right] = \operatorname*{argmin}_{\gamma\in\mathbb{R}^{d_h}} E[\mathrm{Var}(b - \gamma' h|\psi)].
\end{aligned}$$

The final line by characterization of γ_d above and linearity of $Z \to \mathrm{argmin}_\gamma E[\mathrm{Var}(Z - \gamma' h|\psi)]$. Finally note that

$$\begin{aligned}
\mathrm{Var}(W|X,D) &= E\left[\left(\frac{D_i\epsilon_i^1}{p} - \frac{(1-D_i)\epsilon_i^0}{1-p}\right)^2\middle|X,D\right] = E\left[\frac{D_i(\epsilon_i^1)^2}{p^2} + \frac{(1-D_i)(\epsilon_i^0)^2}{(1-p)^2}\middle|X_i,D_i\right] \\
&= \frac{D_i\sigma_1^2(X_i)}{p^2} + \frac{(1-D_i)\sigma_0^2(X_i)}{(1-p)^2}.
\end{aligned}$$

Then $E[\mathrm{Var}(W|X,D)] = E\left[\frac{\sigma_1^2(X_i)}{p} + \frac{\sigma_0^2(X_i)}{1-p}\right]$. Comparing with Equation 3.3 finishes the proof. □

A.6 Proofs for Section 3.3

Proof of Theorem 3.9. By Theorem 3.2, the middle term of the asymptotic variance is $E[\mathrm{Var}(b - \beta' h|\psi)]$ with $\beta = \mathrm{Var}(h)^{-1}\mathrm{Cov}(h,b)$. This is the OLS coefficient from the population regression $b = a + \beta' h + e = a + \alpha' z + \gamma' w + e$ with $E[e(1,w,z)] = 0$ and $h = (w,z)$. Denote $\tilde{b} = b - E[b]$ and similarly for $\tilde{w}, \tilde{z}$. By Frisch-Waugh we have $\tilde{b} = \alpha'\tilde{z} + \gamma'\tilde{w} + e$. Let $\breve{w} = \tilde{w} - (E[\tilde{z}\tilde{z}']^{-1}E[\tilde{z}\tilde{w}'])'\tilde{z}$. Then again by Frisch-Waugh the coefficient of interest is $\gamma = E[\breve{w}\breve{w}']^{-1}E[\breve{w}b]$. Next, we characterize this coefficient.

By assumption, $E[w|\psi] = c + \Lambda z$. De-meaning both sides gives $E[\tilde{w}|\psi] = \Lambda\tilde{z}$. Write $\tilde{u} = \tilde{w} - E[\tilde{w}|\psi] = \tilde{w} - \Lambda\tilde{z}$ with $E[\tilde{u}|\psi] = 0$. Then we have

$$E[\tilde{z}\tilde{w}'] = E[\tilde{z}(\tilde{w} - E[\tilde{w}|\psi] + E[\tilde{w}|\psi])'] = E[\tilde{z}\tilde{u}'] + E[\tilde{z}\tilde{z}'\Lambda'] = E[\tilde{z}\tilde{z}']\Lambda'.$$

Then $\check{w} = \tilde{w} - (E[\tilde{z}\tilde{z}']^{-1}E[\tilde{z}\tilde{z}']\Lambda')'\tilde{z} = \tilde{w} - \Lambda\tilde{z} = \tilde{u}$. We have now shown that

$$\gamma = E[\tilde{u}\tilde{u}']^{-1}E[\tilde{u}b] = E[\text{Var}(\tilde{w}|\psi)]^{-1}E[\text{Cov}(\tilde{w}, b|\psi)] = E[\text{Var}(w|\psi)]^{-1}E[\text{Cov}(w, b|\psi)].$$

In particular, the coefficient $\beta = (\alpha, \gamma)$ is optimal

$$\begin{aligned} E[\text{Var}(b - \beta' h|\psi)] &= E[\text{Var}(b - \gamma' w|\psi)] = \min_{\tilde{\gamma}} E[\text{Var}(b - \tilde{\gamma}' w|\psi)] \\ &= \min_{\tilde{\alpha},\tilde{\gamma}} E[\text{Var}(b - \tilde{\alpha}' z - \tilde{\gamma}' w|\psi)] = \min_{\beta} E[\text{Var}(b - \beta' h|\psi)]. \end{aligned}$$

The second equality since $z = z(\psi)$. This completes the proof. □

A.7 Proofs for Section 3.4

Proof of Theorem 3.15. By Frisch-Waugh $\check{Y}_i = \widehat{\theta}_{FE}\check{D}_i + \widehat{\gamma}_{FE}'\check{h}_i + e_i$ with $\check{D}_i = D_i - k^{-1}\sum_{j\in g(i)} D_j = D_i - p$ and $\check{h}_i = h_i - k^{-1}\sum_{j\in g(i)} h_j$. Applying Frisch-Waugh again, the estimator is $\widehat{\theta}_{FE} = E_n[(\bar{D}_i)^2]^{-1}E_n[\bar{D}_i Y_i]$ with $\bar{D}_i = (D_i - p) - (E_n[\check{h}_i\check{h}_i']^{-1}E_n[\check{h}_i(D_i - p)])'\check{h}_i$. By Lemma A.8 we have $E_n[\check{h}_i\check{h}_i'] \xrightarrow{p} \frac{k-1}{k}E[\text{Var}(h|\psi)] \succ 0$, so that $E_n[\check{h}_i\check{h}_i']^{-1} = O_p(1)$. By the definition of stratification, $E_n[(D_i - p)\mathbb{1}(g(i) = g)] = 0$ for all g. Then defining $\bar{h}_g \equiv k^{-1}\sum_{j\in g} h_j$ we may write

$$\begin{aligned} E_n[(D_i - p)\check{h}_i] &= E_n\left[(D_i - p)\left(h_i - \sum_g \mathbb{1}(g(i) = g)\bar{h}_g\right)\right] \\ &= E_n[(D_i - p)h_i] = O_p(n^{-1/2}). \end{aligned}$$

The final equality since $E[|h|_2^2] < \infty$ and by Lemma A.2 of Cytrynbaum (2023). Then apparently $E_n[(\tilde{D}_i)^2] = E_n[(D_i - p)^2] + O_p(n^{-1})$ so that $E_n[(\tilde{D}_i)^2]^{-1} = (p - p^2)^{-1} + O_p(n^{-1})$. Then we have shown that

$$\begin{aligned} \widehat{\theta}_{FE} &= \frac{E_n[(D_i - p)Y_i]}{p - p^2} - \frac{E_n[\check{h}_i(D_i - p)]'E_n[\check{h}_i\check{h}_i']^{-1}E_n[\check{h}_iY_i]}{p - p^2} + O_p(n^{-1}) \\ &= \widehat{\theta} - (\bar{h}_1 - \bar{h}_0)'E_n[\check{h}_i\check{h}_i']^{-1}E_n[\check{h}_iY_i] + O_p(n^{-1}). \end{aligned}$$

By Lemma A.8 we have

$$\begin{aligned} E_n[\check{h}_iY_i] &= E_n[\check{h}_iD_iY_i(1)] + E_n[\check{h}_i(1 - D_i)Y_i(0)] \\ &= \frac{p(k-1)}{k}E[\text{Cov}(h, Y(1)|\psi)] + \frac{(1-p)(k-1)}{k}E[\text{Cov}(h, Y(0)|\psi)] + o_p(1) \\ &= \frac{(k-1)}{k}E\left[\text{Cov}\left(h, p \cdot m_1(X) + (1-p) \cdot m_0(X)|\psi\right)\right] + o_p(1). \end{aligned}$$

Putting this together, we have $c_p^{-1} E_n[\check{h}_i \check{h}_i{}']^{-1} E_n[\check{h}_i Y_i] \xrightarrow{p} E[\mathrm{Var}(h|\psi)]^{-1} E[\mathrm{Cov}(h, f|\psi)] = \operatorname{argmin}_\gamma E[\mathrm{Var}(f - \gamma' h|\psi)]$. Similar reasoning shows that $\widehat{\gamma}_{FE} = E_n[\check{h}_i \check{h}_i{}']^{-1} E_n[\check{h}_i Y_i] + O_p(n^{-1/2})$. Then we have representation $\widehat{\theta}_{FE} = \widehat{\theta} - (c_p^{-1} \widehat{\gamma}_{FE})'(\bar{h}_1 - \bar{h}_0)c_p + o_p(n^{-1/2})$. The efficiency claims follow identically to the reasoning in Theorem A.4. This finishes the proof. □

Proof of Theorem 3.23 (Part I). Consider the regression $Y_i \sim D_i(1, \check{h}_i) + (1 - D_i)(1, \check{h}_i)$ with $\check{h}_i = h_i - k^{-1} \sum_{j \in g(i)} h_j$. Denote the OLS coefficients by $(\widehat{c}_1, \widehat{\alpha}_1)$ and $(\widehat{c}_0, \widehat{\alpha}_0)$ respectively. By Frisch-Waugh, the coefficient $(\widehat{c}_1, \widehat{\alpha}_1)$ is given by the equation $Y_i = \widehat{c}_1 + \widehat{\alpha}_1' \check{h}_i + e_i$ with $E_n[e_i(1, \check{h}_i)|D_i = 1] = 0$. By the usual OLS formula $\widehat{\alpha}_1 = \mathrm{Var}_n(\check{h}_i|D_i = 1)^{-1} \mathrm{Cov}_n(\check{h}_i, Y_i|D_i = 1)$. Observe that by definition of stratification

$$P_n(g(i) = g|D_i = 1) = \frac{P_n(D_i = 1|g(i) = g)P_n(g(i) = g)}{P_n(D_i = 1)} = P_n(g(i) = g).$$

This shows that $E_n[E_n[h_i|g(i)]|D_i = 1] = E_n[E_n[h_i|g(i)]] = E_n[h_i]$, so that $E_n[\check{h}_i|D_i = 1] = E_n[h_i|D_i = 1] - E_n[h_i] = E_n[p^{-1}(D_i - p)h_i] = O_p(n^{-1/2})$ as above. Then we have

$$\begin{aligned} \mathrm{Var}_n(\check{h}_i|D_i = 1) &= E_n[\check{h}_i \check{h}_i'|D_i = 1] - E_n[\check{h}_i|D_i = 1]E_n[\check{h}_i|D_i = 1]' \\ &= E_n[\check{h}_i \check{h}_i'|D_i = 1] + O_p(n^{-1}). \end{aligned}$$

Similarly, $\mathrm{Cov}_n(\check{h}_i, Y_i|D_i = 1) = E_n[\check{h}_i Y_i|D_i = 1] + O_p(n^{-1/2})$. Then we have

$$\begin{aligned} \widehat{\alpha}_1 &= E_n[\check{h}_i \check{h}_i'|D_i = 1]^{-1} E_n[\check{h}_i Y_i|D_i = 1] + O_p(n^{-1/2}) \\ &= \frac{k-1}{k} \frac{k}{k-1} E[\mathrm{Var}(h|\psi)]^{-1} E[\mathrm{Cov}(h, Y(1)|\psi)] + o_p(1) \end{aligned}$$

by Lemma A.8. Similarly, $\widehat{\alpha}_0 = E[\mathrm{Var}(h|\psi)]^{-1} E[\mathrm{Cov}(h, Y(0)|\psi)] + o_p(1)$. By the usual OLS formula, the constant term $\widehat{c}_1$ has form $\widehat{c}_1 = E_n[Y_i|D_i = 1] - \widehat{\alpha}_1' E_n[\check{h}_i|D_i = 1]$ and similarly for $\widehat{c}_0$. By change of variables used in the proof of Theorem 3.2, our estimator

$$\begin{aligned} \tilde{\theta} = \widehat{c}_1 - \widehat{c}_0 &= E_n[Y_i|D_i = 1] - E_n[Y_i|D_i = 0] - \left[\widehat{\alpha}_1' E_n[\check{h}_i|D_i = 1] - \widehat{\alpha}_0' E_n[\check{h}_i|D_i = 0] \right] \\ &= \widehat{\theta} - E_n \left[\frac{\widehat{\alpha}_1' h_i(D_i - p)}{p} + \frac{\widehat{\alpha}_0' h_i(D_i - p)}{1 - p} \right] \\ &= \widehat{\theta} - \left[\widehat{\alpha}_1 \sqrt{\frac{1-p}{p}} + \widehat{\alpha}_0 \sqrt{\frac{p}{1-p}} \right]' E_n \left[\frac{h_i(D_i - p)}{\sqrt{p - p^2}} \right]. \end{aligned}$$

Define $\widehat{\gamma} = \widehat{\alpha}_1\sqrt{\frac{1-p}{p}} + \widehat{\alpha}_0\sqrt{\frac{p}{1-p}}$. Then by work above

$$\begin{aligned}\widehat{\gamma} &= E[\mathrm{Var}(h|\psi)]^{-1}E\left[\mathrm{Cov}\left(h, \sqrt{\frac{1-p}{p}}Y(1) + \sqrt{\frac{p}{1-p}}Y(0)|\psi\right)\right] + o_p(1) \\ &= E[\mathrm{Var}(h|\psi)]^{-1}E\left[\mathrm{Cov}\left(h, b|\psi\right)\right] + o_p(1) = \operatorname*{argmin}_{\gamma} E[\mathrm{Var}(b - \gamma' h|\psi)] + o_p(1).\end{aligned}$$

Then applying Theorem 3.4 completes the proof. As before, $\widehat{\alpha}_1 = \widehat{a}_1 + \widehat{a}_0$ and $\widehat{\alpha}_0 = \widehat{a}_0$ by change of variables. □

Proof of Theorem 3.23 (Part II). Next, we analyze the group OLS estimator. By Theorem 3.4, it suffices to show that $\widehat{\gamma}_G = \mathrm{Var}_g(h_g)^{-1}\mathrm{Cov}_g(h_g, y_g) = c_p{\cdot}E[\mathrm{Var}(h|\psi)]^{-1}E[\mathrm{Cov}(h, b|\psi)] + o_p(1)$. For the first term, note that $E_g[h_g] = O_p(n^{-1/2})$ as above, so that $\mathrm{Var}(h_g) = E_g[h_g h_g'] - E_g[h_g]E_g[h_g]' = E_g[h_g h_g'] + O_p(n^{-1})$. Similarly, $\mathrm{Cov}_g(h_g, y_g) = E_g[h_g y_g] + O_p(n^{-1/2})$. Applying Lemma A.7 to each component of $h_i h_i'$ shows that

$$E_g[h_g h_g'] = \frac{k}{n}\sum_g\left(k^{-1}\sum_{i\in g}\frac{h_i(D_i - p)}{p - p^2}\right)\left(k^{-1}\sum_{i\in g}\frac{h_i'(D_i - p)}{p - p^2}\right) = \frac{kE[\mathrm{Var}(h|\psi)]}{a(k-a)} + o_p(1).$$

Using the fundamental expansion of the IPW estimator, we have

$$\begin{aligned}E_g[y_g h_g] &= \frac{k}{n}\sum_g\left(k^{-1}\sum_{i\in g}\frac{h_i(D_i - p)}{p - p^2}\right)\left(k^{-1}\sum_{i\in g}\frac{Y_i(D_i - p)}{p - p^2}\right) \\ &= \frac{k}{n}\sum_g\left(k^{-1}\sum_{i\in g}\frac{h_i(D_i - p)}{p - p^2}\right)\left(k^{-1}\sum_{i\in g}c(X_i) + \frac{b_i(D_i - p)}{\sqrt{p - p^2}} + \frac{D_i\epsilon_i^1}{p} - \frac{(1 - D_i)\epsilon_i^0}{1-p}\right) \\ &\equiv A_n + B_n + C_n.\end{aligned}$$

First, note that $A_n = O_p(n^{-1/2})$ and $C_n = O_p(n^{-1/2})$ by Lemma A.7. Moreover, we have

$$\begin{aligned}B_n &= \frac{k}{n}\sum_g\left(k^{-1}\sum_{i\in g}\frac{h_i(D_i - p)}{p - p^2}\right)\left(k^{-1}\sum_{i\in g}\frac{b_i(D_i - p)}{\sqrt{p - p^2}}\right) \\ &= \frac{k\sqrt{p - p^2}}{a(k-a)}E[\mathrm{Cov}(h, b|\psi)] + o_p(1) = \frac{E[\mathrm{Cov}(h, b|\psi)]}{\sqrt{a(k-a)}} + o_p(1).\end{aligned}$$

Putting this together, by continuous mapping we have

$$\begin{aligned}\widehat{\gamma}_G = \mathrm{Var}_g(h_g)^{-1}\mathrm{Cov}_g(h_g, y_g) &= \frac{a(k-a)}{k}\frac{1}{\sqrt{a(k-a)}}E[\mathrm{Var}(h|\psi)]^{-1}E[\mathrm{Cov}(h, b|\psi)] + o_p(1) \\ &= \sqrt{p - p^2}E[\mathrm{Var}(h|\psi)]^{-1}E[\mathrm{Cov}(h, b|\psi)] + o_p(1).\end{aligned}$$

Applying Theorem 3.4 completes the proof. □

Proof of Theorem 3.23 (Part III). Finally, we analyze the ToM estimator. From the work in part I of this proof we have

$$\begin{aligned}\widehat{\gamma}_{PL} &= \operatorname{Var}_n(\check{h}_i|D_i=1)^{-1}\operatorname{Cov}_n(\check{h}_i,Y_i|D_i=1)\sqrt{\frac{1-p}{p}}\\ &+\operatorname{Var}_n(\check{h}_i|D_i=0)^{-1}\operatorname{Cov}_n(\check{h}_i,Y_i|D_i=0)\sqrt{\frac{p}{1-p}}\end{aligned}$$

Comparing with Equation 3.10, it suffices to show that $\operatorname{Var}_n(\check{h}_i|D_i=1)^{-1}\operatorname{Var}_n(\check{h}_i)=o_p(1)$ and $\operatorname{Var}_n(\check{h}_i|D_i=0)^{-1}\operatorname{Var}_n(\check{h}_i)=o_p(1)$. This follows immediately from Lemma A.8. Applying Theorem 3.4 completes the proof. □

Proof of Theorem 3.24. First, consider the fixed effects estimator with

$$Y_i=\widehat{c}+\widehat{\tau}_{FE}D_i+\widehat{\gamma}_{FE}'\check{h}_i+\widehat{\gamma}_z'z_i+e_{i,1}.$$

Note that $\tilde{D}_i=D_i-p$ and $\check{h}_i-E_n[\check{h}_i]=\check{h}_i-(E_n[h_i]-E_n[E_n[h_i|g_i=g])=\check{h}_i$. By Frisch-Waugh, we may instead study $Y_i=\widehat{\tau}_{FE}(D_i-p)+\widehat{\gamma}_{FE}'\check{h}_i+\widehat{\gamma}_z'\tilde{z}_i+e_{i,2}$. Let $\breve{w}_i=(\check{h}_i,\tilde{z}_i)$ and $w_i=(h_i,z_i)$. Then by work in Theorem 3.15, $\widehat{\tau}_{FE}=E_n[(\bar{D}_i)^2]^{-1}E_n[\bar{D}_iY_i]$ with

$$\bar{D}_i=(D_i-p)-(E_n[\breve{w}_i\breve{w}_i']^{-1}E_n[\breve{w}_i(D_i-p)])'\breve{w}_i.$$

Previous work suffices to show that $E_n[\breve{w}_i(D_i-p)]=O_p(n^{-1/2})$. Then as before, $E_n[(\bar{D}_i)^2]^{-1}=(p-p^2)^{-1}+O_p(n^{-1})$. Then we have

$$\begin{aligned}\widehat{\tau}_{FE}&=\widehat{\theta}-(p-p^2)^{-1}(E_n[\breve{w}_i\breve{w}_i']^{-1}E_n[\breve{w}_i(D_i-p)])'E_n[\breve{w}_iY_i]\\&=\widehat{\theta}-(\bar{w}_1-\bar{w}_0)'E_n[\breve{w}_i\breve{w}_i']^{-1}E_n[\breve{w}_iY_i].\end{aligned}$$

The second equality uses $E_n[\check{h}_i(D_i-p)]=E_n[h_i(D_i-p)]$ and $E_n[\tilde{z}_i(D_i-p)]=E_n[z_i(D_i-p)]$ as noted before. This shows the claim about estimator representation.

Next, consider $\widehat{\gamma}_{FE}$. Define $g_i=(D_i-p,\tilde{z}_i)$. Let $\bar{h}_i=\check{h}_i-(E_n[g_ig_i']^{-1}E_n[g_i\check{h}_i])'g_i$. Then by Frisch-Waugh $\widehat{\gamma}_{FE}=E_n[\bar{h}_i\bar{h}_i{}']^{-1}E_n[\bar{h}_iY_i]$. Consider $E_n[\tilde{z}_i\check{h}_i]=E_n[z_i\check{h}_i]$ since $E_n[\check{h}_i]=0$. We have $E_n[z_i\check{h}_i]=o_p(1)$ by Lemma A.8. Then by previous work $E_n[g_i\check{h}_i]=o_p(1)$. Then $E_n[\bar{h}_i\bar{h}_i{}']=E_n[\check{h}_i\check{h}_i']+o_p(1)$. Similarly, $E_n[\bar{h}_iY_i]=E_n[\check{h}_iY_i]+o_p(1)$. Then by continuous mapping $\widehat{\gamma}_{FE}=E_n[\bar{h}_i\bar{h}_i{}']^{-1}E_n[\bar{h}_iY_i]=E_n[\check{h}_i\check{h}_i']^{-1}E_n[\check{h}_iY_i]+o_p(1)$, the coefficient from the regression without strata variables z_i included shown in Theorem 3.15. Consider the coefficient $\widehat{\gamma}_z$ on $z(\psi)$. Let $q_i=(D_i-p,\check{h}_i)$ and $\bar{z}_i=\tilde{z}_i-(E_n[q_iq_i']^{-1}E_n[q_i\tilde{z}_i])'q_i$. We just showed that $E_n[q_i\tilde{z}_i]=o_p(1)$. Then by similar reasoning as above and Frisch-Waugh

$$\begin{aligned}\widehat{\gamma}_z&=E_n[\bar{z}_i\bar{z}_i{}']^{-1}E_n[\bar{z}_iY_i]=E_n[\tilde{z}_i\tilde{z}_i']^{-1}E_n[\tilde{z}_iY_i]+o_p(1)\\&=\operatorname{Var}(z)^{-1}\operatorname{Cov}(z,pm_1+(1-p)m_0)+o_p(1)=c_p\operatorname{Var}(z)^{-1}\operatorname{Cov}(z,f)+o_p(1).\end{aligned}$$

Our work so far also shows that $E_n[\check{w}_i \check{w}_i'] \xrightarrow{p} \text{Diag}(E_n[\check{h}_i \check{h}_i'], E_n[\tilde{z}_i \tilde{z}_i'])$. Then it's easy to see from our expression for $\widehat{\tau}_{FE}$ that we may identify $\widehat{\gamma}_z = \widehat{\alpha}_1 + o_p(1)$. This finishes the proof for $\widehat{\tau}_{FE}$. The proofs for the modified partialled Lin estimator $\widehat{\tau}_{PL}$ and modified ToM estimators are similar and omitted for brevity.

□

A.8 Proofs for Section 4

Proof of Theorem 4.1. Define population augmented potential outcomes $Y^b(d) = Y(d) - c_p\gamma' h(X)$ for $d \in \{0, 1\}$ with outcomes $Y_i^b = Y_i^b(D_i) = Y_i - c_p\gamma' h_i$. The proof of Theorem 3.4 showed that $\widehat{\theta}_{adj} = \bar{Y}_1^b - \bar{Y}_0^b + o_p(n^{-1/2})$. Define $\widehat{v}_1^b$, $\widehat{v}_0^b$, and $\widehat{v}_{10}^b$ to be the analogues of $\widehat{v}_1$, $\widehat{v}_0$, and $\widehat{v}_{10}$ substituting Y_i^b for Y_i^a. By applying Theorem 6.1 of Cytrynbaum (2023) to $\widehat{\theta}_b \equiv \bar{Y}_1^b - \bar{Y}_0^b$, we have $\widehat{V}_b = V + o_p(1)$ for variance estimator

$$\widehat{V}_b = \text{Var}_n\left(\frac{(D_i - p)Y_i^b}{p - p^2}\right) - \widehat{v}_1^b - \widehat{v}_0^b - 2\widehat{v}_{10}^b.$$

Then it suffices to show the following claim: $\widehat{V} - \widehat{V}_b = o_p(1)$. We prove a slight generalization, letting $h_i(d)$ possibly have a potential outcomes structure and setting $h_i = h_i(D_i)$. The case with $h_i(1) = h_i(0) = h_i$ is a special case.

We work term by term. Define the weights $L_i = (D_i - p)/(p - p^2)$. Then we have $\text{Var}_n(L_i Y_i^b) - \text{Var}_n(L_i Y_i^a) = E_n[L_i^2 (Y_i^b)^2] - E_n[L_i Y_i^b]^2 - E_n[L_i^2 (Y_i^a)^2] + E_n[L_i Y_i^a]^2$. We have $E_n[L_i Y_i^a]^2 - E_n[L_i Y_i^b]^2 = \text{ATE}^2 - \text{ATE}^2 + o_p(1) = o_p(1)$ by previous work. Next, we have $|E_n[L_i^2 (Y_i^b)^2] - E_n[L_i^2 (Y_i^a)^2]| = |E_n[L_i^2 (Y_i^b - Y_i^a)(Y_i^b + Y_i^a)]| \lesssim E_n[(Y_i^b - Y_i^a)^2]^{1/2} E_n[(Y_i^b + Y_i^a)^2]^{1/2}$. It's easy to see that $E_n[(Y_i^b + Y_i^a)^2]^{1/2} = O_p(1)$. We have $E_n[(Y_i^b - Y_i^a)^2] = c_p^2 E_n[(\gamma' h_i - \widehat{\gamma}' h_i)^2] = c_p^2 (\widehat{\gamma} - \gamma)' E_n[h_i h_i'](\widehat{\gamma} - \gamma) = o_p(1)$. This shows that $\text{Var}_n(L_i Y_i^b) - \text{Var}_n(L_i Y_i^a) = o_p(1)$, completing the proof for the first term.

Next consider $\widehat{v}_1^b - \widehat{v}_1$. We may expand

$$\widehat{v}_1^b - \widehat{v}_1 = n^{-1} \sum_{g \in \mathcal{G}_n^\nu} \frac{1}{a(g) - 1} \frac{1 - p}{p^2} \sum_{i \neq j \in g} D_i D_j (Y_i^a Y_j^a - Y_i^b Y_j^b).$$

Note that $Y_i^a Y_j^a - Y_i^b Y_j^b = (Y_i^a - Y_i^b) Y_j^a + Y_i^b (Y_j^a - Y_j^b) = c_p(\widehat{\gamma} - \gamma)'(h_i Y_j^a + Y_i^b h_j)$. Then by triangle inequality and Cauchy-Schwarz

$$\begin{aligned} |\widehat{v}_1^b - \widehat{v}_1| &= \left| c_p(\widehat{\gamma} - \gamma)' n^{-1} \sum_{g \in \mathcal{G}_n^\nu} \frac{1}{a(g) - 1} \frac{1 - p}{p^2} \sum_{i \neq j \in g} D_i D_j (h_i Y_j^a + Y_i^b h_j) \right| \\ &\lesssim |\widehat{\gamma} - \gamma|_2 \left(n^{-1} \sum_{g \in \mathcal{G}_n^\nu} \sum_{i \neq j \in g} |h_i|_2 |Y_j^a| + |Y_i^b| |h_j|_2 \right) \end{aligned}$$

Observe that

$$\sum_{i\neq j\in g} |h_i|_2|Y_j^a| \leq (1/2)\sum_{i\neq j\in g} |h_i|_2^2 + |Y_j^a|^2 = \frac{k-1}{2}\sum_{i\in g} |h_i|_2^2 + |Y_i^a|^2$$

Then since $\mathcal{G}_n^\nu$ is a partition of $[n]$ we have $|\widehat{v}_1^b - \widehat{v}_1| \lesssim |\widehat{\gamma} - \gamma|_2 E_n[|h_i|_2^2 + |Y_i^a|^2] = o_p(1)O_p(1) = o_p(1)$. Then by symmetry $\widehat{v}_0^b - \widehat{v}_0 = o_p(1)$ as well. A similar calculation shows that $\widehat{v}_{10}^b - \widehat{v}_{10} = o_p(1)$. Then we have shown that $\widehat{V}_b - \widehat{V} = o_p(1)$, which completes the proof. □

A.9 Proofs of Noncompliance Theorems

Proof of Theorems A.1, A.2, A.3. First we show Theorem A.1. Define $\widehat{\theta}^W(\alpha) = \bar{W}_1 - \bar{W}_0 - \alpha'(\bar{h}_1 - \bar{h}_0)c_p$ and similary for $\widehat{\theta}^D(\alpha)$. We claim that $\widehat{\tau}_{adj} = \widehat{\theta}^W(\gamma_W)/\widehat{\theta}^D(\gamma_D) + o_p(n^{-1/2})$. By algebra, we have

$$\widehat{\tau}_{adj} - \frac{\widehat{\theta}^W(\gamma_W)}{\widehat{\theta}^D(\gamma_D)} = \frac{\widehat{\theta}^D(\gamma_D)(\widehat{\gamma}_W - \gamma_W)'(\bar{h}_1 - \bar{h}_0)c_p + \widehat{\theta}^W(\gamma_W)(\gamma_D - \widehat{\gamma}_D)'(\bar{h}_1 - \bar{h}_0)c_p}{\widehat{\theta}^D(\gamma_D)\widehat{\theta}^D(\widehat{\gamma}_D)}$$

By Theorem 3.4, $\widehat{\theta}^D(\gamma_D), \widehat{\theta}^D(\widehat{\gamma}_D) = \tau_D + o_p(1)$ with $\tau_D > 0$, so the denominator is $O_p(1)$. The numerator is $o_p(n^{-1/2})$ since $\widehat{\theta}^D(\gamma_D), \widehat{\theta}^W(\gamma_W) = O_p(1)$ and $(\widehat{\gamma}_A - \gamma_A)'(\bar{h}_1 - \bar{h}_0)c_p = o_p(n^{-1/2})$ for $A = D, W$ by the first line of the proof of Theorem 3.4. Next, recall the potential outcomes $Q(z) = W(z) - \tau_L D(z)$ and define $\gamma_Q = \gamma_W - \tau_L\gamma_D$. Then we have

$$\frac{\widehat{\theta}^W(\gamma_W)}{\widehat{\theta}^D(\gamma_D)} - \tau_L = \frac{\widehat{\theta}^W(\gamma_W) - \tau_L\widehat{\theta}^D(\gamma_D)}{\widehat{\theta}^D(\gamma_D)} = \frac{\widehat{\theta}^Q(\gamma_Q)}{\widehat{\theta}^D(\gamma_D)}.$$

The ATE-like quantity $E[Q(1) - Q(0)] = 0$ by definition of τ_L. Then by Theorem 3.4, we have $\sqrt{n}\widehat{\theta}^Q(\gamma_Q) \Rightarrow \mathcal{N}(0, V_Q)$ with variance

$$V_Q = \mathrm{Var}(c_Q) + E\Big[\mathrm{Var}(b_Q - h'\gamma_Q|\psi)\Big] + E\left[\frac{\sigma_{1Q}^2(X)}{p} + \frac{\sigma_{0Q}^2(X)}{1-p}\right]. \tag{A.4}$$

The claim now follows by Slutsky since $\widehat{\theta}^D(\gamma_D) = E[D(1) - D(0)] + o_p(1)$ so that $\sqrt{n}(\widehat{\tau}_{adj} - \tau_L) = \sqrt{n}\widehat{\theta}^Q(\gamma_Q)/\widehat{\theta}^D(\gamma_D) + o_p(1) = \sqrt{n}\widehat{\theta}^Q(\gamma_Q)/E[D(1) - D(0)] + o_p(1)$.

Next, we prove Theorem A.2. By linearity of the balance function (Equation 2.2), we have $b_Q = b_W - \tau_L b_D$. The optimal coefficient is $\gamma_Q^* = E[\mathrm{Var}(h|\psi)]^{-1}E[\mathrm{Cov}(h, b_Q|\psi)] = E[\mathrm{Var}(h|\psi)]^{-1}(E[\mathrm{Cov}(h, b_W|\psi)] - \tau_L E[\mathrm{Cov}(h, b_D|\psi)]) = \gamma_W^* - \tau_L\gamma_D^*$. This shows that $\widehat{\tau}_{adj}$ is efficient if and only if $\gamma_W - \tau_L\gamma_D = \gamma_W^* - \tau_L\gamma_D^*$. In particular, this holds if $\gamma_W = \gamma_W^*$ and $\gamma_D = \gamma_D^*$. By the estimator representations in Section 3.4, the estimator $\widehat{\theta}_k^W = \bar{W}_1 - \bar{W}_0 - \widehat{\gamma}_{W,k}'(\bar{h}_1 - \bar{h}_0)c_p$ for $\widehat{\gamma}_{W,k} = \gamma_W^* + o_p(1)$ for $k \in \{PL, GO, TM\}$, and similarly for $\widehat{\theta}_k^D$. Then $\widehat{\tau}_L^k$ is efficient for each $k \in \{PL, GO, TM\}$.

Finally, we show Theorem A.3. With $\gamma_Q = \gamma_W - \tau_L \gamma_D$, define the "population" augmented potential outcomes $Q^b(z) = Q(z) - h'\gamma_Q$ and outcomes $Q_i^b = Q_i - h_i'\gamma_Q$. Let $\widehat{V}_Q^a$ denote the bracketed term in Equation 4.1, and let $\widehat{V}_Q^b$ denote the bracketed term with Q_i^a replaced by the population version Q_i^b. Note that we showed above that $\sqrt{n}(\bar{Q}_1^b - \bar{Q}_0^b) \Rightarrow N(0, V_Q)$. Then $\widehat{V}_Q^b = V_Q + o_p(1)$ by Theorem 4.1. Then it suffices to show that $\widehat{V}_Q^b - \widehat{V}_Q^a = o_p(1)$. To see this, note that we may write $Q_i^b = W_i - \beta' S_i$ and $Q_i^a = W_i - \widehat{\beta}' S_i$ with $\widehat{\beta} = \beta + o_p(1)$ for $\widehat{\beta} = (\widehat{\tau}_L^k, \widehat{\gamma}_Q)$, $\beta = (\tau_L, \gamma_Q)$ and $S_i = (D_i, h_i)$. Then the fact that $\widehat{V}_Q^b - \widehat{V}_Q^a = o_p(1)$ for outcomes of this form and $\widehat{\beta} = \beta + o_p(1)$ is exactly what we showed in the main claim in the proof of Theorem 4.1. This finishes the proof. □

A.10 Technical Lemmas

Lemma A.5 (Conditional Convergence)**.** *Let $(\mathcal{G}_n)_{n\geq 1}$ and $(A_n)_{n\geq 1}$ a sequence of σ-algebras and RV's. Then the following results hold*

(i) $E[|A_n||\mathcal{G}_n] = o_p(1)/O_p(1) \implies A_n = o_p(1)/O_p(1)$.

(ii) $\text{Var}(A_n|\mathcal{G}_n) = o_p(c_n^2)/O_p(c_n^2) \implies A_n - E[A_n|\mathcal{G}_n] = o_p(c_n)/O_p(c_n)$ *for any positive sequence* $(c_n)_n$.

(iii) If $(A_n)_{n\geq 1}$ *has* $A_n \leq \bar{A} < \infty$ $\mathcal{G}_n$*-a.s.* $\forall n$ *and* $A_n = o_p(1) \implies E[|A_n||\mathcal{G}_n] = o_p(1)$.

See Appendix C of Cytrynbaum (2023) for the proof.

Lemma A.6. *Let $(a_i), (b_i), (c_i)$ be positive scalar arrays for $i \in I$ for some index set I. Then we have $\sum_{\substack{i,j,s\in I \\ i\neq j, j\neq s}} a_i b_j c_s \leq 3\sum_{i\in I}(a_i^3 + b_i^3 + c_i^3)$.*

Proof. Note that by AM-GM inequality and Jensen, for non-negative x, y, z we have $xyz \leq ((1/3)(x+y+z))^3 \leq (1/3)(x^3+y^3+z^3)$. Applying this gives

$$\sum_{\substack{i,j,s \\ i\neq j, j\neq s}} a_i b_j c_s \leq \left(\sum_i a_i\right)\left(\sum_j b_j\right)\left(\sum_s c_s\right)$$
$$\leq (1/3)\left[\left(\sum_i a_i\right)^3 + \left(\sum_j b_j\right)^3 + \left(\sum_s c_s\right)^3\right] \leq 3\sum_i (a_i^3 + b_i^3 + c_i^3).$$

□

Lemma A.7 (Group OLS)**.** *Let $h, w : \mathcal{X} \to \mathbb{R}$. Denote $h_i = h(X_i)$ and $w_i = w(X_i)$ and suppose $E[h_i|\psi_i = \psi]$ and $E[w_i|\psi_i = \psi]$ are Lipschitz continuous. Suppose $E[h_i^4] < \infty$*

and $E[w_i^4] < \infty$. *Let* $\epsilon_i^d = Y_i(d) - m_d(X_i)$ *for* $d \in \{0,1\}$. *Then we have*

$$A_n = n^{-1}\sum_g \left(k^{-1}\sum_{i\in g}\frac{h_i(D_i-p)}{p-p^2}\right)\left(k^{-1}\sum_{i\in g}\frac{w_i(D_i-p)}{p-p^2}\right) = \frac{E[\mathrm{Cov}(h,w|\psi)]}{a(k-a)} + o_p(1).$$

$$B_n = n^{-1}\sum_g \left(k^{-1}\sum_{i\in g}\frac{h_i(D_i-p)}{p-p^2}\right)\left(k^{-1}\sum_{i\in g} w_i\right) = O_p(n^{-1/2}).$$

$$C_n = \sum_g \left(k^{-1}\sum_{i\in g}\frac{h_i(D_i-p)}{p-p^2}\right)\left(k^{-1}\sum_{i\in g}\frac{D_i\epsilon_i^1}{p} - \frac{(1-D_i)\epsilon_i^0}{1-p}\right) = O_p(n^{-1/2}).$$

Proof. Define $\bar{h}_{g1} = a^{-1}\sum_{i\in g} h_i \mathbb{1}(D_i = 1)$, $\bar{h}_{g0} = (k-a)^{-1}\sum_{i\in g} h_i\mathbb{1}(D_i = 0)$, and $\bar{w}_g = k^{-1}\sum_{i\in g} w_i$. Recall that $g \in \sigma(\psi_{1:n}, \pi_n)$ for each g and $D_{1:n} \in \sigma(\psi_{1:n}, \pi_n, \tau)$ for an exogenous variable $\tau \perp\!\!\!\perp (X_{1:n}, Y(0)_{1:n}, Y(1)_{1:n})$ used to randomize treatments. Notice that $k^{-1}\sum_{i\in g}\frac{h_i(D_i-p)}{p-p^2} = \bar{h}_{g1} - \bar{h}_{g0}$. First consider B_n. By Lemma C.10 of Cytrynbaum (2023), we have $E[B_n|X_{1:n}, \pi_n] = 0$. Next, we have

$$E[B_n^2|X_{1:n},\pi_n] = E\left[n^{-2}\sum_{g,g'}\left(k^{-1}\sum_{i\in g}\frac{h_i(D_i-p)}{p-p^2}\right)\left(k^{-1}\sum_{i\in g'}\frac{h_i(D_i-p)}{p-p^2}\right)\bar{w}_g\bar{w}_{g'}\middle| X_{1:n},\pi_n\right]$$
$$= E\left[n^{-2}\sum_g\left(k^{-1}\sum_{i\in g}\frac{h_i(D_i-p)}{p-p^2}\right)^2 \bar{w}_g^2 \middle| X_{1:n},\pi_n\right].$$

The second equality follows by Lemma C.10 of Cytrynbaum (2023), since $\mathrm{Cov}(D_i, D_j|X_{1:n}, \pi_n) = 0$ if i, j are in different groups. We may calculate

$$E\left[\left(k^{-1}\sum_{i\in g}\frac{h_i(D_i-p)}{p-p^2}\right)^2\middle| X_{1:n},\pi_n\right] = \frac{1}{k^2(p-p^2)^2}\sum_{i\in g} h_i^2\,\mathrm{Var}(D_i|X_{1:n},\pi_n)$$
$$+\frac{1}{k^2(p-p^2)^2}\sum_{i\neq j\in g} h_ih_j\,\mathrm{Cov}(D_i,D_j|X_{1:n},\pi_n) = \frac{1}{k^2(p-p^2)}\left[\sum_{i\in g}h_i^2 - (k-1)^{-1}\sum_{i\neq j\in g}h_ih_j\right].$$

Note that $\sum_{i\neq j\in g}|h_ih_j| \le \left(\sum_{i\in g}|h_i|\right)^2 = k^2\left(k^{-1}\sum_{i\in g}|h_i|\right)^2 \le k\sum_{i\in g}|h_i|^2$. The final inequality by Jensen. Then by triangle inequality, a simple calculation gives

$$\frac{1}{k^2}\left|\sum_{i\in g}h_i^2 - (k-1)^{-1}\sum_{i\neq j\in g}h_ih_j\right| \le \frac{1}{k^2}\frac{2k-1}{k-1}\sum_{i\in g}h_i^2 \le 3k^{-2}\sum_{i\in g}h_i^2.$$

Then continuing from above

$$E[B_n^2|X_{1:n},\pi_n] \lesssim k^{-2}n^{-2}\sum_g \left(\sum_{i\in g} h_i^2\right)\left(\sum_{i\in g} w_i\right)^2 \le \frac{1}{kn^2}\sum_g \left(\sum_{i\in g} h_i^2\right)\left(\sum_{i\in g} w_i^2\right)$$
$$\le \frac{1}{2kn^2}\sum_g \left[\left(\sum_{i\in g} h_i^2\right)^2 + \left(\sum_{i\in g} w_i^2\right)^2\right] = (2n)^{-1}E_n[h_i^4 + w_i^4] = O_p(n^{-1}).$$

The second inequality follows from Jensen, and the third by Young's inequality. The first equality by Jensen and final equality by our moment assumption. Then by Lemma A.5, $B_n = O_p(n^{-1/2})$.

Next, consider A_n. Using the within-group covariances above, we compute

$$E[A_n|X_{1:n},\pi_n] = \frac{1}{nk^2(p-p^2)^2}\sum_g \sum_{i,j\in g} \mathrm{Cov}(D_i, D_j|X_{1:n},\pi_n)h_i w_j$$
$$= \frac{1}{nk^2(p-p^2)^2}\sum_g \left(\sum_{i\in g}(p-p^2)h_i w_i - \sum_{i\neq j\in g}\frac{a(k-a)}{k^2(k-1)}h_i w_j\right)$$
$$= \frac{1}{k^2(p-p^2)}\left(E_n[h_i w_i] - \frac{1}{n(k-1)}\sum_g \sum_{i\neq j\in g} h_i w_j\right).$$

Define $u_i = w_i - E[w_i|\psi_i]$ and $v_i = h_i - E[h_i|\psi_i]$. Consider the second term. We have

$$n^{-1}\sum_g \sum_{i\neq j\in g} h_i w_j = n^{-1}\sum_g \sum_{i\neq j\in g}(E[h_i|\psi_i] + v_i)(E[w_j|\psi_j] + u_j) \equiv \sum_{l=1}^4 A_{n,l}.$$

First, note that for any scalars $a_i b_j + a_j b_i = a_i b_i + a_j b_j + (a_i - a_j)(b_j - b_i)$. Then we have

$$A_{n,1} \equiv n^{-1}\sum_g \sum_{i\neq j\in g} E[h_i|\psi_i]E[w_j|\psi_j] = n^{-1}\sum_g \sum_{i<j\in g} E[h_i|\psi_i]E[w_j|\psi_j] + E[h_j|\psi_j]E[w_i|\psi_i]$$
$$= n^{-1}\sum_g \sum_{i<j\in g} E[h_i|\psi_i]E[w_i|\psi_i] + E[h_j|\psi_j]E[w_j|\psi_j]$$
$$+ n^{-1}\sum_g \sum_{i<j\in g}(E[h_i|\psi_i] - E[h_j|\psi_j])(E[w_j|\psi_j] - E[w_i|\psi_i]) \equiv B_{n,1} + C_{n,1}.$$

By counting ordered tuples (i,j), it's easy to see that

$$B_{n,1} = n^{-1}\sum_g \sum_{i\in g}(k-1)E[h_i|\psi_i]E[w_i|\psi_i] = (k-1)E_n[E[h_i|\psi_i]E[w_i|\psi_i]]$$
$$= (k-1)E[E[h_i|\psi_i]E[w_i|\psi_i]] + o_p(1) = (k-1)(E[h_i w_i] - E[v_i u_i]) + o_p(1).$$

For the second term, by our Lipschitz assumptions we have $|C_{n,1}| \lesssim n^{-1}\sum_g \sum_{i<j\in g}|\psi_i -$

$\psi_j|_2^2 = o_p(1)$. Next, claim that $A_{n,l} = o_p(1)$ for $l = 2, 3, 4$. For instance, we have

$$E[A_{n,2}|\psi_{1:n}, \pi_n] = n^{-1} \sum_g \sum_{i \neq j \in g} E\left[E[h_i|\psi_i] u_j | \psi_{1:n}, \pi_n\right] = 0.$$

Since $E[u_j|\psi_{1:n}, \pi_n] = E[u_j|\psi_j] = 0$ by Lemma 9.21 of Cytrynbaum (2022). Moreover, we have

$$E[A_{n,2}^2|\psi_{1:n}, \pi_n] = n^{-2} \sum_{g,g'} \sum_{i \neq j \in g} \sum_{s \neq t \in g'} E[h_i|\psi_i] E[h_s|\psi_s] E[u_j u_t|\psi_{1:n}, \pi_n].$$

For $j \neq t$, we have $E[u_j u_t|\psi_{1:n}, \pi_n] = E[u_j|\psi_j] E[u_t|\psi_t] = 0$ by Lemma 9.21 of the paper above. Since the groups g are disjoint, and using $E[u_j^2|\psi_{1:n}, \pi_n] = E[u_j^2|\psi_j]$

$$\begin{aligned}
E[A_{n,2}^2|\psi_{1:n}, \pi_n] &= n^{-2} \sum_g \sum_{\substack{i,j,s \in g \\ i \neq j, j \neq s}} E[h_i|\psi_i] E[h_s|\psi_s] E[u_j^2|\psi_j] \\
&\leq 3n^{-2} \sum_g \sum_{i \in g} 2E[h_i|\psi_i]^3 + E[u_i^2|\psi_i]^3 \\
&= 3n^{-1} E_n[2E[h_i|\psi_i]^3 + E[u_i^2|\psi_i]^3] = O_p(n^{-1}).
\end{aligned}$$

Then we have shown $A_{n,2} = O_p(n^{-1/2})$ by Lemma A.5. The proof for $l = 3, 4$ is almost identical. Summarizing, the work above has shown that

$$\begin{aligned}
E[A_n|X_{1:n}, \pi_n] &= \frac{1}{k^2(p - p^2)} \left(E_n[h_i w_i] - \frac{1}{k-1}(k-1)(E[h_i w_i] - E[v_i u_i]) \right) + o_p(1) \\
&= \frac{1}{k^2(p - p^2)} E[v_i u_i] + o_p(1) = \frac{E[\mathrm{Cov}(h, w|\psi)]}{a(k-a)} + o_p(1).
\end{aligned}$$

Next, we claim that $\mathrm{Var}(A_n|X_{1:n}, \pi_n) = o_p(1)$. Define $\Delta_{h,g} = k^{-1} \sum_{i \in g} \frac{h_i(D_i - p)}{p - p^2}$, then

$$\mathrm{Var}(A_n|X_{1:n}, \pi_n) = n^{-2} \sum_{g,g'} \mathrm{Cov}\left(\Delta_{h,g}\Delta_{w,g}, \Delta_{h,g'}\Delta_{w,g'}|X_{1:n}, \pi_n\right).$$

Note that $\Delta_{h,g}\Delta_{w,g} \perp\!\!\!\perp \Delta_{h,g'}\Delta_{w,g'}|X_{1:n}, \pi_n$ for $g \neq g'$, since treatment assignments are (conditionally) independent between groups. Then the on-diagonal terms are

$$\begin{aligned}
\mathrm{Var}(A_n|X_{1:n}, \pi_n) &= n^{-2} \sum_g \mathrm{Var}\left(\left(k^{-1} \sum_{i \in g} \frac{h_i(D_i - p)}{p - p^2} \right) \left(k^{-1} \sum_{i \in g} \frac{w_i(D_i - p)}{p - p^2} \right) \Bigg| X_{1:n}, \pi_n \right) \\
&= n^{-2} k^{-4} (p - p)^{-4} \sum_g \mathrm{Var}\left(\sum_{i,j \in g} h_i w_j (D_i - p)(D_j - p) \Bigg| X_{1:n}, \pi_n \right).
\end{aligned}$$

The inner variance term can be expanded as

$$\sum_{i,j\in g}\sum_{s,t\in g} h_i w_j h_s w_t \operatorname{Cov}\left((D_i-p)(D_j-p),(D_s-p)(D_t-p)\middle| X_{1:n},\pi_n\right).$$

We have $|\operatorname{Cov}((D_i-p)(D_j-p),(D_s-p)(D_t-p)|X_{1:n},\pi_n)|\leq 2$ since $|(D_i-p)|\leq 1$ for all $i\in[n]$. Using Lemma 9.17 in Cytrynbaum (2022), the previous display is bounded above by $\sum_{i,j\in g}\sum_{s,t\in g}|h_i w_j h_s w_t|\cdot 2\leq 2k^3\sum_{i\in g}(h_i^4+w_i^4)$. Putting this all together,

$$\begin{aligned}\operatorname{Var}(A_n|X_{1:n},\pi_n) &\leq 2n^{-2}k^{-4}(p-p)^{-4}k^3\sum_g\sum_{i\in g}(h_i^4+w_i^4)\\ &= 2n^{-1}k^{-1}(p-p)^{-4}E_n[h_i^4+w_i^4]=O_p(n^{-1})\end{aligned}$$

By conditional Markov, this shows that $A_n-E[A_n|X_{1:n},\pi_n]=O_p(n^{-1/2})$. Then we have shown that $A_n=\frac{E[\operatorname{Cov}(h,w|\psi)]}{a(k-a)}+o_p(1)$.

Finally, we consider C_n. Note that $g,D_{1:n}\in\sigma(X_{1:n},\pi_n,\tau)$ and $E[\epsilon_i^d|X_{1:n},\pi_n,\tau]=E[\epsilon_i^d|X_i]=0$ for $d=0,1$ by Lemma 9.21 of Cytrynbaum (2022), so we have $E[C_n|X_{1:n},\pi_n,\tau]=0$. Next, we claim that $E[C_n^2|X_{1:n},\pi_n,\tau]=O_p(n^{-1})$. Note that C_n^2 can be written

$$\frac{1}{n^2k^4}\sum_{g,g'}\left(\sum_{i,j\in g}\sum_{s,t\in g'}\frac{h_i(D_i-p)}{p-p^2}\left(\frac{D_j\epsilon_j^1}{p}-\frac{(1-D_j)\epsilon_j^0}{1-p}\right)\frac{h_s(D_s-p)}{p-p^2}\left(\frac{D_t\epsilon_t^1}{p}-\frac{(1-D_t)\epsilon_t^0}{1-p}\right)\right).$$

We have $E[\epsilon_j^d\epsilon_t^{d'}|X_{1:n},\pi_n,\tau]=E[\epsilon_j^d|X_j]E[\epsilon_t^{d'}|X_t]=0$ for any $j\neq t$ by Lemma 9.21 of Cytrynbaum (2022). By group disjointness, the term $E[C_n^2|X_{1:n},\pi_n,\tau]$ simplifies to

$$\frac{1}{n^2k^4}\sum_g\left(\sum_{i,j,s\in g}\frac{h_i(D_i-p)}{p-p^2}\frac{h_s(D_s-p)}{p-p^2}E\left[\left(\frac{D_j\epsilon_j^1}{p}-\frac{(1-D_j)\epsilon_j^0}{1-p}\right)^2\middle| X_{1:n},\pi_n,\tau\right]\right).$$

We have $E[(\epsilon_i^d)^2|X_{1:n},\pi_n,\tau]=E[(\epsilon_i^d)^2|X_i]=\sigma_d^2(X_i)$. Then by Young's inequality and Lemma 9.21 of the paper above

$$E\left[\left(\frac{D_j\epsilon_j^1}{p}-\frac{(1-D_j)\epsilon_j^0}{1-p}\right)^2\middle| X_{1:n},\pi_n,\tau\right]\leq 2(p\wedge(1-p))^{-1}(\sigma_1^2(X_j)+\sigma_0^2(X_j)).$$

Taking the absolute value of the second to last display and using triangle inequality gives the upper bound

$$
\begin{aligned}
&2[n^2k^4(p-p^2)^2(p\wedge(1-p))]^{-1}\sum_g\left(\sum_{i,j,s\in g}|h_ih_s|(\sigma_1^2(X_j)+\sigma_0^2(X_j))\right)\\
&\lesssim n^{-2}\sum_g\left(\sum_{i,j,s\in g}|h_ih_s|^2+(\sigma_1^2(X_j)+\sigma_0^2(X_j))^2\right)\\
&\le n^{-1}k^2E_n[(\sigma_1^2(X_i)+\sigma_0^2(X_i))^2]+n^{-2}k\sum_g\sum_{i,s\in g}|h_ih_s|^2.
\end{aligned}
$$

By Young's inequality and assumption $E[E_n[(\sigma_1^2(X_i)+\sigma_0^2(X_i))^2]]\le 2E[\sigma_1^2(X_i)^2+\sigma_0^2(X_i)^2]<\infty$. For the second term, using Jensen we have

$$
n^{-1}\sum_g\sum_{i,s\in g}|h_ih_s|^2=n^{-1}\sum_g\left(\sum_{i\in g}|h_i|^2\right)^2\le kn^{-1}E_n[h_i^4]=O_p(1).
$$

Then we have shown that $E[C_n^2|X_{1:n},\pi_n,\tau]=O_p(n^{-1})$, so by conditional Markov inequality in Lemma A.5, $C_n=O_p(n^{-1/2})$. This finishes the proof. □

Lemma A.8 (Partialled Lin). *Under assumptions, $E_n[\check{h}_iz_i]=o_p(1)$. Also, we have*

$$
\begin{aligned}
E_n[D_i\check{h}_i\check{h}_i']&=\frac{p(k-1)}{k}E[\mathrm{Var}(h|\psi)]+o_p(1)\quad E_n[\check{h}_i\check{h}_i']=\frac{k-1}{k}E[\mathrm{Var}(h|\psi)]+o_p(1)\\
E_n[D_i\check{h}_iY_i]&=\frac{p(k-1)}{k}E[\mathrm{Cov}(h,m_1|\psi)]+o_p(1)\\
E_n[(1-D_i)\check{h}_iY_i]&=\frac{(1-p)(k-1)}{k}E[\mathrm{Cov}(h,m_0|\psi)]+o_p(1).
\end{aligned}
$$

Proof. First, observe that

$$
\check{h}_i=h_i-k^{-1}\sum_{j\in g(i)}h_j=\frac{k-1}{k}\cdot h_i-k^{-1}\sum_{j\in g(i)\setminus\{i\}}h_j=k^{-1}\sum_{j\in g(i)\setminus\{i\}}(h_i-h_j).
$$

Note that $E_n[D_i\check{h}_i\check{h}_i]=E_n[(D_i-p)\check{h}_i\check{h}_i]+pE_n[\check{h}_i\check{h}_i]$. We claim that $E_n[(D_i-p)\check{h}_i\check{h}_i]=O_p(n^{-1/2})$. For $1\le t,t'\le d_h$, by Lemma A.2 of Cytrynbaum (2022) and Cauchy-Schwarz we have $\mathrm{Var}(\sqrt{n}E_n[(D_i-p)\check{h}_{it}\check{h}_{it'}]|X_{1:n},\pi_n)\le 2E_n[\check{h}_{it}^2\check{h}_{it'}^2]\le 2E_n[\check{h}_{it}^4]^{1/2}E_n[\check{h}_{it'}^4]^{1/2}$. Next, note that by Jensen's followed by Young's inequality

$$
\begin{aligned}
\check{h}_{it}^4&=\frac{(k-1)^4}{k^4}\left(\frac{1}{k-1}\sum_{j\in g(i)\setminus\{i\}}(h_{it}-h_{jt})\right)^4\le\frac{(k-1)^3}{k^4}\sum_{j\in g(i)\setminus\{i\}}(h_{it}-h_{jt})^4\\
&\le 8\frac{(k-1)^3}{k^4}\sum_{j\in g(i)\setminus\{i\}}(h_{it}^4+h_{jt}^4)\le 8\frac{(k-1)^3}{k^4}\left((k-1)h_{it}^4+\sum_{j\in g(i)\setminus\{i\}}h_{jt}^4\right).
\end{aligned}
$$

By counting, we have $E_n\left[\sum_{j\in g(i)\setminus\{i\}} h_{jt}^4\right] = (k-1)E_n[h_{it}^4]$. Putting this all together, $E_n[\check{h}_{it}^4] \lesssim E_n[h_{it}^4] = O_p(1)$. Then $\text{Var}(\sqrt{n}E_n[(D_i - p)\check{h}_{it}\check{h}_{it'}]|X_{1:n}, \pi_n) = O_p(1)$ so that $E_n[(D_i - p)\check{h}_{it}\check{h}_{it'}] = O_p(n^{-1/2})$ by Lemma A.5. Then it suffices to show the claim for $E_n[\check{h}_i\check{h}_i]$. Let $f_{it} = E[h_t(X_i)|\psi_i]$ and write $h_{it} = f_{it} + u_{it}$. Then we have

$$
\begin{aligned}
E_n[\check{h}_{it}\check{h}_{it'}] =& \frac{1}{nk^2}\sum_i \left(\sum_{j\in g(i)\setminus\{i\}} h_{it} - h_{jt}\right)\left(\sum_{l\in g(i)\setminus\{i\}} h_{it'} - h_{lt'}\right) \\
=& \frac{1}{nk^2}\sum_i D_i \sum_{j,l\in g(i)\setminus\{i\}} (h_{it} - h_{jt})(h_{it'} - h_{lt'}).
\end{aligned}
$$

We can expand the expression above as

$$
\begin{aligned}
&\frac{1}{nk^2}\sum_i \sum_{j,l\in g(i)\setminus\{i\}} \Big[(f_{it} - f_{jt})(f_{it'} - f_{lt'}) + (f_{it} - f_{jt})(u_{it'} - u_{lt'}) \\
&+ (u_{it} - u_{jt})(f_{it'} - f_{lt'}) + (u_{it} - u_{jt})(u_{it'} - u_{lt'})\Big] \equiv A_n + B_n + C_n + D_n.
\end{aligned}
$$

First consider A_n. By the Lipschitz assumption in 3.1 and Young's inequality

$$
\begin{aligned}
|A_n| &\le \frac{1}{nk^2}\sum_i \sum_{j,l\in g\setminus\{i\}} |f_{it} - f_{jt}||f_{it'} - f_{lt'}| \lesssim \frac{1}{nk^2}\sum_i \sum_{j,l\in g\setminus\{i\}} |\psi_i - \psi_j|_2|\psi_i - \psi_l|_2 \\
&\le \frac{2}{nk^2}\sum_i \sum_{j,l\in g\setminus\{i\}} (|\psi_i - \psi_j|_2^2 + |\psi_i - \psi_l|_2^2) = \frac{4(k-1)}{nk^2}\sum_g \sum_{i,j\in g} |\psi_i - \psi_j|_2^2 = o_p(1).
\end{aligned}
$$

The second to last equality by counting and the final equality by Assumption 2.1. Next consider B_n. Note that each $g \in \sigma(\psi_{1:n}, \pi_n)$ and $E[u_{it}|\psi_{1:n}, \pi_n] = E[u_{it}|\psi_i] = 0$, so $E[B_n|\psi_{1:n}, \pi_n] = 0$. We can rewrite the sum

$$
\sum_i \sum_{j,l\in g\setminus\{i\}} (f_{it} - f_{jt})(u_{it'} - u_{lt'}) = \sum_g \sum_{\substack{i,j,l\in g \\ j,l\neq i}} (f_{it} - f_{jt})(u_{it'} - u_{lt'}).
$$

Then we may compute $\text{Var}(\sqrt{n}B_n|\psi_{1:n}, \pi_n) = E[nB_n^2|\psi_{1:n}, \pi_n]$ as follows. By Lemma 9.21 of Cytrynbaum (2022), $E[u_{it'}u_{jt'}|\psi_{1:n}, \pi_n] = 0$ for any $g(i) \neq g(j)$, so we only consider the diagonal

$$
\begin{aligned}
0 &\le \frac{1}{nk^4}\sum_g \sum_{\substack{i,j,l\in g\\ j,l\neq i}} \sum_{\substack{a,b,c\in g\\ b,c\neq a}} E[(f_{it}-f_{jt})(f_{at}-f_{bt})(u_{it'}-u_{lt'})(u_{at'}-u_{ct'})|\psi_{1:n},\pi_n] \\
&\le n^{-1}\sum_g \sum_{\substack{i,j,l\in g\\ j,l\neq i}} \sum_{\substack{a,b,c\in g\\ b,c\neq a}} |f_{it}-f_{jt}||f_{at}-f_{bt}||E[(u_{it'}-u_{lt'})(u_{at'}-u_{ct'})|\psi_{1:n},\pi_n]| \\
&\lesssim n^{-1}\sum_g \max_{i,j\in g}|\psi_i-\psi_j|_2^2 \sum_{\substack{i,j,l\in g\\ j,l\neq i}} \sum_{\substack{a,b,c\in g\\ b,c\neq a}} |E[(u_{it'}-u_{lt'})(u_{at'}-u_{ct'})|\psi_{1:n},\pi_n]|.
\end{aligned}
$$

Next, by Lemma 9.21 of Cytrynbaum (2022), $E[(u_{it'}-u_{lt'})(u_{at'}-u_{ct'})|\psi_{1:n},\pi_n]$ is

$$
\delta_{ai}E[u_{it'}^2|\psi_i]-\delta_{la}E[u_{at'}^2|\psi_a]-\delta_{ci}E[u_{it'}^2|\psi_i]+\delta_{lc}E[u_{lt'}^2|\psi_l].
$$

Applying the triangle inequality and summing out using this relation, the above is

$$
\begin{aligned}
&\le \frac{4k(k-1)^3}{n}\sum_g \max_{i,j\in g}|\psi_i-\psi_j|_2^2 \sum_{i\in g} E[u_{it'}^2|\psi_i] \\
&\lesssim n^{-1}\sum_g \left(\max_{i,j\in g}|\psi_i-\psi_j|_2^4 + \sum_{i\in g} E[u_{it'}^2|\psi_i]^2\right) \\
&\le n^{-1}\sum_g \mathrm{Diam}(\mathrm{Supp}(\psi))^2 \sum_{i,j\in g}|\psi_i-\psi_j|_2^2 + E_n[E[u_{it'}^2|\psi_i]^2].
\end{aligned}
$$

We claim that $E[u_{it'}^4]<\infty$. Note that $E[u_{it'}^4]=E[(h_{it'}-f_{it'})^4]\le 8E[h_{it'}^4]+8E[f_{it'}^4]$ by Young's inequality. We have $E[h_{it'}^4]<\infty$ by assumption. Note that $E[f_{it'}^4]\le C_f|\psi_i|^4\le C_f\,\mathrm{Diam}(\mathrm{Supp}(\psi))^4<\infty$ by Assumption 3.1, with Lipschitz constant C_f. Then $E[u_{it'}^4]<\infty$, so $E[E_n[E[u_{it'}^2|\psi_i]^2]] = E[E[u_{it'}^2|\psi_i]^2] \le E[u_{it'}^4] < \infty$. The inequality follows by conditional Jensen and tower law. Then $E_n[E[u_{it'}^2|\psi_i]^2 = O_p(1)$ by Markov inequality. Then using Assumption 2.1 in the display above, we have shown $E[nB_n^2|\psi_{1:n},\pi_n]=O_p(1)$ and by Lemma A.5 we have shown $B_n=O_p(n^{-1/2})$. We have $C_n=O_p(n^{-1/2})$ by symmetry. Finally, consider D_n. By Lemma 9.21 of Cytrynbaum (2022) compute $E[(u_{it}-u_{jt})(u_{it'}-u_{lt'})|\psi_{1:n},\pi_n]=E[u_{it}u_{it'}|\psi_i]+E[u_{jt}u_{jt'}|\psi_j]\delta_{jl}$ for $j,l\neq i$. Then we calculate

$$
\begin{aligned}
E[D_n|\psi_{1:n},\pi_n] &= \frac{1}{nk^2}\sum_i \sum_{j,l\in g(i)\setminus\{i\}} E[u_{it}u_{it'}|\psi_i]+E[u_{jt}u_{jt'}|\psi_j]\mathbb{1}(j=l) \\
= \frac{1}{nk^2}&\sum_i (k-1)^2E[u_{it}u_{it'}|\psi_i] + \frac{1}{nk^2}\sum_i \sum_{j\in g(i)\setminus\{i\}} E[u_{jt}u_{jt'}|\psi_j] \\
= \frac{(k-1)^2}{nk^2}&\sum_i E[u_{it}u_{it'}|\psi_i] + \frac{k-1}{nk^2}\sum_i E[u_{it}u_{it'}|\psi_i] = \frac{k(k-1)}{nk^2}\sum_i E[u_{it}u_{it'}|\psi_i].
\end{aligned}
$$

Now $E[E[u_{it}u_{it'}|\psi_i]^2] \le E[u_{it}^2u_{it'}^2] \le 2E[u_{it}^4]+2E[u_{it'}^4] < \infty$ by Jensen, tower law, Young's, and work above. Then by Chebyshev $\frac{(k-1)}{nk}\sum_i E[u_{it}u_{it'}|\psi_i] = \frac{k-1}{k}E[u_{it}u_{it'}] + O_p(n^{-1/2}) = \frac{k-1}{k}E[\mathrm{Cov}(h_{it}, h_{it'}|\psi_i)] + O_p(n^{-1/2})$. Then we have shown $E[D_n|\psi_{1:n}, \pi_n] = \frac{k-1}{k}E[\mathrm{Cov}(h_{it}, h_{it'}|\psi_i)] + O_p(n^{-1/2})$. Next, we claim that $\mathrm{Var}(\sqrt{n}D_n|\psi_{1:n}, \pi_n) = O_p(1)$. Following the steps above for B_n replacing terms shows that $\mathrm{Var}(\sqrt{n}D_n|\psi_{1:n}, \pi_n)$ is

$$0 \le \frac{1}{nk^4}\sum_g \sum_{\substack{i,j,l\in g\\ j,l\neq i}} \sum_{\substack{a,b,c\in g\\ b,c\neq a}} \mathrm{Cov}((u_{it}-u_{jt})(u_{it'}-u_{lt'}), (u_{at}-u_{bt})(u_{at'}-u_{ct'})|\psi_{1:n}, \pi_n).$$

For any variables A, B, $|\mathrm{Cov}(A,B)| \le |E[AB]| + |E[A]E[B]| \le 2|A|_2|B|_2$ by Cauchy-Schwarz and increasing $L_p(\mathbb{P})$ norms. By Young's inequality, $(a-b)^4 \le 8(a^4+b^4)$ for any $a, b \in \mathbb{R}$. Then using these facts

$$\begin{aligned}
&|\mathrm{Cov}((u_{it}-u_{jt})(u_{it'}-u_{lt'}), (u_{at}-u_{bt})(u_{at'}-u_{ct'})|\psi_{1:n}, \pi_n)| \\
&\le 2E[(u_{it}-u_{jt})^2(u_{it'}-u_{lt'})^2|\psi_{1:n}, \pi_n]^{1/2}E[(u_{at}-u_{bt})^2(u_{at'}-u_{ct'})^2|\psi_{1:n}, \pi_n]^{1/2} \\
&\le 4E[(u_{it}-u_{jt})^2(u_{it'}-u_{lt'})^2|\psi_{1:n}, \pi_n] + 4E[(u_{at}-u_{bt})^2(u_{at'}-u_{ct'})^2|\psi_{1:n}, \pi_n] \\
&\le 2E[(u_{it}-u_{jt})^4 + (u_{it'}-u_{lt'})^4|\psi_{1:n}, \pi_n] + 2E[(u_{at}-u_{bt})^4 + (u_{at'}-u_{ct'})^4|\psi_{1:n}, \pi_n] \\
&\le 16(E[u_{it}^4 + u_{jt}^4 + u_{it'}^4 + u_{lt'}^4|\psi_{1:n}, \pi_n] + E[u_{at}^4 + u_{bt}^4 + u_{at'} + u_{ct'}^4|\psi_{1:n}, \pi_n]) \\
&= 16(2E[u_{it}^4|\psi_i] + E[u_{jt}^4|\psi_j] + E[u_{lt'}^4|\psi_l] + 2E[u_{at}^4|\psi_a] + E[u_{bt}^4|\psi_b] + E[u_{ct'}^4|\psi_c]).
\end{aligned}$$

Plugging this bound in above and summing out gives

$$\mathrm{Var}(\sqrt{n}D_n|\psi_{1:n}, \pi_n) \le \frac{32k^5}{nk^4}\sum_g\sum_{i\in g} E[u_{it}^4|\psi_i] \asymp E_n[E[u_{it}^4|\psi_i]] = O_p(1).$$

The final equality by Markov since $E[u_{it}^4] < \infty$. Then by conditional Markov A.5 we have $D_n = \frac{k-1}{k}E[\mathrm{Cov}(h_{it}, h_{it'}|\psi_i)] + O_p(n^{-1/2})$. Since t, t' were arbirary, this shows $E_n[\check{h}_i\check{h}_i'] = E[\mathrm{Var}(h|\psi)] + o_p(1)$.

Next, consider $E_n[D_i\check{h}_iY_i] = E_n[(D_i-p)\check{h}_iY_i(1)] + pE_n[\check{h}_iY_i(1)]$. We claim that $E_n[(D_i - p)\check{h}_iY_i(1)] = O_p(n^{-1/2})$. For $1 \le t \le d_h$, by Lemma A.2 of Cytrynbaum (2023), and Cauchy-Schwarz

$$\mathrm{Var}(\sqrt{n}E_n[(D_i-p)\check{h}_{it}Y_i(1)]|X_{1:n}, Y(1)_{1:n}, \pi_n) \le 2E_n[\check{h}_{it}^2Y_i(1)^2] \le 2E_n[\check{h}_{it}^4]^{1/2}E_n[Y_i(1)^4]^{1/2}.$$

Note that $E_n[Y_i(1)^4] = O_p(1)$ by Markov inequality and Assumption 3.1 and $E_n[\check{h}_{it}^4] = O_p(1)$ was shown above. Then by Lemma A.5 (conditional Markov), this shows the claim. Then it suffices to analyze $E_n[\check{h}_iY_i(1)]$. Let $g_i = E[Y_i(1)|\psi_i]$ and $v_i = Y_i(1) - g_i$

with $E[v_i|\psi_i] = 0$. Then as above we may expand

$$
\begin{aligned}
E_n[\check{h}_i Y_i(1)] &= \frac{1}{nk}\sum_i \left(\sum_{j\in g(i)\setminus\{i\}} f_{it} - f_{jt} + u_{it} - u_{jt}\right)(g_i + v_i) \\
&= \frac{1}{nk}\sum_i \sum_{j\in g(i)\setminus\{i\}} (f_{it} - f_{jt})g_i + (f_{it} - f_{jt})v_i + (u_{it} - u_{jt})g_i + (u_{it} - u_{jt})v_i \\
&\equiv H_n + J_n + K_n + L_n.
\end{aligned}
$$

First consider H_n. By Assumption 3.1, $\psi \to g(\psi)$ is continuous and $\text{Supp}(\psi) \subseteq \bar{B}(0, K)$ compact, so $\sup_{\psi\in\bar{B}(0,K)} |g(\psi)| \equiv K' < \infty$ and $|g_i| \leq K'$ a.s. Then we have

$$
|H_n| \lesssim n^{-1}\sum_i \sum_{j\in g(i)\setminus\{i\}} |\psi_i - \psi_j|_2 |g_i| \lesssim n^{-1}\sum_g \sum_{i,j\in g} |\psi_i - \psi_j|_2 = o_p(1).
$$

For the final equality, note that here we have the unsquared norm, different from Assumption 2.1. Proposition 8.6 of Cytrynbaum (2022) showed that this quantity is also $o_p(1)$. By substituting z_i for g_i, which satisfies the same conditions, this also shows that $E_n[z_i\check{h}_i'] = o_p(1)$. The proof that $J_n, K_n = O_p(n^{-1/2})$ are similar to the terms B_n, C_n above. Next, consider L_n. We have

$$
\begin{aligned}
E[L_n|\psi_{1:n}, \pi_n] &= \frac{1}{nk}\sum_i \sum_{j\in g(i)\setminus\{i\}} E[(u_{it} - u_{jt})v_i|\psi_{1:n}, \pi_n] \\
&= \frac{1}{nk}\sum_i \sum_{j\in g(i)\setminus\{i\}} E[u_{it}v_i|\psi_i] = \frac{k-1}{k} E_n[E[u_{it}v_i|\psi_i]] \\
&= \frac{k-1}{k} E[\text{Cov}(h_{it}, Y_i(1)|\psi_i)] + O_p(n^{-1/2}).
\end{aligned}
$$

The second equality follows since $j \neq i$ and by Lemma 9.21 of Cytrynbaum (2022). The third equality by counting. For the last equality, note that by Jensen, tower law, Young's inequality $E[E[u_{it}v_i|\psi_i]^2] \leq E[u_{it}^2 v_i^2] \leq (1/2)(E[u_{it}^4] + E[v_i^4])$. We showed $E[u_{it}^4] < \infty$ above and a similar proof applies to v_i. Then the final equality above follows by Chebyshev. The proof that $\text{Var}(L_n|\psi_{1:n}, \pi_n) = O_p(n^{-1/2})$ is similar to our analysis of D_n above. Then we have shown $E_n[D_i\check{h}_i Y_i] = p\frac{k-1}{k} E[\text{Cov}(h, Y(1)|\psi)] + o_p(1)$. The conclusion for $E_n[(1 - D_i)\check{h}_i Y_i]$ follows by symmetry. This finishes the proof. □