Covariate Adjustment and Post-stratification for Treatment Effect Estimation

Jorn Dammann

April 3, 2024

Abstract

This paper studies covariate-adjusted estimation methods for potential outcomes in randomized experiments. We consider a general framework with multiple treatment arms and consider the cases of both covariate-dependent and fixed treatment propensities. Though linear covariate adjustment is a known technique for reducing estimator variance, its interaction with post-stratification is less well-explored. We derive asymptotically optimal linear covariate adjustments for three types of estimators: (i) inverse propensity weighted estimator, (ii) difference-in-means estimator, and (iii) post-stratified difference-in-means estimator. Finally, we provide asymptotically valid confidence intervals for each optimally adjusted estimator. The results illustrate the utility of combining post-stratification and covariate adjustment in randomized experiments for improved inference on potential outcomes and, in turn, arbitrary contrasts of these potential outcomes (e.g., average treatment effects).

I thank my advisor, Max Cytrynbaum, for exceptionally generous guidance. Any errors are my own.

1 Introduction

In this paper, we study covariate-adjusted estimation methods for the population means of potential outcomes in randomized experiments. We consider a general framework with multiple treatment arms and consider both covariate-dependent and fixed treatment propensity cases. Considering multiple treatment arms allows for an analysis of optimal inference on arbitrary linear contrasts of expected potential outcomes. However, we focus primarily on optimal inference for individual expected potential outcomes, as taking linear contrasts of the resulting estimators provides estimators for the contrasts of corresponding expected potential outcomes. Throughout the paper, we examine three classes of estimators, for which we each derive asymptotically optimal linear covariate adjustment. First, we consider the case of treatment propensities that are a function of the underlying covariates. This assumption of varying treatment propensities is well-motivated theoretically. For example, Hahn et al. (2011) considered multiple stage designs, where, in the second stage, assignment to different treatments is randomized on observed characteristics of individuals (i.e., covariates). The idea is that results from first-stage data can suggest how to best alter conditional assignment probabilities for the second wave of the experiment. We consider an inverse propensity weighted (IPW) estimator for this case of varying treatment propensities and derive a linear covariate adjustment that minimizes the asymptotic variance of our estimator.

Next, we consider the case of fixed treatment propensities. Hirano et al. (2003) showed that using a nonparametric estimate of the treatment propensity is more efficient than using the true known treatment propensity. Motivated by this result, we consider a difference-in-means estimator that incorporates a particular nonparametric propensity score. Rather than using an IPW estimator based on known treatment propensities, unadjusted difference-in-means estimates the population mean of a chosen potential outcome using the sample average among units assigned to that treatment arm. As for the IPW case, we also derive an optimal linear covariate adjustment coefficient and the corresponding asymptotic variance.

The third leg of our results considers estimation with post-stratification. Post-stratification is a technique by which the data are grouped into a discrete set of strata, often based on one or two particularly predictive underlying covariates. For example, Miratrix et al. (2013) gave the example of a medical trial experiment in which the efficacy of a cancer drug is being tested. In their example, stratifying based on the cancer stage could improve estimation efficiency by "conditioning away" variance explained by the cancer stage. Our post-stratified estimator extends the difference-in-means estimator in the sense that the difference-in-means estimator corresponds to the post-stratified estimator with only a single stratum. Here, we pay particular attention to the interaction between post-stratification and the optimal covariate adjustment coefficient.

Fourth, we provide asymptotically valid confidence intervals for each optimally adjusted estimator. The results illustrate the utility of combining post-stratification and covariate adjustment in randomized experiments for improved inference on potential outcomes and, in turn, arbitrary contrasts of these potential outcomes (e.g., average treatment effects).

2 Framework and Setup

We consider a randomized experiment with k treatment arms and n experimental units. Each $i \in \{1, \ldots, n\}$ is associated with a k-vector of potential outcomes $Y_i = (Y_{1i}, Y_{2i}, \ldots, Y_{ki})' \in \mathbf{R}^k$ and a p-vector of pre-treatment covariates $X_i = (X_{1i}, X_{2i}, \ldots, X_{pi})' \in \mathbf{R}^p$. We assume

$$(Y_i, X_i) \overset{iid}{\sim} F$$

where F is unknown. Once the data is drawn from the population and the covariate vectors observed, a k-vector of assignment indicators $\mathbb{R}_i = (\mathbb{R}_{1i}, \mathbb{R}_{2i}, \ldots, \mathbb{R}_{ki})'$ is independently assigned to each unit i according to known propensities $\pi(X_i) = (\pi_{1i}, \ldots, \pi_{ki})$. Each unit is assigned to exactly one treatment arm. That is, $\sum_{j=1}^{k} \mathbb{R}_{ji} = 1$ with probability one. Note that this setup implies that $(Y_i, X_i, \mathbb{R}_i)$ are iid. We maintain the following set of assumptions throughout the paper:

Assumption 2.1 (Setup Assumptions)**.**

1. The random assignment indicator vector is conditionally independent of the potential outcomes given the p-vector of pre-treatment covariates.

2. $\pi_{ai} = E[\mathbb{R}_{ai}|X_i] > 0$ for all realizations of X_i and $a = 1, \ldots, k$.

3. Each unit in the sample receives precisely one of the k treatments. Note that if there are $k-1$ different treatments being tested, one of the arms corresponds to no treatment.

Assumption 2.2 (Moment Conditions)**.**

1. For each $a = 1, \ldots, k$, $\mathrm{E}[Y_{ai}^4] < \infty$.

2. $\mathrm{E}[\|X_i\|^4] < \infty$

We are interested in estimating the expected potential outcome for each treatment arm. That is, for each $a \in \{1, \ldots, k\}$, we are interested in estimating $\mathrm{E}[Y_{ai}]$.

Remark 2.3 (Connection to Average Treatment Effects)**.** We can extend the resulting tools to estimate quantities of the form $\mathrm{E}[c'Y_i]$, where $c \in \mathbf{R}^k$ is a contrast vector of choice. For example, if $k = 2$ and we are interested in estimating $\mathrm{E}[Y_{1i} - Y_{0i}]$, we would choose $c = (1, -1)'$.

2.1 Additional Notation

Beyond the notation introduced above, we will use some additional notation. We let $\mathrm{E}_n[\cdot]$ denote sample averages and $\mathrm{E}_n[\cdot|\cdot]$ denote conditional sample averages. For instance, $\mathrm{E}_n[Y_i] = n^{-1}\sum_{i=1}^n Y_i$, and $\mathrm{E}_n[Y_i|X_i] = f(X_i)$ where $f(x) = \frac{\sum_{i=1}^n Y_i \mathbb{1}\{X_i=x\}}{\sum_{i=1}^n \mathbb{1}\{X_i=x\}}$. LIE refers to the Law of Iterated Expectations and LOTP to the Law of Total Probability.

3 Main Results

3.1 Optimal Covariate Adjustment for IPW Estimator

The first setting we consider is one in which treatment propensities are allowed to vary as functions of covariates. This setting is not foreign to empirical practice, particularly in two-stage designs. For example, Karlan & Wood (2017) estimate how donors respond to mailed reports of charities effectiveness using a two-stage design. In their second wave, donors' probability of treatment assignment is a function of stratification variables related to previous donation history and experimental status in the first wave, chosen to maximize statistical power.

Assumption 3.1 (Varying Treatment Propensities)**.** As outlined in the setup, the probability of assignment to a given treatment arm is a function of the covariate vector X_i for each unit i. That is, we have that $\Pr(\mathbb{R}_{ai} = 1|X_i) = \pi(X_i)_a := \pi_{ai}$, where a is a chosen treatment arm.

We consider inverse propensited weighted (IPW) estimators. The simple, unadjusted IPW estimator has the form

$$\widehat{\mu}_{an}^{\text{simple}} = \mathrm{E}_n\left[\frac{Y_{ai}\mathbb{R}_{ai}}{\pi_{ai}}\right]$$

where we recall that $Y_{ai}\mathbb{R}_{ai}$ takes value Y_{ai} if unit i is assigned to treatment arm a and 0 otherwise. The inverse weighting by treatment propensities, i.e. the $\frac{1}{\pi_{ai}}$ scaling, ensures both consistency and unbiasedness. Recall, importantly, that π_{ai} is defined as a function of X_i. We can improve on $\widehat{\mu}_{an}^{\text{simple}}$ by modifying our estimator via a linear covariate adjustment that reduces asymptotic variance by incorporating into the estimator variation in potential outcomes predictable by the baseline covariates. In particular, we consider the following family of estimators:

Definition 3.2 (Family of Linearly Adjusted IPW Estimators)**.** We consider the family of linearly adjusted IPW estimators $\mathcal{F}_{\text{IPW}}$ given by

$$\mathcal{F}_{\text{IPW}} = \left\{\widehat{\mu}_{an} = \mathrm{E}_n\left[\frac{Y_{ai}\mathbb{R}_{ai}}{\pi_{ai}}\right] - \widehat{\gamma}_{an}'\left(\mathrm{E}_n\left[\frac{X_i(\mathbb{R}_{ai}-\pi_{ai})}{\pi_{ai}}\right]\right) : \widehat{\gamma}_{an} \xrightarrow{p} \gamma \text{ and } \gamma \in \mathbf{R}^p\right\}$$

where the first term of each estimator in the family is the unadjusted IPW estimator and the second term is the covariate adjustment term.

Next, we characterize the asymptotic distribution of each estimator in $\mathcal{F}_{\text{IPW}}$.

Theorem 3.3 (Asymptotic Distribution of the Adjusted IPW Estimator)**.** *Under Assumption 3.2, we have for each $\widehat{\mu}_{an} \in \mathcal{F}_{IPW}$ that*

$$\sqrt{n}\left(\widehat{\mu}_{an} - \mathrm{E}\left[Y_{ai}\right]\right) \xrightarrow{d} \mathcal{N}\left(0, V(\gamma_a)\right)$$

where

$$V(\gamma_a) = \mathrm{Var}\left(Y_{ai}\right) + \mathrm{E}\left[\left(\frac{1-\pi_{ai}}{\pi_{ai}}\right)\left(Y_{ai} - \gamma_a' X_i\right)^2\right].$$

The theorem above establishes the asymptotic distribution of an arbitrary estimator in the family of estimators defined above. It is useful to examine the form of the asymptotic variance, noting in particular that it takes the form of the variance of the potential outcome plus a squared expectation. In the case of fixed treatment assignments, we can pull that propensity factor $\frac{1-\pi_{ai}}{\pi_{ai}}$ out of the expectation. We will observe as we progress to the difference-in-means section that squared expectation term improves to a variance term. Note also that the choice of γ_a matters: a strong choice of γ_a will explain much of the variation in Y_{ai} and thus reduce the expectation term, whereas a very poor choice of γ would increase the asymptotic variance. Of course, we are particularly interested in finding the estimator which yields the lowest asymptotic variance. This motivates the following definition:

Definition 3.4 (Optimally Adjusted Estimator)**.** An estimator $\widehat{\mu}^*_{an} \in \mathcal{F}_{\text{IPW}}$ is optimally adjusted if it has the form

$$\widehat{\mu}^*_{an} = \mathrm{E}_n\left[\frac{Y_{ai}\mathbb{R}_{ai}}{\pi_{ai}}\right] - \widehat{\gamma}^{*\prime}_{an}\left(\mathrm{E}_n\left[\frac{X_i\left(\mathbb{R}_{ai} - \pi_{ai}\right)}{\pi_{ai}}\right]\right)$$

where

$$\widehat{\gamma}^*_{an} \xrightarrow{p} \gamma^*_a \in \arg\min_{\gamma_a \in \mathbf{R}^p} V(\gamma_a).$$

That is, an optimally adjusted estimator utilizes a constistent estimator for a covariate adjustment coefficient that minimizes the asymptotic variance of the estimator. Since we have found the form of the asymptotic variance in Theorem 3.3, we can solve for an optimally adjusted estimator by taking first order conditions.

Assumption 3.5. The quantity $\mathrm{E}\left[\left(\frac{1-\pi_{ai}}{\pi_{ai}}\right)X_i X_i'\right]$ is invertible.

Theorem 3.6 (Optimal Covariate Adjustment)**.** *Under Assumption 3.6, an optimally adjusted estimator under the above framework is*

$$\widehat{\mu}^*_{an} = \mathrm{E}_n\left[\frac{Y_{ai}\mathbb{R}_{ai}}{\pi_{ai}}\right] - \widehat{\gamma}^{*\prime}_{an}\left(\mathrm{E}_n\left[\frac{X_i\left(\mathbb{R}_{ai}-\pi_{ai}\right)}{\pi_{ai}}\right]\right)$$

where

$$\widehat{\gamma}^*_{an} = \left(\mathrm{E}_n\left[\left(\frac{1-\pi_{ai}}{\pi_{ai}}\right)X_iX_i'\right]\right)^{-1}\mathrm{E}_n\left[\left(\frac{1-\pi_{ai}}{\pi_{ai}}\right)\frac{Y_{ai}\mathbb{R}_{ai}}{\pi_{ai}}X_i\right]$$

and

$$\widehat{\gamma}^*_{an} \xrightarrow{p} \gamma^*_a = \left(\mathrm{E}\left[\left(\frac{1-\pi_{ai}}{\pi_{ai}}\right)X_iX_i'\right]\right)^{-1}\mathrm{E}\left[\left(\frac{1-\pi_{ai}}{\pi_{ai}}\right)Y_{ai}X_i\right]$$

as $n \to \infty$*.*

In the above theorems we have established a covariate-adjusted IPW estimator for $\mathrm{E}\left[Y_{ai}\right]$ that achieves the lowest asymptotic variance among estimators in the family $\mathcal{F}_{\mathrm{IPW}}$ defined above.

Remark 3.7 (Optimal Covariate Adjustment for Arbitrary Contrasts)**.** What if, instead of estimating $\mathrm{E}\left[Y_{ai}\right]$, we are interested in estimating $\mathrm{E}\left[c'Y_{ai}\right]$ for some $c \in \mathbf{R}^k$? A natural estimator to try given the above results would be $c'\widehat{\mu}^*_{an}$. However, given our setup with treatment propensities as functions of the covariates, the optimal covariate adjustment may depend on the choice of contrast vector. That is, we should be able to do even better than simply using the optimal covariate adjustment for each arm. Though we do not solve for an explicit optimal covariate adjustment, as it is generally not unique, we find a single equation that the optimal covariate adjustment must satisfy. See Appendix B.

3.2 Optimal Covariate Adjustment for Difference-in-Means

In the previous section, we operated in a more general setting of treatment propensities that vary by X_i. We now transition to a setting with constant treatment propensities.

Assumption 3.8 (Constant Treatment Propensities)**.** The treatment propensities are no longer a function of X_i. That is, $(\pi_1, \ldots, \pi_k)$ is constant.

Though it may seem intuitive to continue using an IPW estimator as defined in the previous section, we find that a difference-in-means estimator that utilizes a non-parametric estimate of treatment propensities rather than the known treatment propensities actually achieves a weakly smaller asymptotic variance for the same choice of covariate adjustment coefficient. Our simple, unadjusted difference-in-means estimator for $\mathrm{E}\left[Y_{ai}\right]$ has the form

$$\widehat{\mu}^{\text{simple}}_{an} = \mathrm{E}_n\left[\frac{Y_{ai}\mathbb{R}_{ai}}{\mathrm{E}_n\left[\mathbb{R}_{ai}\right]}\right].$$

Note that this is identical to the estimator from the previous section except that $\pi_{ai}(X_i)$ has been replaced by $\mathrm{E}_n\left[\mathbb{R}_{ai}\right]$. Though it may seem counterintuitive, we will show that replacing π_a with the estimator $\mathrm{E}_n\left[\mathbb{R}_{ai}\right]$ reduces asymptotic variance. As in the previous section, we consider linear covariate adjustment.

Definition 3.9 (Family of Linearly Adjusted Difference-in-Means Estimators)**.** We consider the family of linearly adjusted difference-in-means estimators $\mathcal{F}_{\mathrm{DIM}}$ given by

$$\mathcal{F}_{\mathrm{DIM}} = \left\{ \widehat{\mu}_{an} = \mathrm{E}_n\left[\frac{Y_{ai}\mathbb{R}_{ai}}{\mathrm{E}_n\left[\mathbb{R}_{ai}\right]}\right] - \widehat{\gamma}_{an}'\left(\mathrm{E}_n\left[\frac{X_i\left(\mathbb{R}_{ai} - \mathrm{E}_n\left[\mathbb{R}_{ai}\right]\right)}{\mathrm{E}_n\left[\mathbb{R}_{ai}\right]}\right]\right) : \widehat{\gamma}_{an} \xrightarrow{p} \gamma \text{ and } \gamma \in \mathbf{R}^p \right\}$$

where the first term of each estimator in the family is the unadjusted difference-in-means estimator and the second term is the covariate adjustment term.

Note that the covariate adjustment coefficient $\widehat{\gamma}_{an}$ is an estimated quantity. However, as we will see in the following theorem, the asymptotic variance of our estimator depends only on the probability limit of $\widehat{\gamma}_{an}$, denoted γ_a. This fact implies that when we find an optimal adjustment coefficient that is unknown and needs to be estimated, an estimator based on a consistent estimator of this optimal adjustment coefficient enjoys the same asymptotic variance as the infeasible estimator that utilizes the true optimal adjustment coefficient.

Theorem 3.10 (Asymptotic Distribution of the Adjusted Difference-in-Means Estimator)**.** *Under Assumptions 2.1, 2.2, and 3.8, we have for each* $\widehat{\mu}_{an} \in \mathcal{F}_{DIM}$ *that*

$$\sqrt{n}\left(\widehat{\mu}_{an} - \mathrm{E}\left[Y_{ai}\right]\right) \xrightarrow{d} \mathcal{N}\left(0, V\left(\gamma_a\right)\right)$$

where

$$V(\gamma_a) = \mathrm{Var}\left(Y_{ai}\right) + \left(\frac{1-\pi_a}{\pi_a}\right)\mathrm{Var}\left(Y_{ai} - \gamma_a' X_i\right).$$

Remark 3.11 (Difference-in-Means achieves lower variance than IPW)**.** Let's compare the distribution of the IPW estimator from Theorem 3.3 with that of the difference-in-means estimator from the above theorem under the assumption of constant treatment propensities. Note that they are identical, except that in the former, the second term is $\mathrm{E}\left[\widetilde{Y}_{ai}^2\right]$, whereas in the latter, it is $\mathrm{Var}\left(\widetilde{Y}_{ai}\right)$, where $\widetilde{Y}_{ai} = Y_{ai} - \gamma_a' X_i$. Since $\mathrm{Var}\left(\widetilde{Y}_{ai}\right) = \mathrm{E}\left[\widetilde{Y}_{ai}^2\right] - \left(\mathrm{E}\left[\widetilde{Y}_{ai}\right]\right)^2$, this confirms that using an estimate of the treatment propensity is in fact superior to using the true treatment propensity in this setting.

We proceed by defining and deriving the optimally adjusted difference-in-means estimator.

Definition 3.12 (Optimally Adjusted Estimator)**.** An estimator $\widehat{\mu}_{an}^* \in \mathcal{F}_{\mathrm{DIM}}$ is optimally adjusted if it has the form

$$\widehat{\mu}_{an}^* = \mathrm{E}_n\left[\frac{Y_{ai}\mathbb{R}_{ai}}{\mathrm{E}_n\left[\mathbb{R}_{ai}\right]}\right] - \widehat{\gamma}_{an}^{*\prime}\left(\mathrm{E}_n\left[\frac{X_i\left(\mathbb{R}_{ai} - \mathrm{E}_n\left[\mathbb{R}_{ai}\right]\right)}{\mathrm{E}_n\left[\mathbb{R}_{ai}\right]}\right]\right)$$

where

$$\widehat{\gamma}_{an}^{*} \xrightarrow{p} \gamma_a^* \in \arg\min_{\gamma_a \in \mathbf{R}^p} V(\gamma_a).$$

Assumption 3.13. The quantity $\mathrm{Var}(X_i)$ is invertible.

Theorem 3.14 (Optimal Covariate Adjustment)**.** *An optimally adjusted estimator under the above framework is*

$$\widehat{\mu}_{an}^{*} = \mathrm{E}_n\left[\frac{Y_{ai}\mathbb{R}_{ai}}{\mathrm{E}_n[\mathbb{R}_{ai}]}\right] - \widehat{\gamma}_{an}^{*\prime}\left(\mathrm{E}_n\left[\frac{X_i(\mathbb{R}_{ai} - \mathrm{E}_n[\mathbb{R}_{ai}])}{\mathrm{E}_n[\mathbb{R}_{ai}]}\right]\right)$$

where

$$\widehat{\gamma}_{an}^{*} = \mathrm{E}_n\left[(X_i - \mathrm{E}_n[X_i])(X_i - \mathrm{E}_n[X_i])'\right]^{-1}\left(\frac{\mathrm{E}_n[X_i Y_{ai}\mathbb{R}_{ai}]}{\mathrm{E}_n[\mathbb{R}_{ai}]} - \mathrm{E}_n[X_i]\frac{\mathrm{E}_n[Y_{ai}\mathbb{R}_{ai}]}{\mathrm{E}_n[\mathbb{R}_{ai}]}\right)$$

and

$$\widehat{\gamma}_{an}^{*} \xrightarrow{p} \gamma_a^* = \mathrm{Var}(X_i)^{-1}\,\mathrm{Cov}(X_i, Y_{ai})$$

as $n \to \infty$*.*

It is worth comparing the optimal adjustment coefficient above to that of the IPW estimator under the assumption of constant treatment propensities. In particular, we have as a direct consequence of Theorem 3.6 that with constant treatment propensities, the optimal covariate adjustment coefficient for IPW is

$$\gamma_a^{*,\,\mathrm{IPW}} = (\mathrm{E}[X_i X_i'])^{-1}\,\mathrm{E}[Y_{ai}X_i]$$

whereas in Theorem 3.14 above we found the optimal covariate adjustment for difference-in-means

$$\begin{aligned}\gamma_a^{*,\,\mathrm{DIM}} &= \mathrm{Var}(X_i)^{-1}\,\mathrm{Cov}(X_i, Y_{ai}) \\ &= (\mathrm{E}[X_i X_i'] - \mathrm{E}[X_i]\,\mathrm{E}[X_i'])^{-1}(\mathrm{E}[X_i Y_{ai}] - \mathrm{E}[X_i]\,\mathrm{E}[Y_{ai}])\end{aligned}$$

Thus, $\gamma_a^{*,\,\mathrm{IPW}}$ corresponds precisely to $\gamma_a^{*,\,\mathrm{DIM}}$ when the covariates X_i have mean zero.

3.3 Optimal Covariate Adjustment for Post-Stratification

Maintaining the constant treatment propensity assumption from the previous section, we take difference-in-means further by considering post-stratification. Post-stratification is a form of adjustment: we stratify the experimental units using a variable S_i that is a function of the information available prior to treatment (such as covariates), estimate the expected potential outcome within each stratum, and then take a weighted average of these in-stratum estimates to get the overall estimator. The benefits of post-stratification are substantial. Miratrix et al. (2013) show that post-stratification can substantially improve upon the simple difference-in-means estimator under various randomization schemes if the stratification variable is well-chosen, including the Bernoulli scheme considered in this paper. Moreover, they find that post-stratification is asymptotically equivalent to stratifying units prior to treatment into pre-defined blocks and then randomizing within blocks, a technique known as blocking. The ability of post-stratification to improve estimator efficiency has prompted work on applying the technique to various types of estimators. For instance, Pashley et al. (2023) apply post-stratification to improve the bias, variance, and standard error estimates of instrumental variables estimators. Another direction of research examines how to choose the post-stratification rule optimally: Kim et al. (2023) propose a predictive regression model-based method to determine mappings from the covariate space to the stratification variable.

We will focus on the interaction between post-stratification and optimal linear covariate adjustment. To implement post-stratification, we introduce a stratification variable S_i with discrete, finite support. Our simple, unadjusted post-stratified estimator for $\mathrm{E}\left[Y_{ai}\right]$ has the form

$$\widehat{\mu}_{an}^{\text{simple}} = \mathrm{E}_n\left[\frac{Y_{ai}\mathbb{R}_{ai}}{\mathrm{E}_n\left[\mathbb{R}_{ai}|S_i\right]}\right]$$

which is identical to the simple difference-in-means estimator from the previous estimator from the previous section, except that we utilize a conditional treatment propensity $\mathrm{E}_n\left[\mathbb{R}_{ai}|S_i\right]$ in the place of $\mathrm{E}_n\left[\mathbb{R}_{ai}\right]$. Thus, instead of re-estimating a general propensity score, we re-estimate treatment propensities for each stratum.

Assumption 3.15 (Finite Strata)**.** We assume that there are a finite number of strata $\{1, \ldots, k\}$. This denotes the support of S_i.

Definition 3.16 (Family of Linearly Adjusted Post-stratified Estimators)**.** We consider the family of linearly adjusted post-stratified estimators $\mathcal{F}_{\text{PS}}$ given by

$$\mathcal{F}_{\text{PS}} = \left\{\widehat{\mu}_{an} = \mathrm{E}_n\left[\frac{Y_{ai}\mathbb{R}_{ai}}{\mathrm{E}_n\left[\mathbb{R}_{ai}|S_i\right]}\right] - \widehat{\gamma}_{an}'\left(\mathrm{E}_n\left[\frac{X_i\left(\mathbb{R}_{ai} - \mathrm{E}_n\left[\mathbb{R}_{ai}|S_i\right]\right)}{\mathrm{E}_n\left[\mathbb{R}_{ai}|S_i\right]}\right]\right) : \widehat{\gamma}_{an} \xrightarrow{p} \gamma \text{ and } \gamma \in \mathbf{R}^p\right\}$$

where the first term of each estimator in the family is the unadjusted post-stratified estimator and the second term is the covariate adjustment term.

Theorem 3.17 (Asymptotic Distribution of the Adjusted Post-Stratified Estimator)**.** *Under Assumptions 2.1, 2.2, 3.8, and 3.15, we have for each $\widehat{\mu}_{an} \in \mathcal{F}_{PS}$ that*

$$\sqrt{n}\left(\widehat{\mu}_{an} - \mathrm{E}\left[Y_{ai}\right]\right) \xrightarrow{d} \mathcal{N}\left(0, V(\gamma_a)\right)$$

as $n \to \infty$, where

$$V(\gamma_a) = \mathrm{Var}(Y_{ai}) + \left(\frac{1-\pi_a}{\pi_a}\right) \mathrm{E}\left[\mathrm{Var}\left(Y_{ai} - \gamma_a' X_i | S_i\right)\right].$$

Remark 3.18 (Post-Stratification Weakly Improves on Difference-in-Means Asymptotically)**.** We should take a moment to analyze how post-stratification improves upon the asymptotic variance of the generic difference-in-means estimator from the previous section. Comparing Theorem 3.17 to 3.12, we see that the two asymptotic variances are identical except that $\mathrm{Var}(Y_{ai} - \gamma_a' X_i)$ has been replaced by the conditional expectation of this variance given S_i. Since

$$\begin{aligned} \mathrm{E}\left[\mathrm{Var}\left(Y_{ai} - \gamma_a' X_i | S_i\right)\right] &= \mathrm{Var}\left(Y_{ai} - \gamma_a' X_i\right) - \mathrm{Var}\left(\mathrm{E}\left[Y_{ai} - \gamma_a' X_i | S_i\right]\right) \quad \text{by the LOTV} \\ &\leq \mathrm{Var}\left(Y_{ai} - \gamma_a' X_i\right), \end{aligned}$$

Theorem 3.17 implies that post-stratification yields a weakly smaller asymptotic variance than regular difference-in-means.

Of course, as Miratrix et al. (2013) pointed out, post-stratification can adversely affect estimator variance in finite samples if the number of strata is too large and the stratification variable is poorly chosen.

Definition 3.19 (Optimally Adjusted Estimator)**.** An estimator $\widehat{\mu}_{an}^* \in \mathcal{F}_{\mathrm{PS}}$ is optimally adjusted if it has the form

$$\widehat{\mu}_{an}^* = \mathrm{E}_n\left[\frac{Y_{ai}\mathbb{R}_{ai}}{\mathrm{E}_n\left[\mathbb{R}_{ai}|S_i\right]}\right] - \widehat{\gamma}_{an}^{*\prime}\left(\mathrm{E}_n\left[\frac{X_i\left(\mathbb{R}_{ai} - \mathrm{E}_n\left[\mathbb{R}_{ai}|S_i\right]\right)}{\mathrm{E}_n\left[\mathbb{R}_{ai}|S_i\right]}\right]\right)$$

where

$$\widehat{\gamma}_{an}^* \xrightarrow{p} \gamma_a^* \in \arg\min_{\gamma_a \in \mathbf{R}^p} V(\gamma_a).$$

Assumption 3.20. The quantity $\mathrm{E}\left[\mathrm{Var}\left(X_i|S_i\right)\right]$ is invertible.

Theorem 3.21 (Optimal Covariate Adjustment)**.** *An optimally adjusted estimator under the above framework is*

$$\widehat{\mu}_{an}^* = \mathrm{E}_n\left[\frac{Y_{ai}\mathbb{R}_{ai}}{\mathrm{E}_n\left[\mathbb{R}_{ai}|S_i\right]}\right] - \widehat{\gamma}_{an}^{*\prime}\left(\mathrm{E}_n\left[\frac{X_i\left(\mathbb{R}_{ai} - \mathrm{E}_n\left[\mathbb{R}_{ai}|S_i\right]\right)}{\mathrm{E}_n\left[\mathbb{R}_{ai}|S_i\right]}\right]\right)$$

where

$$\widehat{\gamma}_{an}^* = \left(\mathrm{E}_n\left[X_iX_i'\right] - \mathrm{E}_n\left[\mathrm{E}_n\left[X_i|S_i\right]\mathrm{E}_n\left[X_i'|S_i\right]\right]\right)^{-1}\mathrm{E}_n\left[\frac{\mathrm{E}_n\left[X_iY_{ai}\mathbb{R}_{ai}|S_i\right]}{\mathrm{E}_n\left[\mathbb{R}_{ai}|S_i\right]} - \mathrm{E}_n\left[X_i|S_i\right]\frac{\mathrm{E}_n\left[Y_{ai}\mathbb{R}_{ai}|S_i\right]}{\mathrm{E}_n\left[\mathbb{R}_{ai}|S_i\right]}\right]$$

and

$$\widehat{\gamma}_{an}^* \xrightarrow{p} \gamma_a^* = \left(\mathrm{E}\left[\mathrm{Var}\left(X_i|S_i\right)\right]\right)^{-1} \mathrm{E}\left[\mathrm{Cov}\left(X_i, Y_{ai}|S_i\right)\right]$$

as $n \to \infty$*.*

Remark 3.22. Note that the optimal covariate adjustment coefficient is identical to that from the generic difference-in-means section but with variances replaced by expectations of conditional variances and covariances replaced by expectations of conditional covariances.

Intuitively, post-stratification is a powerful tool for reducing the estimator variance because it eliminates fluctuations in the potential outcome of interest predictable by S. In particular, building on Remark 3.21, we observe that the maximum improvement is achieved asymptotically relative to the generic difference-in-means estimator when $\mathrm{Var}\left(\mathrm{E}\left[Y_{ai} - \gamma_a' X_i | S_i\right]\right)$ is maximized. Thus, intuitively, S_i should be chosen so as to maximize the heterogeneity of $\mathrm{E}\left[Y_{ai} - \gamma_a' X_i | S_i\right]$ across strata. That is $\mathrm{E}\left[Y_{ai} - \gamma_a' X_i | S_i = s\right]$ should vary greatly for different values of s, where $s \in \{1, \ldots, k\}$.

4 Inference

Finally, we derive asymptotically valid confidence intervals for $E\left[Y_{ai}\right]$ using each of the three estimators considered. In so doing, we harness the efficiency gains from covariate adjustment and post-stratification for superior inference.

Theorem 4.1 (Inference for IPW)**.** *Let* $\widehat{\mu}_{an}$ *be the covariate-adjusted IPW estimator from Section 3.1. Then an asymptotically valid* $1-\alpha$ *CI for* $\mathrm{E}\left[Y_{ai}\right]$ *is*

$$\left[\widehat{\mu}_{an} - z_{1-\alpha/2}\sqrt{\frac{\widehat{V}_n\left(\widehat{\gamma}_{an}\right)}{n}},\ \widehat{\mu}_{an} + z_{1-\alpha/2}\sqrt{\frac{\widehat{V}_n\left(\widehat{\gamma}_{an}\right)}{n}}\right].$$

where

$$\widehat{V}_n\left(\widehat{\gamma}_{an}\right) = \mathrm{E}_n\left[\frac{Y_{ai}^2 \mathbb{R}_{ai}}{\pi_{ai}}\right] - \left(\mathrm{E}_n\left[\frac{Y_{ai}\mathbb{R}_{ai}}{\pi_{ai}}\right]\right)^2 + \mathrm{E}_n\left[\left(\frac{1-\pi_{ai}}{\pi_{ai}}\right)\left(\frac{\mathbb{R}_{ai}Y_{ai}^2}{\pi_{ai}} - 2\widehat{\gamma}_{an}' X_i \frac{\mathbb{R}_{ai}Y_{ai}}{\pi_{ai}} + \left(\widehat{\gamma}_{an}' X_i\right)^2\right)\right].$$

Theorem 4.2 (Inference for Difference-in-Means)**.** *Let* $\widehat{\mu}_{an}$ *be the covariate-adjusted difference-in-means estimator from Section 3.2. Then an asymptotically valid* $1-\alpha$ *CI for* $\mathrm{E}\left[Y_{ai}\right]$ *is*

$$\left[\widehat{\mu}_{an} - z_{1-\alpha/2}\sqrt{\frac{\widehat{V}_n\left(\widehat{\gamma}_{an}\right)}{n}},\ \widehat{\mu}_{an} + z_{1-\alpha/2}\sqrt{\frac{\widehat{V}_n\left(\widehat{\gamma}_{an}\right)}{n}}\right]$$

where

$$\widehat{V}_n\left(\widehat{\gamma}_{an}\right) = A_n + \left(\frac{1-\pi_a}{\pi_a}\right)\left(A_n + B_n + 2C_n\right)$$

with

$$\begin{aligned}
A_n &= \mathrm{E}_n\left[\frac{\mathbb{R}_{ai}Y_{ai}^2}{\pi_a}\right] - \left(\mathrm{E}_n\left[\frac{\mathbb{R}_{ai}Y_{ai}}{\pi_a}\right]\right)^2 \\
B_n &= \widehat{\gamma}_{an}'\,\mathrm{E}_n\left[\left(X_i - \mathrm{E}_n\left[X_i\right]\right)\left(X_i - \mathrm{E}_n\left[X_i\right]\right)'\right]\widehat{\gamma}_{an} \\
C_n &= \widehat{\gamma}_{an}'\left(\mathrm{E}_n\left[X_i\frac{Y_{ai}\mathbb{R}_{ai}}{\pi_a}\right] - \mathrm{E}_n\left[X_i\right]\mathrm{E}_n\left[\frac{Y_{ai}\mathbb{R}_{ai}}{\pi}\right]\right).
\end{aligned}$$

Theorem 4.3 (Inference for Post-Stratified Estimator)**.** *Let $\widehat{\mu}_{an}$ be the post-stratified estimator from Section 3.3. Then an asymptotically valid $1-\alpha$ CI for $\mathrm{E}\left[Y_{ai}\right]$ is*

$$\left[\widehat{\mu}_{an} - z_{1-\alpha/2}\sqrt{\frac{\widehat{V}_n\left(\widehat{\gamma}_{an}\right)}{n}},\ \widehat{\mu}_{an} + z_{1-\alpha/2}\sqrt{\frac{\widehat{V}_n\left(\widehat{\gamma}_{an}\right)}{n}}\right]$$

where

$$\widehat{V}_n\left(\widehat{\gamma}_{an}\right) = A_n + \left(\frac{1-\pi_a}{\pi_a}\right)\left(B_n + C_n + 2D_n\right)$$

with

$$\begin{aligned}
A_n &= \mathrm{E}_n\left[\frac{\mathbb{R}_{ai}Y_{ai}^2}{\pi_a}\right] - \left(\mathrm{E}_n\left[\frac{\mathbb{R}_{ai}Y_{ai}}{\pi_a}\right]\right)^2 \\
B_n &= \mathrm{E}_n\left[\mathrm{E}_n\left[\left.\frac{Y_{ai}^2\mathbb{R}_{ai}}{\pi_a}\right|S_i\right] - \left(\mathrm{E}_n\left[\left.\frac{Y_{ai}\mathbb{R}_{ai}}{\pi_a}\right|S_i\right]\right)^2\right] \\
C_n &= \widehat{\gamma}_{an}'\,\mathrm{E}_n\left[\mathrm{E}_n\left[X_iX_i'|S_i\right] - \mathrm{E}_n\left[X_i|S_i\right]\mathrm{E}_n\left[X_i'|S_i\right]\right]\widehat{\gamma}_{an} \\
D_n &= \widehat{\gamma}_{an}\,\mathrm{E}_n\left[\mathrm{E}_n\left[\left.X_i\frac{Y_{ai}\mathbb{R}_{ai}}{\pi_a}\right|S_i\right] - \mathrm{E}_n\left[X_i|S_i\right]\mathrm{E}_n\left[\left.\frac{Y_{ai}\mathbb{R}_{ai}}{\pi_a}\right|S_i\right]\right].
\end{aligned}$$

Remark 4.4. The above inference theorems are formulated to work for a generic choice of linear adjustment coefficient so as to retain flexibility. To achieve the tightest confidence intervals, a researcher should utilize an optimal covariate adjustment coefficient $\widehat{\gamma}_{an} = \widehat{\gamma}_{an}^*$. These are derived in Theorems 3.6, 3.14, and 3.21, respectively.

5 Concluding Remarks

In this paper, we introduce three types of covariate-adjusted estimation for a chosen expected potential outcome of the form $\mathrm{E}\left[Y_{ai}\right]$: inverse propensity weighted estimation, difference-in-means estimation, and post-stratified estimation. We show that each estimator improves upon the former in a constant marginal treatment propensities framework by achieving a weakly smaller asymptotic variance. This result allows us to construct increasingly tighter confidence intervals for each family of estimators considered. Throughout, we derive asymptotically optimal linear covariate adjustments, which improve estimator efficiency relative to the baseline. The optimally adjusted post-stratified estimator combines the tools of post-stratification and linear covariate adjustment for maximal effect: post-stratifying on highly predictive categories while employing linear covariate adjustment with an adjustment coefficient that is predictive within strata, researchers can significantly reduce the standard errors of their estimates. Though we focus on estimating a single potential outcome throughout, the results can be utilized to estimate ATE-like quantities. For instance, in a binary treatment scenario where the researcher is interested in $\mathrm{E}\left[Y_{1i} - Y_{0i}\right]$, they can estimate this quantity by $\widehat{\theta}^*_{1n} - \widehat{\theta}^*_{0n}$, where $\widehat{\theta}^*_{an}$ is the optimally adjusted post-stratified estimator from Section 3.3.

References

Hahn, J., Hirano, K., & Karlan, D. (2011). Adaptive experimental design using the propensity score. *Journal of Business & Economic Statistics*, *29*(1), 96–108.

Hirano, K., Imbens, G. W., & Ridder, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica*, *71*(4), 1161–1189.

Karlan, D., & Wood, D. H. (2017). The effect of effectiveness: Donor response to aid effectiveness in a direct mail fundraising experiment. *Journal of Behavioral and Experimental Economics*, *66*, 1–8.

Kim, T., Wang, L., Lai, R., & Yoon, S. (2023). Efficient treatment effect estimation with out-of-bag post-stratification. *arXiv preprint arXiv:2307.08150*.

Miratrix, L. W., Sekhon, J. S., & Yu, B. (2013). Adjusting treatment effect estimates by post-stratification in randomized experiments. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *75*(2), 369–396.

Pashley, N. E., Keele, L., & Miratrix, L. W. (2023). Improving instrumental variable estimators with post-stratification. *arXiv preprint arXiv:2303.10016*.

A Appendix

A.1 Proofs for Section 3.1

Proof of Theorem 3.3. Note that we can rewrite $\widehat{\mu}_{an}$ as

$$\begin{aligned}
\widehat{\mu}_{an} &= \mathrm{E}_n\left[Y_{ai}\right] + \mathrm{E}_n\left[\frac{\left(Y_{ai} - \widehat{\gamma}'_{an} X_i\right)\left(\mathbb{R}_{ai} - \pi_{ai}\right)}{\pi_{ai}}\right] \\
&= \mathrm{E}_n\left[Y_{ai}\right] + \mathrm{E}_n\left[\frac{\left(Y_{ai} - \gamma'_{a} X_i\right)\left(\mathbb{R}_{ai} - \pi_{ai}\right)}{\pi_{ai}}\right] + \mathrm{E}_n\left[\frac{\left(\gamma'_a - \widehat{\gamma}'_{an}\right) X_i\left(\mathbb{R}_{ai} - \pi_{ai}\right)}{\pi_{ai}}\right].
\end{aligned}$$

We can first check that the third term is $o_p\left(1/\sqrt{n}\right)$. In particular,

$$\sqrt{n}\,\mathrm{E}_n\left[\frac{\left(\gamma_a'-\widehat{\gamma}_{an}'\right)X_i\left(\mathbb{R}_{ai}-\pi_{ai}\right)}{\pi_{ai}}\right]=\left(\gamma_a'-\widehat{\gamma}_{an}'\right)\sqrt{n}\,\mathrm{E}_n\left[\frac{X_i\left(\mathbb{R}_{ai}-\pi_{ai}\right)}{\pi_{ai}}\right].$$

Since $\left(\gamma_a'-\widehat{\gamma}_{an}'\right)\xrightarrow{p}0$ as $n\to\infty$ by assumption, and $\sqrt{n}\,\mathrm{E}_n\left[\frac{X_i(\mathbb{R}_{ai}-\pi_{ai})}{\pi_{ai}}\right]\xrightarrow{d}Z\sim\mathcal{N}\left(0,\mathrm{Var}\left(\frac{X_i(\mathbb{R}_{ai}-\pi_{ai})}{\pi_{ai}}\right)\right)$ as $n\to\infty$ under our moment conditions by the CLT, combining results gives

$$\sqrt{n}\,\mathrm{E}_n\left[\frac{\left(\gamma_a'-\widehat{\gamma}_{an}'\right)X_i\left(\mathbb{R}_{ai}-\pi_{ai}\right)}{\pi_{ai}}\right]\xrightarrow{p}0Z=0$$

as $n\to\infty$. Thus, we can focus our analysis on the first two terms. We have

$$\begin{aligned}\mathrm{E}\left[\frac{\left(Y_{ai}-\gamma_a'X_i\right)\left(\mathbb{R}_{ai}-\pi_{ai}\right)}{\pi_{ai}}\right]&=\mathrm{E}\left[E\left[\frac{\left(Y_{ai}-\gamma_a'X_i\right)\left(\mathbb{R}_{ai}-\pi_{ai}\right)}{\pi_{ai}}\middle|X_i\right]\right]\quad\text{by LIE}\\&=\mathrm{E}\left[\frac{1}{\pi_{ai}}\mathrm{E}\left[Y_{ai}-\gamma_a'X_i|X_i\right]\mathrm{E}\left[\mathbb{R}_{ai}-\pi_{ai}|X_i\right]\right]\quad\text{by conditional independence}\\&=0\quad\text{because }\mathrm{E}\left[\mathbb{R}_{ai}-\pi_{ai}|X_i\right]=0\end{aligned}$$

Therefore, since $(Y_{ai},\mathbb{R}_{ai},X_i)$ are iid, $E\left[Y_{ai}^2\right]<\infty$, and $\mathrm{E}\left[\left(\frac{(Y_{ai}-\gamma_a'X_i)(\mathbb{R}_{ai}-\pi_{ai})}{\pi_{ai}}\right)^2\right]<\infty$ by Assumption 3.2, we have by the CLT that

$$\sqrt{n}\left(\mathrm{E}_n\begin{bmatrix}Y_{ai}\\\frac{(Y_{ai}-\gamma_a'X_i)(\mathbb{R}_{ai}-\pi_{ai})}{\pi_{ai}}\end{bmatrix}-\mathrm{E}\begin{bmatrix}Y_{ai}\\0\end{bmatrix}\right)\xrightarrow{d}Z\sim\mathcal{N}\left(0,\Sigma\right)$$

as $n\to\infty$, where

$$\Sigma=\mathrm{Var}\begin{pmatrix}Y_{ai}\\\frac{(Y_{ai}-\gamma_a'X_i)(\mathbb{R}_{ai}-\pi_{ai})}{\pi_{ai}}\end{pmatrix}.$$

Note that Σ is diagonal because

$$\begin{aligned}&\mathrm{Cov}\left(Y_{ai},\frac{\left(Y_{ai}-\gamma_a'X_i\right)\left(\mathbb{R}_{ai}-\pi_{ai}\right)}{\pi_{ai}}\right)\\&=\mathrm{E}\left[Y_{ai}\right]\mathrm{E}\left[\frac{\left(Y_{ai}-\gamma_a'X_i\right)\left(\mathbb{R}_{ai}-\pi_{ai}\right)}{\pi_{ai}}\right]-\mathrm{E}\left[\frac{Y_{ai}\left(Y_{ai}-\gamma_a'X_i\right)\left(\mathbb{R}_{ai}-\pi_{ai}\right)}{\pi_{ai}}\right]\\&=0-\mathrm{E}\left[\frac{1}{\pi_{ai}}\mathrm{E}\left[Y_{ai}\left(Y_{ai}-\gamma_a'X_i\right)|X_i\right]\mathrm{E}\left[\mathbb{R}_{ai}-\pi_{ai}|X_i\right]\right]\quad\text{by LIE and conditional independence}\\&=0.\end{aligned}$$

It follows that

$$\sqrt{n}\left(\widehat{\mu}_{an}-\mathrm{E}\left[Y_{ai}\right]\right)\xrightarrow{d}a'Z\sim\mathcal{N}\left(a'0,a'\Sigma a\right)$$

as $n \to \infty$, where $a = (1,1)'$. Using the result above that Σ is diagonal, we have $a'\Sigma a = \text{Var}(Y_{ai}) + \text{Var}\left(\frac{(Y_{ai}-\gamma_a' X_i)(\mathbb{R}_{ai}-\pi_{ai})}{\pi_{ai}}\right)$. Let's simplify the variance term. We have

$$\begin{aligned}
&\text{Var}\left(\frac{(Y_{ai}-\gamma_a' X_i)(\mathbb{R}_{ai}-\pi_{ai})}{\pi_{ai}}\right) \\
&= \text{E}\left[\left(\frac{(Y_{ai}-\gamma_a' X_i)(\mathbb{R}_{ai}-\pi_{ai})}{\pi_{ai}}\right)^2\right] \\
&= \text{E}\left[\text{E}\left[\left(\frac{(Y_{ai}-\gamma_a' X_i)(\mathbb{R}_{ai}-\pi_{ai})}{\pi_{ai}}\right)^2 \middle| X_i\right]\right] \quad \text{by LIE} \\
&= \text{E}\left[\text{E}\left[(Y_{ai}-\gamma_a' X_i)^2 | X_i\right] \frac{1}{\pi_{ai}^2} \text{E}\left[(\mathbb{R}_{ai}-\pi_{ai})^2 | X_i\right]\right] \quad \text{by conditional independence} \\
&= \text{E}\left[\text{E}\left[(Y_{ai}-\gamma_a' X_i)^2 | X_i\right] \frac{\pi_{ai}(1-\pi_{ai})}{\pi_{ai}^2}\right] \quad \text{using the variance of a Bernoulli r.v.} \\
&= \text{E}\left[\text{E}\left[(Y_{ai}-\gamma_a' X_i)^2 \left(\frac{1-\pi_{ai}}{\pi_{ai}}\right) \middle| X_i\right]\right] \\
&= \text{E}\left[(Y_{ai}-\gamma_a' X_i)^2 \left(\frac{1-\pi_{ai}}{\pi_{ai}}\right)\right] \quad \text{by LIE}
\end{aligned}$$

Combining results, we have

$$\sqrt{n}\left(\widehat{\mu}_{an} - \text{E}[Y_{ai}]\right) \xrightarrow{d} \mathcal{N}\left(0, \text{Var}(Y_{ai}) + \text{E}\left[(Y_{ai}-\gamma_a' X_i)^2 \left(\frac{1-\pi_{ai}}{\pi_{ai}}\right)\right]\right)$$

as $n \to \infty$. □

Proof of Theorem 3.6. By Theorem 3.3, minimizing the asymptotic variance of $\widehat{\mu}_{an}$ is equivalent to solving

$$\arg\min_{\gamma_a} \text{E}\left[(Y_{ai}-\gamma_a' X_i)^2 \left(\frac{1-\pi_{ai}}{\pi_{ai}}\right)\right].$$

The gradient vector of the function being minimized is

$$\begin{aligned}
\nabla_{\gamma_a} \text{E}\left[(Y_{ai}-\gamma_a' X_i)^2 \left(\frac{1-\pi_{ai}}{\pi_{ai}}\right)\right] &= \text{E}\left[2(Y_{ai}-\gamma_a' X_i)(-X_i)\left(\frac{1-\pi_{ai}}{\pi_{ai}}\right)\right] \\
&= -2\,\text{E}\left[(Y_{ai}X_i - X_i X_i' \gamma_a)\left(\frac{1-\pi_{ai}}{\pi_{ai}}\right)\right]
\end{aligned}$$

so the FOC is

$$\begin{aligned}
&\text{E}\left[(Y_{ai}X_i - X_i X_i' \gamma_a^*)\left(\frac{1-\pi_{ai}}{\pi_{ai}}\right)\right] = 0_p \\
\Longrightarrow\ &\text{E}\left[\left(\frac{1-\pi_{ai}}{\pi_{ai}}\right) Y_{ai}X_i\right] = \text{E}\left[\left(\frac{1-\pi_{ai}}{\pi_{ai}}\right) X_i X_i'\right]\gamma_a^* \\
\Longrightarrow\ &\gamma_a^* = \text{E}\left[\left(\frac{1-\pi_{ai}}{\pi_{ai}}\right) X_i X_i'\right]^{-1} \text{E}\left[\left(\frac{1-\pi_{ai}}{\pi_{ai}}\right) Y_{ai}X_i\right]
\end{aligned}$$

under the assumption that $\mathrm{E}\left[\left(\frac{1-\pi_{ai}}{\pi_{ai}}\right) X_i X_i'\right]$ is invertible.

Of course, γ_a^* needs to be estimated. A sensible estimator is

$$\widehat{\gamma}_{an}^* = \left(\mathrm{E}_n\left[\left(\frac{1-\pi_{ai}}{\pi_{ai}}\right) X_i X_i'\right]\right)^{-1} \mathrm{E}_n\left[\left(\frac{1-\pi_{ai}}{\pi_{ai}}\right) \frac{Y_{ai}\mathbb{R}_{ai}}{\pi_{ai}} X_i\right]$$

Let's check that $\widehat{\gamma}_{an}^*$ is consistent for γ_a^*. By WLLN's,

$$\mathrm{E}_n\left[\left(\frac{1-\pi_{ai}}{\pi_{ai}}\right) X_i X_i'\right] \xrightarrow{p} \mathrm{E}\left[\left(\frac{1-\pi_{ai}}{\pi_{ai}}\right) X_i X_i'\right] \quad \text{as } n \to \infty$$

and

$$\begin{aligned}
\mathrm{E}_n\left[\left(\frac{1-\pi_{ai}}{\pi_{ai}}\right) \frac{Y_{ai}\mathbb{R}_{ai}}{\pi_{ai}} X_i\right] &\xrightarrow{p} \mathrm{E}\left[\left(\frac{1-\pi_{ai}}{\pi_{ai}}\right) \frac{Y_{ai}\mathbb{R}_{ai}}{\pi_{ai}} X_i\right] \quad \text{as } n \to \infty \\
&= \mathrm{E}\left[\mathrm{E}\left[\left(\frac{1-\pi_{ai}}{\pi_{ai}}\right) \frac{Y_{ai}\mathbb{R}_{ai}}{\pi_{ai}} X_i \middle| X_i\right]\right] \\
&= \mathrm{E}\left[\left(\frac{1-\pi_{ai}}{\pi_{ai}}\right) \frac{\mathrm{E}\left[Y_{ai}|X_i\right] \mathrm{E}\left[\mathbb{R}_{ai}|X_i\right]}{\pi_{ai}}\right] \quad \text{by conditional independence} \\
&= \mathrm{E}\left[\mathrm{E}\left[\left(\frac{1-\pi_{ai}}{\pi_{ai}}\right) Y_{ai}|X_i\right]\right] \quad \text{because } \pi_{ai} = \mathrm{E}\left[\mathbb{R}_{ai}|X_i\right].
\end{aligned}$$

Under the assumption that

$$\mathrm{E}\left[\left(\frac{1-\pi_{ai}}{\pi_{ai}}\right) X_i X_i'\right]$$

is invertible, we have by Slutsky's Theorem with $h(a,b) = a^{-1}b$ that

$$\left(\mathrm{E}_n\left[\left(\frac{1-\pi_{ai}}{\pi_{ai}}\right) X_i X_i'\right]\right)^{-1} \mathrm{E}_n\left[\left(\frac{1-\pi_{ai}}{\pi_{ai}}\right) \frac{Y_{ai}\mathbb{R}_{ai}}{\pi_{ai}} X_i\right] \xrightarrow{p} \mathrm{E}\left[\left(\frac{1-\pi_{ai}}{\pi_{ai}}\right) X_i X_i'\right]^{-1} \mathrm{E}\left[\left(\frac{1-\pi_{ai}}{\pi_{ai}}\right) Y_{ai} X_i\right]$$

as $n \to \infty$. That is, $\widehat{\gamma}_{an}^* \xrightarrow{p} \gamma_a^*$ as $n \to \infty$. $\square$

A.2 Proofs for Section 3.2

Proof of Theorem 3.10. We can rewrite the covariate-adjusted estimator $\widehat{\mu}_{an}$ as

$$\begin{aligned}
\widehat{\mu}_{an} &= \mathrm{E}_n\left[\frac{Y_{ai}\mathbb{R}_{ai}}{\mathrm{E}_n\left[\mathbb{R}_{ai}\right]}\right] - \widehat{\gamma}_{an}'\left(\mathrm{E}_n\left[\frac{X_i\left(\mathbb{R}_{ai} - \mathrm{E}_n\left[\mathbb{R}_{ai}\right]\right)}{\mathrm{E}_n\left[\mathbb{R}_{ai}\right]}\right]\right) \\
&= \frac{\mathrm{E}_n\left[Y_{ai}\mathbb{R}_{ai}\right]}{\mathrm{E}_n\left[\mathbb{R}_{ai}\right]} - \widehat{\gamma}_{an}' \frac{\left(\mathrm{E}_n\left[X_i\mathbb{R}_{ai}\right] - \mathrm{E}_n\left[X_i\right]\mathrm{E}_n\left[\mathbb{R}_{ai}\right]\right)}{\mathrm{E}_n\left[\mathbb{R}_{ai}\right]} \\
&= \mathrm{E}_n\left[Y_{ai}\right] + \mathrm{E}_n\left[\frac{Y_{ai}\left(\mathbb{R}_{ai} - \mathrm{E}_n\left[\mathbb{R}_{ai}\right]\right)}{\mathrm{E}_n\left[\mathbb{R}_{ai}\right]}\right] - \widehat{\gamma}_{an}' \frac{\left(\mathrm{E}_n\left[X_i\mathbb{R}_{ai}\right] - \mathrm{E}_n\left[X_i\right]\mathrm{E}_n\left[\mathbb{R}_{ai}\right]\right)}{\mathrm{E}_n\left[\mathbb{R}_{ai}\right]}
\end{aligned}$$

We to find the asymptotic variance of the estimator defined above. That is, we want to find the limiting variance of $\sqrt{n}\left(\widehat{\mu}_{an} - \mathrm{E}\left[Y_{ai}\right]\right)$.

The third term in the final expression above can be rewritten as

$$\begin{aligned}
&\widehat{\gamma}_{an}' \frac{\left(\mathrm{E}_n\left[X_i \mathbb{R}_{ai}\right] - \mathrm{E}_n\left[X_i\right]\mathrm{E}_n\left[\mathbb{R}_{ai}\right]\right)}{\mathrm{E}_n\left[\mathbb{R}_{ai}\right]} \\
&= \gamma_a' \frac{\left(\mathrm{E}_n\left[X_i \mathbb{R}_{ai}\right] - \mathrm{E}_n\left[X_i\right]\mathrm{E}_n\left[\mathbb{R}_{ai}\right]\right)}{\mathrm{E}_n\left[\mathbb{R}_{ai}\right]} + \left(\widehat{\gamma}_{an} - \gamma_a\right)' \frac{\left(\mathrm{E}_n\left[X_i \mathbb{R}_{ai}\right] - \mathrm{E}_n\left[X_i\right]\mathrm{E}_n\left[\mathbb{R}_{ai}\right]\right)}{\mathrm{E}_n\left[\mathbb{R}_{ai}\right]}
\end{aligned}$$

Note that

$$\begin{aligned}
&\sqrt{n}\left(\mathrm{E}_n\left(\frac{X_i\left(\mathbb{R}_{ai} - \mathrm{E}_n\left[\mathbb{R}_{ai}\right]\right)}{\mathrm{E}_n\left[\mathbb{R}_{ai}\right]}\right]\right) \\
=&\sqrt{n}\left(\mathrm{E}_n\left(\frac{\left(X_i - \mathrm{E}\left[X_i\right]\right)\left(\mathbb{R}_{ai} - \mathrm{E}_n\left[\mathbb{R}_{ai}\right]\right)}{\mathrm{E}_n\left[\mathbb{R}_{ai}\right]}\right]\right) \\
=&\sqrt{n}\left(\mathrm{E}_n\left(\frac{\left(X_i - \mathrm{E}\left[X_i\right]\right)\left(\mathbb{R}_{ai} - \pi_a\right)}{\mathrm{E}_n\left[\mathbb{R}_{ai}\right]}\right]\right) + \sqrt{n}\left(\mathrm{E}_n\left(\frac{\left(X_i - \mathrm{E}\left[X_i\right]\right)\left(\pi_a - \mathrm{E}_n\left[\mathbb{R}_{ai}\right]\right)}{\mathrm{E}_n\left[\mathbb{R}_{ai}\right]}\right]\right) \\
=&\frac{\sqrt{n}\left(\mathrm{E}_n\left[\left(X_i - \mathrm{E}\left[X_i\right]\right)\left(\mathbb{R}_{ai} - \pi_a\right)\right]\right)}{\mathrm{E}_n\left[\mathbb{R}_{ai}\right]} + \frac{\sqrt{n}\left(\mathrm{E}_n\left[X_i - \mathrm{E}\left[X_i\right]\right]\right)\left(\pi_a - \mathrm{E}_n\left[\mathbb{R}_{ai}\right]\right)}{\mathrm{E}_n\left[\mathbb{R}_{ai}\right]} \\
\xrightarrow{d}& \frac{Z_1}{\mathrm{E}\left[\mathbb{R}_{ai}\right]} + \frac{Z_2 \cdot 0}{\mathrm{E}\left[\mathbb{R}_{ai}\right]} \quad \text{as } n \to \infty \\
=&Z_1
\end{aligned}$$

where $Z_1 \sim \mathcal{N}\left(0, \mathrm{Var}\left(\left(X_i - \mathrm{E}\left[X_i\right]\right)\left(\mathbb{R}_{ai} - \pi_a\right)\right)\right)$ and $Z_2 \sim \mathcal{N}\left(0, \mathrm{Var}\left(X_i - \mathrm{E}\left[X_i\right]\right)\right)$. For the second-to-last step, we apply the CLT and WLLNs, and combine using Slutsky's Theorem. Since $\widehat{\gamma}_{an} - \gamma_a \xrightarrow{p} 0$ as $n \to \infty$ by the consistency assumption, combining results gives

$$\left(\widehat{\gamma}_{an} - \gamma_a\right)' \frac{\left(\mathrm{E}_n\left[X_i \mathbb{R}_{ai}\right] - \mathrm{E}_n\left[X_i\right]\mathrm{E}_n\left[\mathbb{R}_{ai}\right]\right)}{\mathrm{E}_n\left[\mathbb{R}_{ai}\right]} \xrightarrow{p} 0' Z_1 = 0.$$

Thus, $\widehat{\mu}_{an}$ has the same asymptotic distribution as

$$\widehat{\theta}_{an,\ \text{infeasible}} = \mathrm{E}_n\left[Y_{ai}\right] + \mathrm{E}_n\left[\frac{Y_{ai}\left(\mathbb{R}_{ai} - \mathrm{E}_n\left[\mathbb{R}_{ai}\right]\right)}{\mathrm{E}_n\left[\mathbb{R}_{ai}\right]}\right] - \gamma_a' \frac{\left(\mathrm{E}_n\left[X_i \mathbb{R}_{ai}\right] - \mathrm{E}_n\left[X_i\right]\mathrm{E}_n\left[\mathbb{R}_{ai}\right]\right)}{\mathrm{E}_n\left[\mathbb{R}_{ai}\right]}$$

which is analogous to our estimator but with $\widehat{\gamma}_{an}$ replaced by its probability limit. Thus, we can focus our attention exclusively on the infeasible estimator above for the rest of the proof as it is asymptotically equivalent to the feasible estimator $\widehat{\theta}_{an}$. Note that

$$\begin{aligned}
\widehat{\theta}_{an,\ \text{infeasible}} &= \mathrm{E}_n\left[Y_{ai}\right] + \mathrm{E}_n\left[\frac{\widetilde{Y}_{ai}\left(\mathbb{R}_{ai} - \mathrm{E}_n\left[\mathbb{R}_{ai}\right]\right)}{\mathrm{E}_n\left[\mathbb{R}_{ai}\right]}\right] \\
&= \mathrm{E}_n\left[Y_{ai}\right] + \frac{\mathrm{E}_n\left[\widetilde{Y}_{ai}\left(\mathbb{R}_{ai} - \mathrm{E}_n\left[\mathbb{R}_{ai}\right]\right)\right]}{\mathrm{E}_n\left[\mathbb{R}_{ai}\right]}
\end{aligned}$$

where $\widetilde{Y}_{ai} = Y_{ai} - \gamma_a' X_i$. Moreover, let $\widetilde{\mu}_a = \mathrm{E}\left[Y_{ai} - \gamma_a' X_i\right]$. We can rewrite the numerator of the second term above as follows:

$$\begin{aligned}
\mathrm{E}_n\left[\widetilde{Y}_{ai}\left(\mathbb{R}_{ai} - \mathrm{E}_n\left[\mathbb{R}_{ai}\right]\right)\right] &= \mathrm{E}_n\left[\left(\widetilde{Y}_{ai} - \mathrm{E}_n\left[\widetilde{Y}_{ai}\right]\right)\left(\mathbb{R}_{ai} - \mathrm{E}_n\left[\mathbb{R}_{ai}\right]\right)\right] \\
&= \mathrm{E}_n\left[\left(\widetilde{Y}_{ai} - \mathrm{E}_n\left[\widetilde{Y}_{ai}\right]\right)\left(\mathbb{R}_{ai} - \pi_a\right)\right] \\
&= \mathrm{E}_n\left[\left(\widetilde{Y}_{ai} - \widetilde{\mu}_a\right)\left(\mathbb{R}_{ai} - \pi_a\right)\right] + \mathrm{E}_n\left[\left(\widetilde{\mu}_a - \mathrm{E}_n\left[\widetilde{Y}_{ai}\right]\right)\left(\mathbb{R}_{ai} - \pi_a\right)\right] \\
&= \mathrm{E}_n\left[\left(\widetilde{Y}_{ai} - \widetilde{\mu}_a\right)\left(\mathbb{R}_{ai} - \pi_a\right)\right] + \left(\widetilde{\mu}_a - \mathrm{E}_n\left[\widetilde{Y}_{ai}\right]\right)\left(\mathrm{E}_n\left[\mathbb{R}_{ai}\right] - \pi_a\right)
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\sqrt{n}\left(\mathrm{E}_n\left[Y_{ai}\right] + \frac{\mathrm{E}_n\left[\widetilde{Y}_{ai}\left(\mathbb{R}_{ai} - \mathrm{E}_n\left[\mathbb{R}_{ai}\right]\right)\right]}{\mathrm{E}_n\left[\mathbb{R}_{ai}\right]} - \widetilde{\mu}_a\right) \\
=&\sqrt{n}\left(\mathrm{E}_n\left[Y_{ai}\right] - \mathrm{E}\left[Y_{ai}\right]\right) + \sqrt{n}\left(\frac{\mathrm{E}_n\left[\widetilde{Y}_{ai}\left(\mathbb{R}_{ai} - \mathrm{E}_n\left[\mathbb{R}_{ai}\right]\right)\right]}{\mathrm{E}_n\left[\mathbb{R}_{ai}\right]}\right) \\
=&\sqrt{n}\left(\mathrm{E}_n\left[Y_{ai}\right] - \mathrm{E}\left[Y_{ai}\right]\right) + \sqrt{n}\left(\frac{\mathrm{E}_n\left[\left(\widetilde{Y}_{ai} - \widetilde{\mu}_a\right)\left(\mathbb{R}_{ai} - \mathrm{E}_n\left[\mathbb{R}_{ai}\right]\right)\right]}{\mathrm{E}_n\left[\mathbb{R}_{ai}\right]}\right) \\
=&\sqrt{n}\left(\mathrm{E}_n\left[Y_{ai}\right] - \mathrm{E}\left[Y_{ai}\right]\right) + \frac{\sqrt{n}\left(\mathrm{E}_n\left[\left(\widetilde{Y}_{ai} - \widetilde{\mu}_a\right)\left(\mathbb{R}_{ai} - \pi_a\right)\right] + \left(\widetilde{\mu}_a - \mathrm{E}_n\left[\widetilde{Y}_{ai}\right]\right)\left(\mathrm{E}_n\left[\mathbb{R}_{ai}\right] - \pi_a\right)\right)}{\mathrm{E}_n\left[\mathbb{R}_{ai}\right]} \\
=&\sqrt{n}\left(\mathrm{E}_n\left[Y_{ai}\right] - \mathrm{E}\left[Y_{ai}\right]\right) + \frac{\sqrt{n}\,\mathrm{E}_n\left[\left(\widetilde{Y}_{ai} - \widetilde{\mu}_a\right)\left(\mathbb{R}_{ai} - \pi_a\right)\right]}{\mathrm{E}_n\left[\mathbb{R}_{ai}\right]} + \frac{\sqrt{n}\left(\widetilde{\mu}_a - \mathrm{E}_n\left[\widetilde{Y}_{ai}\right]\right)\left(\mathrm{E}_n\left[\mathbb{R}_{ai}\right] - \pi_a\right)}{\mathrm{E}_n\left[\mathbb{R}_{ai}\right]}
\end{aligned}$$

Now we find the asymptotic variance of the above expression. The first two terms can be analyzed by applying multivariate CLT and Slutsky's Theorem. We also show that the third term goes to zero in probability. By the multivariate CLT, we have

$$\sqrt{n}\begin{pmatrix} \mathrm{E}_n\left[Y_{ai}\right] - \mathrm{E}\left[Y_{ai}\right] \\ \mathrm{E}_n\left[\left(\widetilde{Y}_{ai} - \widetilde{\mu}_a\right)\left(\mathbb{R}_{ai} - \pi_a\right)\right] \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0, \mathrm{Var}\begin{pmatrix} Y_{ai} \\ \left(\widetilde{Y}_{ai} - \widetilde{\mu}_a\right)\left(\mathbb{R}_{ai} - \pi_a\right) \end{pmatrix}\right)$$

Note that

$$\begin{aligned}
\mathrm{Cov}\left(Y_{ai}, \left(\widetilde{Y}_{ai} - \widetilde{\mu}_a\right)\left(\mathbb{R}_{ai} - \pi_a\right)\right) &= \mathrm{E}\left[Y_{ai}\left(\widetilde{Y}_{ai} - \widetilde{\mu}_a\right)\left(\mathbb{R}_{ai} - \pi_a\right)\right] - \mathrm{E}\left[Y_{ai}\right]\mathrm{E}\left[\left(\widetilde{Y}_{ai} - \widetilde{\mu}_a\right)\left(\mathbb{R}_{ai} - \pi_a\right)\right] \\
&= \mathrm{E}\left[Y_{ai}\left(\widetilde{Y}_{ai} - \widetilde{\mu}_a\right)\right]\mathrm{E}\left[\left(\mathbb{R}_{ai} - \pi_a\right)\right] - \mathrm{E}\left[Y_{ai}\right]\mathrm{E}\left[\left(\widetilde{Y}_{ai} - \widetilde{\mu}_a\right)\right]\mathrm{E}\left[\left(\mathbb{R}_{ai} - \pi_a\right)\right] \\
&= 0
\end{aligned}$$

where we use the independence assumptions in the second step. Therefore, the off-diagonal entries of

$$\mathrm{Var}\begin{pmatrix} Y_{ai} \\ \left(\widetilde{Y}_{ai}-\widetilde{\mu}_a\right)\left(\mathbb{R}_{ai}-\pi_a\right)\end{pmatrix}$$

are zero. Moreover,

$$\begin{pmatrix} 1 \\ \frac{1}{\mathrm{E}_n[\mathbb{R}_{ai}]}\end{pmatrix} \xrightarrow{p} \begin{pmatrix} 1 \\ \frac{1}{\pi_a}\end{pmatrix}$$

by the WLLNs and Slutsky's Theorem. Combining results using Slutsky's Theorem, we have

$$\begin{aligned}
&\sqrt{n}\left(\mathrm{E}_n\left[Y_{ai}\right]-\mathrm{E}\left[Y_{ai}\right]\right)+\frac{\sqrt{n}\,\mathrm{E}_n\left[\left(\widetilde{Y}_{ai}-\widetilde{\mu}_a\right)\left(\mathbb{R}_{ai}-\pi_a\right)\right]}{\mathrm{E}_n\left[\mathbb{R}_{ai}\right]} \\
&=\begin{pmatrix} 1 \\ \frac{1}{\mathrm{E}_n[\mathbb{R}_{ai}]}\end{pmatrix}'\sqrt{n}\begin{pmatrix} \mathrm{E}_n\left[Y_{ai}\right]-\mathrm{E}\left[Y_{ai}\right] \\ \mathrm{E}_n\left[\left(\widetilde{Y}_{ai}-\widetilde{\mu}_a\right)\left(\mathbb{R}_{ai}-\pi_a\right)\right]\end{pmatrix} \\
&\xrightarrow{d}\mathcal{N}\left(0,\mathrm{Var}\left(Y_{ai}\right)+\frac{\mathrm{Var}\left(\left(\widetilde{Y}_{ai}-\widetilde{\mu}_a\right)\left(\mathbb{R}_{ai}-\pi_a\right)\right)}{\pi_a^2}\right)
\end{aligned}$$

Moreover,

$$\frac{\sqrt{n}\left(\widetilde{\mu}_a-\mathrm{E}_n\left[\widetilde{Y}_{ai}\right]\right)\left(\mathrm{E}_n\left[\mathbb{R}_{ai}\right]-\pi_a\right)}{\mathrm{E}_n\left[\mathbb{R}_{ai}\right]} \xrightarrow{p} 0$$

as $n \to \infty$ because $\sqrt{n}\left(\widetilde{\mu}_a-\mathrm{E}_n\left[\widetilde{Y}_{ai}\right]\right) \xrightarrow{d} Z \sim \mathcal{N}\left(0,\mathrm{Var}\left(\widetilde{Y}_{ai}\right)\right)$ by the CLT, $\left(\mathrm{E}_n\left[\mathbb{R}_{ai}\right]-\pi_a\right) \xrightarrow{p} 0$ by WLLNs, and $\mathrm{E}_n\left[\mathbb{R}_{ai}\right] \xrightarrow{p} \pi_a$ by WLLNs as $n \to \infty$, so combining results using Slutsky's Theorem yields the claim. Thus,

$$\sqrt{n}\left(\widehat{\mu}_{an}-\mathrm{E}\left[Y_{ai}\right]\right) \xrightarrow{d} \mathcal{N}\left(0,\mathrm{Var}\left(Y_{ai}\right)+\frac{\mathrm{Var}\left(\left(\widetilde{Y}_{ai}-\widetilde{\mu}_a\right)\left(\mathbb{R}_{ai}-\pi_a\right)\right)}{\pi_a^2}\right)$$

and where $\widetilde{\mu}_a=\mathrm{E}\left[Y_{ai}-\gamma_a' X_i\right]$.

Note that we can rewrite the second term of the asymptotic variance found above. In particular, we have

$$
\begin{aligned}
\frac{\operatorname{Var}\left(\left(\widetilde{Y}_{ai} - \mathrm{E}\left[\widetilde{Y}_{ai}\right]\right)\left(\mathbb{R}_{ai} - \pi_a\right)\right)}{\pi_a^2} &= \frac{\operatorname{Var}\left(\widetilde{Y}_{ai} - \mathrm{E}\left[\widetilde{Y}_{ai}\right]\right)\operatorname{Var}\left(\mathbb{R}_{ai} - \pi_a\right)}{\pi_a^2} \\
&= \frac{\operatorname{Var}\left(\widetilde{Y}_{ai}\right)\left(\pi_a\right)\left(1 - \pi_a\right)}{\pi_a^2} \\
&= \left(\frac{1 - \pi_a}{\pi_a}\right)\operatorname{Var}\left(\widetilde{Y}_{ai}\right) \\
&= \left(\frac{1 - \pi_a}{\pi_a}\right)\operatorname{Var}\left(Y_{ai} - \gamma_a' X_i\right)
\end{aligned}
$$

where for the first equality we use the fact that $\left(\widetilde{Y}_{ai} - \mathrm{E}\left[\widetilde{Y}_{ai}\right]\right)$ and $\left(\mathbb{R}_{ai} - \pi_a\right)$ are independent mean zero random variables. This completes the proof. □

Proof of Theorem 3.14. From Theorem 3.10, we have that minimizing the asymptotic variance of our estimator amounts to minimzing $\operatorname{Var}\left(\left(\widetilde{Y}_{ai} - \widetilde{\mu}_a\right)\left(\mathbb{R}_{ai} - \pi_a\right)\right)$, where $\widetilde{Y}_{ai} = Y_{ai} - \gamma_a' X_i$ and $\widetilde{\mu}_a = \mathrm{E}\left[\widetilde{Y}_{ai}\right]$. First note that

$$
\begin{aligned}
&\operatorname{Var}\left(\left(\widetilde{Y}_{ai} - \widetilde{\mu}_a\right)\left(\mathbb{R}_{ai} - \pi_a\right)\right) \\
=& \mathrm{E}\left[\operatorname{Var}\left(\left(\widetilde{Y}_{ai} - \widetilde{\mu}_a\right)\left(\mathbb{R}_{ai} - \pi_a\right) | \mathbb{R}_{ai}\right)\right] + \operatorname{Var}\left(\mathrm{E}\left[\left(\widetilde{Y}_{ai} - \widetilde{\mu}_a\right)\left(\mathbb{R}_{ai} - \pi_a\right) | \mathbb{R}_{ai}\right]\right) \quad \text{by LOTV} \\
=& \mathrm{E}\left[\left(\mathbb{R}_{ai} - \pi_a\right)^2 \operatorname{Var}\left(\widetilde{Y}_{ai}\right)\right] + \operatorname{Var}(0) \\
=& \operatorname{Var}\left(\widetilde{Y}_{ai}\right)\operatorname{Var}\left(\mathbb{R}_{ai}\right) \\
=& \operatorname{Var}\left(\widetilde{Y}_{ai}\right)\pi_a\left(1 - \pi_a\right)
\end{aligned}
$$

Thus, minimizing the asymptotic variance of our estimator amounts to minimizing $\operatorname{Var}\left(\widetilde{Y}_{ai}\right)$. We have

$$
\begin{aligned}
\operatorname{Var}\left(\widetilde{Y}_{ai}\right) &= \operatorname{Var}\left(Y_{ai} - \gamma_a' X_i\right) \\
&= \operatorname{Var}\left(Y_{ai}\right) + \operatorname{Var}\left(\gamma_a' X_i\right) - 2\operatorname{Cov}\left(Y_{ai}, \gamma_a' X_i\right) \\
&= \operatorname{Var}\left(Y_{ai}\right) + \gamma_a' \operatorname{Var}\left(X_i\right)\gamma_a - 2\gamma_a' \operatorname{Cov}\left(X_i, Y_{ai}\right)
\end{aligned}
$$

Taking the derivative with respect to γ_a, we have

$$
\frac{\partial}{\partial \gamma_a}\operatorname{Var}\left(\widetilde{Y}_{ai}\right) = 2\operatorname{Var}\left(X_i\right)\gamma_a - 2\operatorname{Cov}\left(X_i, Y_{ai}\right).
$$

Therefore, the FOC is

$$\begin{aligned}
&2\text{Var}\left(X_i\right)\gamma_a^* - 2\,\text{Cov}\left(X_i, Y_{ai}\right) = 0_p \\
\Longrightarrow\ &\text{Var}\left(X_i\right)\gamma_a^* = \text{Cov}\left(X_i, Y_{ai}\right) \\
\Longrightarrow\ &\gamma_a^* = \text{Var}\left(X_i\right)^{-1}\text{Cov}\left(X_i, Y_{ai}\right)
\end{aligned}$$

assuming that $\text{Var}\left(X_i\right)$ is invertible.

A consistent estimator of γ_a^* is given by

$$\widehat{\gamma}_{an}^* = \text{E}_n\left[\left(X_i - \text{E}_n\left[X_i\right]\right)\left(X_i - \text{E}_n\left[X_i\right]\right)'\right]^{-1}\left(\frac{\text{E}_n\left[X_i Y_{ai}\mathbb{R}_{ai}\right]}{\text{E}_n\left[\mathbb{R}_{ai}\right]} - \text{E}_n\left[X_i\right]\frac{\text{E}_n\left[Y_{ai}\mathbb{R}_{ai}\right]}{\text{E}_n\left[\mathbb{R}_{ai}\right]}\right)$$

Let's check that the above estimator is consistent for γ_a^*. We have

$$\begin{aligned}
\text{E}_n\left[\left(X_i - \text{E}_n\left[X_i\right]\right)\left(X_i - \text{E}_n\left[X_i\right]\right)'\right] &= \text{E}_n\left[X_i X_i'\right] - \text{E}_n\left[X_i\right]\text{E}_n\left[X_i\right]' \\
&\xrightarrow{p} E\left[X_i X_i'\right] - \text{E}\left[X_i\right]\text{E}\left[X_i\right]' \quad \text{as } n \to \infty \text{ by WLLNs + Slutsky's Thm} \\
&= \text{Var}\left(X_i\right).
\end{aligned}$$

Moreover, we have also by WLLNs and Slutksy's Theorem that

$$\begin{aligned}
\frac{\text{E}_n\left[X_i Y_{ai}\mathbb{R}_{ai}\right]}{\text{E}_n\left[\mathbb{R}_{ai}\right]} - \text{E}_n\left[X_i\right]\frac{\text{E}_n\left[Y_{ai}\mathbb{R}_{ai}\right]}{\text{E}_n\left[\mathbb{R}_{ai}\right]} &\xrightarrow{p} \frac{\text{E}\left[X_i Y_{ai}\mathbb{R}_{ai}\right]}{\text{E}\left[\mathbb{R}_{ai}\right]} - \text{E}\left[X_i\right]\frac{\text{E}\left[Y_{ai}\mathbb{R}_{ai}\right]}{\text{E}\left[\mathbb{R}_{ai}\right]} \quad \text{as } n \to \infty \\
&= \frac{\text{E}\left[X_i Y_{ai}\right]\text{E}\left[\mathbb{R}_{ai}\right]}{\text{E}\left[\mathbb{R}_{ai}\right]} - \text{E}\left[X_i\right]\frac{\text{E}\left[Y_{ai}\right]\text{E}\left[\mathbb{R}_{ai}\right]}{\text{E}\left[\mathbb{R}_{ai}\right]} \\
&= \text{E}\left[X_i Y_{ai}\right] - \text{E}\left[X_i\right]\text{E}\left[Y_{ai}\right] \\
&= \text{Cov}\left(X_i, Y_{ai}\right)
\end{aligned}$$

Combining results using Slutsky's Theorem with $h(x,y) = x^{-1}y$ yields $\widehat{\gamma}_{an}^* \xrightarrow{p} \gamma_a^*$ as $n \to \infty$, as desired. □

A.3 Proofs for Section 3.3

Proof of Theorem 3.17. Let's begin by rewriting our estimator $\widehat{\mu}_{an}$. We have

$$\begin{aligned}
\widehat{\mu}_{an} &= \text{E}_n\left[\frac{Y_{ai}\mathbb{R}_{ai}}{\text{E}_n\left[\mathbb{R}_{ai}|S_i\right]}\right] - \widehat{\gamma}_{an}'\left(\text{E}_n\left[\frac{X_i\left(\mathbb{R}_{ai} - \text{E}_n\left[\mathbb{R}_{ai}|S_i\right]\right)}{\text{E}_n\left[\mathbb{R}_{ai}|S_i\right]}\right]\right) \\
&= \text{E}_n\left[Y_{ai}\right] + \text{E}_n\left[\frac{Y_{ai}\left(\mathbb{R}_{ai} - \text{E}_n\left[\mathbb{R}_{ai}|S_i\right]\right)}{\text{E}_n\left[\mathbb{R}_{ai}|S_i\right]}\right] - \widehat{\gamma}_{an}'\left(\text{E}_n\left[\frac{X_i\left(\mathbb{R}_{ai} - \text{E}_n\left[\mathbb{R}_{ai}|S_i\right]\right)}{\text{E}_n\left[\mathbb{R}_{ai}|S_i\right]}\right]\right) \\
&= \text{E}_n\left[Y_{ai}\right] + \text{E}_n\left[\frac{\widetilde{Y}_{ai}\left(\mathbb{R}_{ai} - \text{E}_n\left[\mathbb{R}_{ai}|S_i\right]\right)}{\text{E}_n\left[\mathbb{R}_{ai}|S_i\right]}\right] - \left(\widehat{\gamma}_{an} - \gamma_a\right)'\left(\text{E}_n\left[\frac{X_i\left(\mathbb{R}_{ai} - \text{E}_n\left[\mathbb{R}_{ai}|S_i\right]\right)}{\text{E}_n\left[\mathbb{R}_{ai}|S_i\right]}\right]\right)
\end{aligned}$$

where $\widetilde{Y}_{ai} = Y_{ai} - \gamma_a' X_i$. Let's first show that the third term is $o_p(1/\sqrt{n})$, so that we can focus on the first two terms of the final expansion above. By the consistency of $\widehat{\gamma}_{an}$, we have that $(\widehat{\gamma}_{an} - \gamma_a) \xrightarrow{p} 0$ as $n \to \infty$. Let the support of S_i be given by $\{1, \ldots, k\}$. Then

$$
\begin{aligned}
&\mathrm{E}_n \left[\frac{X_i (\mathbb{R}_{ai} - \mathrm{E}_n [\mathbb{R}_{ai}|S_i])}{\mathrm{E}_n [\mathbb{R}_{ai}|S_i]} \right] \\
=& \mathrm{E}_n \left[\mathrm{E}_n \left[\frac{X_i (\mathbb{R}_{ai} - \mathrm{E}_n [\mathbb{R}_{ai}|S_i])}{\mathrm{E}_n [\mathbb{R}_{ai}|S_i]} \middle| S_i \right] \right] \\
=& \sum_{s=1}^{k} \mathrm{E}_n [\mathbb{1}\{S_i = s\}] \, \mathrm{E}_n \left[\frac{X_i (\mathbb{R}_{ai} - \mathrm{E}_n [\mathbb{R}_{ai}|S_i])}{\mathrm{E}_n [\mathbb{R}_{ai}|S_i]} \middle| S_i = s \right] \\
=& \sum_{s=1}^{k} \mathrm{E}_n [\mathbb{1}\{S_i = s\}] \, \frac{\mathrm{E}_n [X_i (\mathbb{R}_{ai} - \mathrm{E}_n [\mathbb{R}_{ai}|S_i]) | S_i = s]}{\mathrm{E}_n [\mathbb{R}_{ai}|S_i = s]} \\
=& \sum_{s=1}^{k} \mathrm{E}_n [\mathbb{1}\{S_i = s\}] \, \frac{\mathrm{E}_n [(X_i - \mathrm{E}_n [X_i|S_i]) (\mathbb{R}_{ai} - \mathrm{E}_n [\mathbb{R}_{ai}|S_i]) | S_i = s]}{\mathrm{E}_n [\mathbb{R}_{ai}|S_i = s]} \\
=& \sum_{s=1}^{k} \mathrm{E}_n [\mathbb{1}\{S_i = s\}] \, \frac{\mathrm{E}_n [(X_i - \mathrm{E}_n [X_i|S_i]) (\mathbb{R}_{ai} - \mathrm{E} [\mathbb{R}_{ai}|S_i]) | S_i = s]}{\mathrm{E}_n [\mathbb{R}_{ai}|S_i = s]} \\
=& \sum_{s=1}^{k} \mathrm{E}_n [\mathbb{1}\{S_i = s\}] \, \frac{\mathrm{E}_n [(X_i - \mathrm{E} [X_i|S_i]) (\mathbb{R}_{ai} - \mathrm{E} [\mathbb{R}_{ai}|S_i]) | S_i = s]}{\mathrm{E}_n [\mathbb{R}_{ai}|S_i = s]} \\
&+ \sum_{s=1}^{k} \mathrm{E}_n [\mathbb{1}\{S_i = s\}] \, \frac{\mathrm{E}_n [(\mathrm{E} [X_i|S_i] - \mathrm{E}_n [X_i|S_i]) (\mathbb{R}_{ai} - \mathrm{E} [\mathbb{R}_{ai}|S_i]) | S_i = s]}{\mathrm{E}_n [\mathbb{R}_{ai}|S_i = s]} \\
=& \sum_{s=1}^{k} \mathrm{E}_n [\mathbb{1}\{S_i = s\}] \, \frac{\mathrm{E}_n [(X_i - \mathrm{E} [X_i|S_i]) (\mathbb{R}_{ai} - \mathrm{E} [\mathbb{R}_{ai}|S_i]) \mathbb{1}\{S_i = s\}] / \mathrm{E}_n [\mathbb{1}\{S_i = s\}]}{\mathrm{E}_n [\mathbb{1}\{S_i = s\} \mathbb{R}_{ai}] / \mathrm{E}_n [\mathbb{1}\{S_i = s\}]} \\
&+ \sum_{s=1}^{k} \mathrm{E}_n [\mathbb{1}\{S_i = s\}] \, \frac{(\mathrm{E} [X_i|S_i = s] - \mathrm{E}_n [X_i|S_i = s]) (\mathrm{E}_n [\mathbb{R}_{ai}|S_i = s] - \mathrm{E} [\mathbb{R}_{ai}|S_i = s])}{\mathrm{E}_n [\mathbb{R}_{ai}|S_i = s]} \\
=& \sum_{s=1}^{k} \frac{\mathrm{E}_n [\mathbb{1}\{S_i = s\}]}{\mathrm{E}_n [\mathbb{1}\{S_i = s\} \mathbb{R}_{ai}]} \mathrm{E}_n [(X_i - \mathrm{E} [X_i|S_i]) (\mathbb{R}_{ai} - \mathrm{E} [\mathbb{R}_{ai}|S_i]) \mathbb{1}\{S_i = s\}] \\
&+ \sum_{s=1}^{k} \frac{(\mathrm{E}_n [\mathbb{1}\{S_i = s\}])^2}{\mathrm{E}_n [\mathbb{R}_{ai} \mathbb{1}\{S_i = s\}]} \left(\frac{\mathrm{E} [X_i \mathbb{1}\{S_i = s\}]}{\mathrm{E} [\mathbb{1}\{S_i = s\}]} - \frac{\mathrm{E}_n [X_i \mathbb{1}\{S_i = s\}]}{\mathrm{E}_n [\mathbb{1}\{S_i = s\}]} \right) \\
&\times \left(\frac{\mathrm{E}_n [\mathbb{R}_{ai} \mathbb{1}\{S_i = s\}]}{\mathrm{E}_n [\mathbb{1}\{S_i = s\}]} - \frac{\mathrm{E} [\mathbb{R}_{ai} \mathbb{1}\{S_i = s\}]}{\mathrm{E} [\mathbb{1}\{S_i = s\}]} \right)
\end{aligned}
$$

Note that $\mathrm{E}_n [\mathbb{1}\{S_i = s\}] \xrightarrow{p} \mathrm{E} [\mathbb{1}\{S_i = s\}]$ and $\mathrm{E}_n [\mathbb{1}\{S_i = s\} \mathbb{R}_{ai}] \xrightarrow{p} \mathrm{E} [\mathbb{1}\{S_i = s\} \mathbb{R}_{ai}]$ as $n \to \infty$ by WLLNs. Therefore, applying Slutsky's Theorem,

$$
\frac{\mathrm{E}_n [\mathbb{1}\{S_i = s\}]}{\mathrm{E}_n [\mathbb{1}\{S_i = s\} \mathbb{R}_{ai}]} \xrightarrow{p} \frac{\mathrm{E} [\mathbb{1}\{S_i = s\}]}{\mathrm{E} [\mathbb{1}\{S_i = s\} \mathbb{R}_{ai}]}
$$

as $n \to \infty$, so $\frac{\mathrm{E}_n[\mathbb{1}\{S_i=s\}]}{\mathrm{E}_n[\mathbb{1}\{S_i=s\}\mathbb{R}_{ai}]} = O_p(1)$. Moreover, noting that

$$\begin{aligned}\mathrm{E}\left[\left(X_i - \mathrm{E}\left[X_i|S_i\right]\right)\left(\mathbb{R}_{ai} - \mathrm{E}\left[\mathbb{R}_{ai}|S_i\right]\right)\mathbb{1}\{S_i = s\}\right] &= \mathrm{E}\left[\left(X_i - \mathrm{E}\left[X_i|S_i\right]\right)\mathbb{1}\{S_i = s\}\right]\mathrm{E}\left[\mathbb{R}_{ai} - \mathrm{E}\left[\mathbb{R}_{ai}\right]\right] \\ &= 0\end{aligned}$$

we have

$$\mathrm{E}_n\left[\left(X_i - \mathrm{E}\left[X_i|S_i\right]\right)\left(\mathbb{R}_{ai} - \mathrm{E}\left[\mathbb{R}_{ai}|S_i\right]\right)\mathbb{1}\{S_i = s\}\right] = O_p\left(\frac{1}{\sqrt{n}}\right)$$

by the CLT. Similarly,

$$\frac{\left(\mathrm{E}_n\left[\mathbb{1}\{S_i = s\}\right]\right)^2}{\mathrm{E}_n\left[\mathbb{R}_{ai}\mathbb{1}\{S_i = s\}\right]} = O_p(1)$$

by a combination of the WLLNs and Slutsky's Theorem. For ease of notation, let $a = \mathrm{E}\left[X_i\mathbb{1}\{S_i = s\}\right]$, $b = \mathrm{E}\left[\mathbb{1}\{S_i = s\}\right]$, $a_n = \mathrm{E}_n\left[X_i\mathbb{1}\{S_i = s\}\right]$, and $b_n = \mathrm{E}_n\left[\mathbb{1}\{S_i = s\}\right]$. Then

$$\begin{aligned}\frac{a}{b} - \frac{a_n}{b_n} &= \frac{ab_n - a_nb}{bb_n} \\ &= \frac{a\left(b_n - b\right) - b\left(a_n - a\right)}{bb_n}.\end{aligned}$$

We know that $a\left(b_n - b\right) = O_p\left(\frac{1}{\sqrt{n}}\right)$ and $b\left(a_n - a\right) = O_p\left(\frac{1}{\sqrt{n}}\right)$ by the CLT. Moreover, $\left(bb_n\right)^{-1} = O_p(1)$ by WLLNs and Slutsky's Theorem. Thus,

$$\begin{aligned}\frac{a}{b} - \frac{a_n}{b_n} &= O_p(1)\left(O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{1}{\sqrt{n}}\right)\right) \\ &= O_p\left(\frac{2}{\sqrt{n}}\right)\end{aligned}$$

The same argument above holds if we let $a = \mathrm{E}\left[\mathbb{R}_{ai}\mathbb{1}\{S_i = s\}\right]$ and $a_n = \mathrm{E}_n\left[\mathbb{R}_{ai}\mathbb{1}\{S_i = s\}\right]$. Compiling results, we have

$$\begin{aligned}&\sum_{s=1}^{k}\frac{\mathrm{E}_n\left[\mathbb{1}\{S_i = s\}\right]}{\mathrm{E}_n\left[\mathbb{1}\{S_i = s\}\mathbb{R}_{ai}\right]}\mathrm{E}_n\left[\left(X_i - \mathrm{E}\left[X_i|S_i\right]\right)\left(\mathbb{R}_{ai} - \mathrm{E}\left[\mathbb{R}_{ai}|S_i\right]\right)\mathbb{1}\{S_i = s\}\right] \\ &+ \sum_{s=1}^{k}\frac{\left(\mathrm{E}_n\left[\mathbb{1}\{S_i = s\}\right]\right)^2}{\mathrm{E}_n\left[\mathbb{R}_{ai}\mathbb{1}\{S_i = s\}\right]}\left(\frac{\mathrm{E}\left[X_i\mathbb{1}\{S_i = s\}\right]}{\mathrm{E}\left[\mathbb{1}\{S_i = s\}\right]} - \frac{\mathrm{E}_n\left[X_i\mathbb{1}\{S_i = s\}\right]}{\mathrm{E}_n\left[\mathbb{1}\{S_i = s\}\right]}\right) \\ &\times\left(\frac{\mathrm{E}_n\left[\mathbb{R}_{ai}\mathbb{1}\{S_i = s\}\right]}{\mathrm{E}_n\left[\mathbb{1}\{S_i = s\}\right]} - \frac{\mathrm{E}\left[\mathbb{R}_{ai}\mathbb{1}\{S_i = s\}\right]}{\mathrm{E}\left[\mathbb{1}\{S_i = s\}\right]}\right) \\ =&\sum_{s=1}^{k}O_p(1)O_p\left(\frac{1}{\sqrt{n}}\right) + \sum_{s=1}^{k}O_p(1)O_p\left(\frac{2}{\sqrt{n}}\right)O_p\left(\frac{2}{\sqrt{n}}\right) \\ =&O_p\left(\frac{k}{\sqrt{n}}\right) + O_p\left(\frac{4k}{n}\right) \\ =&O_p\left(\frac{k}{\sqrt{n}}\right).\end{aligned}$$

Therefore,

$$\begin{aligned}\sqrt{n}\left(\widehat{\gamma}_{an}-\gamma_a\right)'\left(\mathrm{E}_n\left[\frac{X_i\left(\mathbb{R}_{ai}-\mathrm{E}_n\left[\mathbb{R}_{ai}|S_i\right]\right)}{\mathrm{E}_n\left[\mathbb{R}_{ai}|S_i\right]}\right]\right) &= \sqrt{n}o_p(1)O_p\left(\frac{k}{\sqrt{n}}\right)\\ &= ko_p(1)O_p(1)\\ &= o_p(1)\end{aligned}$$

so we can ignore the third term in the expanded for of $\widehat{\mu}_{an}$ in determining its asymptotic distribution.

Next, note that

$$\begin{aligned}&\mathrm{E}_n\left[\frac{\widetilde{Y}_{ai}\left(\mathbb{R}_{ai}-\mathrm{E}_n\left[\mathbb{R}_{ai}|S_i\right]\right)}{\mathrm{E}_n\left[\mathbb{R}_{ai}|S_i\right]}\right]\\ &=\mathrm{E}_n\left[\frac{\left(\widetilde{Y}_{ai}-\mathrm{E}_n\left[\widetilde{Y}_{ai}|S_i\right]\right)\left(\mathbb{R}_{ai}-\mathrm{E}_n\left[\mathbb{R}_{ai}|S_i\right]\right)}{\mathrm{E}_n\left[\mathbb{R}_{ai}|S_i\right]}\right]\\ &=\mathrm{E}_n\left[\frac{\left(\widetilde{Y}_{ai}-\mathrm{E}_n\left[\widetilde{Y}_{ai}|S_i\right]\right)\left(\mathbb{R}_{ai}-\mathrm{E}\left[\mathbb{R}_{ai}|S_i\right]\right)}{\mathrm{E}_n\left[\mathbb{R}_{ai}|S_i\right]}\right]\\ &=\mathrm{E}_n\left[\frac{\left(\widetilde{Y}_{ai}-\mathrm{E}\left[\widetilde{Y}_{ai}|S_i\right]\right)\left(\mathbb{R}_{ai}-\mathrm{E}_n\left[\mathbb{R}_{ai}|S_i\right]\right)}{\mathrm{E}_n\left[\mathbb{R}_{ai}|S_i\right]}\right]+\mathrm{E}_n\left[\frac{\left(\mathrm{E}\left[\widetilde{Y}_{ai}|S_i\right]-\mathrm{E}_n\left[\widetilde{Y}_{ai}|S_i\right]\right)\left(\mathbb{R}_{ai}-\mathrm{E}\left[\mathbb{R}_{ai}|S_i\right]\right)}{\mathrm{E}_n\left[\mathbb{R}_{ai}|S_i\right]}\right]\end{aligned}$$

Let's show that the second term above is $o_p\left(\frac{1}{\sqrt{n}}\right)$, so we can also ignore it asymptotically. We have

$$
\begin{aligned}
&\mathrm{E}_n\left[\frac{\left(\mathrm{E}\left[\widetilde{Y}_{ai}|S_i\right]-\mathrm{E}_n\left[\widetilde{Y}_{ai}|S_i\right]\right)\left(\mathbb{R}_{ai}-\mathrm{E}\left[\mathbb{R}_{ai}|S_i\right]\right)}{\mathrm{E}_n\left[\mathbb{R}_{ai}|S_i\right]}\right] \\
&=\mathrm{E}_n\left[\mathrm{E}_n\left[\left.\frac{\left(\mathrm{E}\left[\widetilde{Y}_{ai}|S_i\right]-\mathrm{E}_n\left[\widetilde{Y}_{ai}|S_i\right]\right)\left(\mathbb{R}_{ai}-\mathrm{E}\left[\mathbb{R}_{ai}|S_i\right]\right)}{\mathrm{E}_n\left[\mathbb{R}_{ai}|S_i\right]}\right|S_i\right]\right] \\
&=\mathrm{E}_n\left[\frac{\left(\mathrm{E}\left[\widetilde{Y}_{ai}|S_i\right]-\mathrm{E}_n\left[\widetilde{Y}_{ai}|S_i\right]\right)\left(\mathrm{E}_n\left[\mathbb{R}_{ai}|S_i\right]-\mathrm{E}\left[\mathbb{R}_{ai}|S_i\right]\right)}{\mathrm{E}_n\left[\mathbb{R}_{ai}|S_i\right]}\right] \\
&=\sum_{s=1}^{k}\mathrm{E}_n\left[\mathbb{1}\{S_i=s\}\right]\mathrm{E}_n\left[\left.\frac{\left(\mathrm{E}\left[\widetilde{Y}_{ai}|S_i\right]-\mathrm{E}_n\left[\widetilde{Y}_{ai}|S_i\right]\right)\left(\mathrm{E}_n\left[\mathbb{R}_{ai}|S_i\right]-\mathrm{E}\left[\mathbb{R}_{ai}|S_i\right]\right)}{\mathrm{E}_n\left[\mathbb{R}_{ai}|S_i\right]}\right|S_i=s\right] \\
&=\sum_{s=1}^{k}\frac{\left(\mathrm{E}_n\left[\mathbb{1}\{S_i=s\}\right]\right)^2}{\mathrm{E}_n\left[\mathbb{R}_{ai}\mathbb{1}\{S_i=s\}\right]}\left(\frac{\mathrm{E}\left[\widetilde{Y}_{ai}\mathbb{1}\{S_i=s\}\right]}{\mathrm{E}\left[\mathbb{1}\{S_i=s\}\right]}-\frac{\mathrm{E}_n\left[\widetilde{Y}_{ai}\mathbb{1}\{S_i=s\}\right]}{\mathrm{E}_n\left[\mathbb{1}\{S_i=s\}\right]}\right) \\
&\quad\times\left(\frac{\mathrm{E}_n\left[\mathbb{R}_{ai}\mathbb{1}\{S_i=s\}\right]}{\mathrm{E}_n\left[\mathbb{1}\{S_i=s\}\right]}-\frac{\mathrm{E}\left[\mathbb{R}_{ai}\mathbb{1}\{S_i=s\}\right]}{\mathrm{E}\left[\mathbb{1}\{S_i=s\}\right]}\right)
\end{aligned}
$$

We showed above that this expression with $\widetilde{Y}_{ai}$ replaced by X_i is $O_p\left(\frac{4k}{n}\right)$. Since the above argument still goes through with $\widetilde{Y}_{ai}$ instead of X_i, we have that the expression $O_p\left(\frac{4k}{n}\right)$. This implies

$$
\sqrt{n}\,\mathrm{E}_n\left[\frac{\left(\mathrm{E}\left[\widetilde{Y}_{ai}|S_i\right]-\mathrm{E}_n\left[\widetilde{Y}_{ai}|S_i\right]\right)\left(\mathbb{R}_{ai}-\mathrm{E}\left[\mathbb{R}_{ai}|S_i\right]\right)}{\mathrm{E}_n\left[\mathbb{R}_{ai}|S_i\right]}\right]\xrightarrow{p}0
$$

as $n\to\infty$, so we can ignore this term when considering the asymptotic variance of $\widehat{\mu}_{an}$.

To summarize, we have shown thus far that

$$
\sqrt{n}\left(\widehat{\mu}_{an}-\mathrm{E}\left[Y_{ai}\right]\right)
$$

has the same limiting distribution as

$$
\sqrt{n}\left(\mathrm{E}_n\left[Y_{ai}\right]+\mathrm{E}_n\left[\frac{\left(\widetilde{Y}_{ai}-\mathrm{E}\left[\widetilde{Y}_{ai}|S_i\right]\right)\left(\mathbb{R}_{ai}-\mathrm{E}\left[\mathbb{R}_{ai}|S_i\right]\right)}{\mathrm{E}_n\left[\mathbb{R}_{ai}|S_i\right]}\right]-\mathrm{E}\left[Y_{ai}\right]\right).
$$

Note that

$$
\begin{aligned}
&\mathrm{E}_n\left[\frac{\left(\widetilde{Y}_{ai}-\mathrm{E}\left[\widetilde{Y}_{ai}|S_i\right]\right)\left(\mathbb{R}_{ai}-\mathrm{E}\left[\mathbb{R}_{ai}|S_i\right]\right)}{\mathrm{E}_n\left[\mathbb{R}_{ai}|S_i\right]}\right] \\
=&\mathrm{E}_n\left[\mathrm{E}_n\left[\frac{\left(\widetilde{Y}_{ai}-\mathrm{E}\left[\widetilde{Y}_{ai}|S_i\right]\right)\left(\mathbb{R}_{ai}-\mathrm{E}\left[\mathbb{R}_{ai}|S_i\right]\right)}{\mathrm{E}_n\left[\mathbb{R}_{ai}|S_i\right]}\Big|S_i\right]\right] \\
=&\sum_{s=1}^{k}\mathrm{E}_n\left[\mathbb{1}\{S_i=s\}\right]\mathrm{E}_n\left[\frac{\left(\widetilde{Y}_{ai}-\mathrm{E}\left[\widetilde{Y}_{ai}|S_i\right]\right)\left(\mathbb{R}_{ai}-\mathrm{E}\left[\mathbb{R}_{ai}|S_i\right]\right)}{\mathrm{E}_n\left[\mathbb{R}_{ai}|S_i\right]}\Big|S_i=s\right] \\
=&\sum_{s=1}^{k}\frac{\mathrm{E}_n\left[\mathbb{1}\{S_i=s\}\right]}{\mathrm{E}_n\left[\mathbb{R}_{ai}|S_i=s\right]}\mathrm{E}_n\left[\left(\widetilde{Y}_{ai}-\mathrm{E}\left[\widetilde{Y}_{ai}|S_i\right]\right)\left(\mathbb{R}_{ai}-\mathrm{E}\left[\mathbb{R}_{ai}|S_i\right]\right)|S_i=s\right].
\end{aligned}
$$

For each $s\in\{1,\ldots,k\}$, we have

$$
\begin{aligned}
&\mathrm{E}_n\left[\left(\widetilde{Y}_{ai}-\mathrm{E}\left[\widetilde{Y}_{ai}|S_i\right]\right)\left(\mathbb{R}_{ai}-\mathrm{E}\left[\mathbb{R}_{ai}|S_i\right]\right)|S_i=s\right] \\
=&\frac{\mathrm{E}_n\left[\mathbb{1}\{S_i=s\}\left(\widetilde{Y}_{ai}-\mathrm{E}\left[\widetilde{Y}_{ai}|S_i=s\right]\right)\left(\mathbb{R}_{ai}-\mathrm{E}\left[\mathbb{R}_{ai}|S_i=s\right]\right)\right]}{\mathrm{E}_n\left[\mathbb{1}\{S_i=s\}\right]}
\end{aligned}
$$

Note that $\mathrm{E}\left[\mathbb{R}_{ai}|S_i=s\right]=\mathrm{E}\left[\mathbb{R}_{ai}\right]=\pi_a$ by our assumptions. Moreover, $\mathbb{R}_{ai}$ is independent of Y_{ai} and $\mathbb{1}\{S_i=s\}$. Thus,

$$
\mathrm{E}\left[\mathbb{1}\{S_i=s\}\left(\widetilde{Y}_{ai}-\mathrm{E}\left[\widetilde{Y}_{ai}|S_i=s\right]\right)\left(\mathbb{R}_{ai}-\mathrm{E}\left[\mathbb{R}_{ai}|S_i=s\right]\right)\right]=0.
$$

Let

$$
g_{sn}=\mathrm{E}_n\left[\mathbb{1}\{S_i=s\}\left(\widetilde{Y}_{ai}-\mathrm{E}\left[\widetilde{Y}_{ai}|S_i=s\right]\right)\left(\mathbb{R}_{ai}-\mathrm{E}\left[\mathbb{R}_{ai}|S_i=s\right]\right)\right]
$$

and

$$
g_{si}=\mathbb{1}\{S_i=s\}\left(\widetilde{Y}_{ai}-\mathrm{E}\left[\widetilde{Y}_{ai}|S_i=s\right]\right)\left(\mathbb{R}_{ai}-\mathrm{E}\left[\mathbb{R}_{ai}|S_i=s\right]\right).
$$

Then by the multivariate CLT,

$$
\sqrt{n}\begin{pmatrix}\mathrm{E}_n\left[Y_{ai}\right]-\mathrm{E}\left[Y_{ai}\right]\\ g_{1n}\\ g_{2n}\\ \vdots\\ g_{kn}\end{pmatrix}\xrightarrow{d}\mathcal{N}\left(0,\Sigma\right)
$$

as $n \to \infty$, where $\Sigma = \mathrm{Var}\left(Y_{ai}, g_{1i}, g_{2i}, \ldots, g_{ki}\right)'$. Note that for $a, b \in \{1, \ldots, k\}$ with $a \neq b$,

$$\begin{aligned}\mathrm{Cov}\Big(&\mathbb{1}\{S_i = a\}\left(\widetilde{Y}_{ai} - \mathrm{E}\left[\widetilde{Y}_{ai}|S_i = a\right]\right)\left(\mathbb{R}_{ai} - \mathrm{E}\left[\mathbb{R}_{ai}|S_i = a\right]\right), \\ &\mathbb{1}\{S_i = b\}\left(\widetilde{Y}_{ai} - \mathrm{E}\left[\widetilde{Y}_{ai}|S_i = b\right]\right)\left(\mathbb{R}_{ai} - \mathrm{E}\left[\mathbb{R}_{ai}|S_i = b\right]\right)\Big) \\ &= 0\end{aligned}$$

because $\mathbb{1}\{S_i = a\}\mathbb{1}\{S_i = b\} = 0$, since a unit can be assigned to only one bucket. We also note that

$$\begin{aligned}&\mathrm{Cov}\left(Y_{ai}, \mathbb{1}\{S_i = s\}\left(\widetilde{Y}_{ai} - \mathrm{E}\left[\widetilde{Y}_{ai}|S_i = s\right]\right)\left(\mathbb{R}_{ai} - \mathrm{E}\left[\mathbb{R}_{ai}|S_i = s\right]\right)\right) \\ =&\mathrm{E}\left[\left(Y_{ai} - \mathrm{E}\left[Y_{ai}\right]\right)\left(\widetilde{Y}_{ai} - \mathrm{E}\left[\widetilde{Y}_{ai}|S_i = s\right]\right)\left(\mathbb{R}_{ai} - \mathrm{E}\left[\mathbb{R}_{ai}|S_i = s\right]\right)\right] \\ =&\mathrm{E}\left[\left(Y_{ai} - \mathrm{E}\left[Y_{ai}\right]\right)\left(\widetilde{Y}_{ai} - \mathrm{E}\left[\widetilde{Y}_{ai}|S_i = s\right]\right)\right]\mathrm{E}\left[\mathbb{R}_{ai} - \mathrm{E}\left[\mathbb{R}_{ai}|S_i = s\right]\right] \\ =&0.\end{aligned}$$

Therefore, Σ is diagonal. Next, note that

$$\frac{\mathrm{E}_n\left[\mathbb{1}\{S_i = s\}\right]}{\mathrm{E}_n\left[\mathbb{R}_{ai}|S_i = s\right]} \cdot \frac{1}{\mathrm{E}_n\left[\mathbb{1}\{S_i = s\}\right]} = \frac{1}{\mathrm{E}_n\left[\mathbb{R}_{ai}|S_i = s\right]}.$$

Define

$$\begin{aligned}c_n &= \left(1, \frac{1}{\mathrm{E}_n\left[\mathbb{R}_{ai}|S_i = 1\right]}, \ldots, \frac{1}{\mathrm{E}_n\left[\mathbb{R}_{ai}|S_i = k\right]}\right)' \\ &= \left(1, \frac{\mathrm{E}_n\left[\mathbb{1}\{S_i = 1\}\right]}{\mathrm{E}_n\left[\mathbb{R}_{ai}\mathbb{1}\{S_i = 1\}\right]}, \ldots, \frac{\mathrm{E}_n\left[\mathbb{1}\{S_i = k\}\right]}{\mathrm{E}_n\left[\mathbb{R}_{ai}\mathbb{1}\{S_i = k\}\right]}\right)'\end{aligned}$$

By WLLNs and Slutsky's Theorem,

$$c_n \xrightarrow{p} c$$

as $n \to \infty$, where

$$\begin{aligned}c &= \left(1, \frac{\mathrm{E}\left[\mathbb{1}\{S_i = 1\}\right]}{\mathrm{E}\left[\mathbb{R}_{ai}\mathbb{1}\{S_i = 1\}\right]}, \ldots, \frac{\mathrm{E}\left[\mathbb{1}\{S_i = k\}\right]}{\mathrm{E}\left[\mathbb{R}_{ai}\mathbb{1}\{S_i = k\}\right]}\right)' \\ &= \left(1, \frac{1}{\mathrm{E}\left[\mathbb{R}_{ai}|S_i = 1\right]}, \ldots, \frac{1}{\mathrm{E}\left[\mathbb{R}_{ai}|S_i = k\right]}\right)' \\ &= \left(1, \frac{1}{\pi_a}, \ldots, \frac{1}{\pi_a}\right)'.\end{aligned}$$

Combing results, the asymptotic distribution $\sqrt{n}\left(\widehat{\mu}_{an}-\mathrm{E}\left[Y_{ai}\right]\right)$ is

$$
\begin{aligned}
&\mathcal{N}\left(c'0, c'\Sigma c\right)\\
=&\mathcal{N}\left(0, \operatorname{Var}\left(Y_{ai}\right)+\sum_{s=1}^{k} c_s^2 \Sigma_{s+1,s+1}\right)\\
=&\mathcal{N}\left(0, \operatorname{Var}\left(Y_{ai}\right)+\sum_{s=1}^{k} \frac{\operatorname{Var}\left(\mathbb{1}\{S_i=s\}\left(\widetilde{Y}_{ai}-\mathrm{E}\left[\widetilde{Y}_{ai}|S_i=s\right]\right)\left(\mathbb{R}_{ai}-\mathrm{E}\left[\mathbb{R}_{ai}|S_i=s\right]\right)\right)}{\pi_a^2}\right).
\end{aligned}
$$

Now we just need to rewrite the second term in the asymptotic variance expression. Note that

$$
\begin{aligned}
&\sum_{s=1}^{k} \operatorname{Var}\left(\mathbb{1}\{S_i=s\}\left(\widetilde{Y}_{ai}-\mathrm{E}\left[\widetilde{Y}_{ai}|S_i=s\right]\right)\left(\mathbb{R}_{ai}-\mathrm{E}\left[\mathbb{R}_{ai}|S_i=s\right]\right)\right)\\
=&\sum_{s=1}^{k} \mathrm{E}\left[\left(\mathbb{1}\{S_i=s\}\left(\widetilde{Y}_{ai}-\mathrm{E}\left[\widetilde{Y}_{ai}|S_i=s\right]\right)\left(\mathbb{R}_{ai}-\mathrm{E}\left[\mathbb{R}_{ai}|S_i=s\right]\right)\right)^2\right]\\
&-\left(\mathrm{E}\left[\mathbb{1}\{S_i=s\}\left(\widetilde{Y}_{ai}-\mathrm{E}\left[\widetilde{Y}_{ai}|S_i=s\right]\right)\left(\mathbb{R}_{ai}-\mathrm{E}\left[\mathbb{R}_{ai}|S_i=s\right]\right)\right]\right)^2\\
=&\sum_{s=1}^{k} \mathrm{E}\left[\left(\mathbb{1}\{S_i=s\}\left(\widetilde{Y}_{ai}-\mathrm{E}\left[\widetilde{Y}_{ai}|S_i=s\right]\right)\left(\mathbb{R}_{ai}-\mathrm{E}\left[\mathbb{R}_{ai}|S_i=s\right]\right)\right)^2\right]\\
=&\sum_{s=1}^{k} \mathrm{E}\left[\left(\mathbb{1}\{S_i=s\}\left(\widetilde{Y}_{ai}-\mathrm{E}\left[\widetilde{Y}_{ai}|S_i\right]\right)\left(\mathbb{R}_{ai}-\pi_a\right)\right)^2\right]\\
=&\sum_{s=1}^{k} \mathrm{E}\left[\mathbb{1}\{S_i=s\}\left(\left(\widetilde{Y}_{ai}-\mathrm{E}\left[\widetilde{Y}_{ai}|S_i\right]\right)\left(\mathbb{R}_{ai}-\pi_a\right)\right)^2\right] \quad \text{because } \mathbb{1}\{S_i=s\}\in\{0,1\}\\
=&\sum_{s=1}^{k} \mathrm{E}\left[\mathbb{1}\{S_i=s\}\right] \frac{\mathrm{E}\left[\mathbb{1}\{S_i=s\}\left(\left(\widetilde{Y}_{ai}-\mathrm{E}\left[\widetilde{Y}_{ai}|S_i\right]\right)\left(\mathbb{R}_{ai}-\pi_a\right)\right)^2\right]}{\mathrm{E}\left[\mathbb{1}\{S_i=s\}\right]}\\
=&\sum_{s=1}^{k} \mathrm{E}\left[\mathbb{1}\{S_i=s\}\right] \mathrm{E}\left[\left(\left(\widetilde{Y}_{ai}-\mathrm{E}\left[\widetilde{Y}_{ai}|S_i\right]\right)\left(\mathbb{R}_{ai}-\pi_a\right)\right)^2 \middle| S_i=s\right]\\
=&\sum_{s=1}^{k} \mathrm{E}\left[\mathbb{1}\{S_i=s\}\right] \operatorname{Var}\left(\left(\widetilde{Y}_{ai}-\mathrm{E}\left[\widetilde{Y}_{ai}|S_i\right]\right)\left(\mathbb{R}_{ai}-\pi_a\right) \middle| S_i=s\right)\\
=&\,\mathrm{E}\left[\operatorname{Var}\left(\left(\widetilde{Y}_{ai}-\mathrm{E}\left[\widetilde{Y}_{ai}|S_i\right]\right)\left(\mathbb{R}_{ai}-\pi_a\right) \middle| S_i\right)\right].
\end{aligned}
$$

Therefore, recalling that $\widetilde{Y}_{ai}=Y_{ai}-\gamma_a' X_i$, we have by substitution that

$$
\sqrt{n}\left(\widehat{\mu}_{an}-\mathrm{E}\left[Y_{ai}\right]\right) \xrightarrow{d} \mathcal{N}\left(0, V\left(\gamma_a\right)\right)
$$

as $n \to \infty$, where

$$V(\gamma_a) = \mathrm{Var}(Y_{ai}) + \frac{\mathrm{E}\left[\mathrm{Var}\left(\left(Y_{ai} - \gamma_a' X_i - \mathrm{E}\left[Y_{ai} - \gamma_a' X_i | S_i\right]\right)(\mathbb{R}_{ai} - \pi_a) | S_i\right)\right]}{\pi_a^2}.$$

We can rewrite the second term in the variance expression above. Note that

$$\begin{aligned}
\frac{\mathrm{E}\left[\mathrm{Var}\left(\left(\widetilde{Y}_{ai} - \mathrm{E}\left[\widetilde{Y}_{ai}|S_i\right]\right)(\mathbb{R}_{ai} - \pi_a)|S_i\right)\right]}{\pi_a^2} &= \frac{\mathrm{E}\left[\mathrm{Var}\left(\widetilde{Y}_{ai} - \mathrm{E}\left[\widetilde{Y}_{ai}|S_i\right]|S_i\right)\mathrm{Var}\left(\mathbb{R}_{ai} - \pi_a|S_i\right)\right]}{\pi_a^2} \\
&= \frac{\mathrm{E}\left[\mathrm{Var}\left(\widetilde{Y}_{ai}|S_i\right)\right]\mathrm{Var}\left(\mathbb{R}_{ai} - \pi_a\right)}{\pi_a^2} \\
&= \left(\frac{1-\pi_a}{\pi_a}\right)\mathrm{E}\left[\mathrm{Var}\left(\widetilde{Y}_{ai}|S_i\right)\right] \\
&= \left(\frac{1-\pi_a}{\pi_a}\right)\mathrm{E}\left[\mathrm{Var}\left(Y_{ai} - \gamma_a' X_i|S_i\right)\right]
\end{aligned}$$

For the first equality, we use the fact that $\left(\widetilde{Y}_{ai} - \mathrm{E}\left[\widetilde{Y}_{ai}|S_i\right]\right)$ and $(\mathbb{R}_{ai} - \pi_a)$ are conditionally independent given S_i and also have conditional mean zero given S_i. This completes the proof. □

Proof of Theorem 3.21. Using the result of Theorem 3.17, note that minimizing the asymptotic variance amounts to minimizing the quantity

$$\mathrm{E}\left[\mathrm{Var}\left(\left(\widetilde{Y}_{ai} - \mathrm{E}\left[\widetilde{Y}_{ai}|S_i\right]\right)\left(\mathbb{R}_{ai} - \mathrm{E}\left[\mathbb{R}_{ai}|S_i\right]\right)|S_i\right)\right].$$

By the Law of Total Variance

$$\begin{aligned}
&\mathrm{E}\left[\mathrm{Var}\left(\left(\widetilde{Y}_{ai} - \mathrm{E}\left[\widetilde{Y}_{ai}|S_i\right]\right)\left(\mathbb{R}_{ai} - \mathrm{E}\left[\mathbb{R}_{ai}|S_i\right]\right)|S_i\right)\right] \\
=&\mathrm{Var}\left(\left(\widetilde{Y}_{ai} - \mathrm{E}\left[\widetilde{Y}_{ai}|S_i\right]\right)\left(\mathbb{R}_{ai} - \mathrm{E}\left[\mathbb{R}_{ai}|S_i\right]\right)\right) - \mathrm{Var}\left(\mathrm{E}\left[\left(\widetilde{Y}_{ai} - \mathrm{E}\left[\widetilde{Y}_{ai}|S_i\right]\right)\left(\mathbb{R}_{ai} - \mathrm{E}\left[\mathbb{R}_{ai}|S_i\right]\right)|S_i\right]\right) \\
=&\mathrm{Var}\left(\left(\widetilde{Y}_{ai} - \mathrm{E}\left[\widetilde{Y}_{ai}|S_i\right]\right)(\mathbb{R}_{ai} - \pi_a)\right) \quad \text{because the second term is just zero} \\
=&\mathrm{Var}\left(\widetilde{Y}_{ai} - \mathrm{E}\left[\widetilde{Y}_{ai}|S_i\right]\right)\mathrm{Var}\left(\mathbb{R}_{ai} - \pi_a\right)
\end{aligned}$$

where the last step follows from the fact that $\left(\widetilde{Y}_{ai} - \mathrm{E}\left[\widetilde{Y}_{ai}|S_i\right]\right)$ and $(\mathbb{R}_{ai} - \pi_a)$ are independent and each have mean zero. Thus, our minimization problem further reduces to minimizing

$$
\begin{aligned}
& \mathrm{Var}\left(\widetilde{Y}_{ai} - \mathrm{E}\left[\widetilde{Y}_{ai}|S_i\right]\right) \\
= & \,\mathrm{E}\left[\mathrm{Var}\left(\widetilde{Y}_{ai} - \mathrm{E}\left[\widetilde{Y}_{ai}|S_i\right] \middle| S_i\right)\right] + \mathrm{Var}\left(\mathrm{E}\left[\widetilde{Y}_{ai} - \mathrm{E}\left[\widetilde{Y}_{ai}|S_i\right] \middle| S_i\right]\right) \quad \text{by LOTV} \\
= & \,\mathrm{E}\left[\mathrm{Var}\left(\widetilde{Y}_{ai}|S_i\right)\right] + 0 \\
= & \,\mathrm{E}\left[\mathrm{Var}\left(\widetilde{Y}_{ai}|S_i\right)\right] \\
= & \,\mathrm{E}\left[\mathrm{Var}\left(Y_{ai}|S_i\right) + \gamma_a' \mathrm{Var}\left(X_i|S_i\right)\gamma_a - 2\gamma_a' \,\mathrm{Cov}\left(X_i, Y_{ai}|S_i\right)\right] \quad \text{by the definition of } \widetilde{Y}_{ai} \\
= & \,\mathrm{E}\left[\mathrm{Var}\left(Y_{ai}|S_i\right)\right] + \gamma_a' \,\mathrm{E}\left[\mathrm{Var}\left(X_i|S_i\right)\right]\gamma_a - 2\gamma_a' \,\mathrm{E}\left[\mathrm{Cov}\left(X_i, Y_{ai}|S_i\right)\right]
\end{aligned}
$$

Taking the derivative with respect to γ_a and forming the FOC, we have

$$2\,\mathrm{E}\left[\mathrm{Var}\left(X_i|S_i\right)\right]\gamma_a^* - 2\,\mathrm{E}\left[\mathrm{Cov}\left(X_i, Y_{ai}|S_i\right)\right] = 0_p.$$

Thus,

$$\gamma_a^* = \mathrm{E}\left[\mathrm{Var}\left(X_i|S_i\right)\right]^{-1} \mathrm{E}\left[\mathrm{Cov}\left(X_i, Y_{ai}|S_i\right)\right].$$

Our proposed estimator for γ_a^* is

$$\widehat{\gamma}_{an}^* = \left(\mathrm{E}_n\left[X_i X_i'\right] - \mathrm{E}_n\left[\mathrm{E}_n\left[X_i|S_i\right]\mathrm{E}_n\left[X_i'|S_i\right]\right]\right)^{-1} \mathrm{E}_n\left[\frac{\mathrm{E}_n\left[X_i Y_{ai}\mathbb{R}_{ai}|S_i\right]}{\mathrm{E}_n\left[\mathbb{R}_{ai}|S_i\right]} - \mathrm{E}_n\left[X_i|S_i\right]\frac{\mathrm{E}_n\left[Y_{ai}\mathbb{R}_{ai}|S_i\right]}{\mathrm{E}_n\left[\mathbb{R}_{ai}|S_i\right]}\right].$$

We need to check that it is consistent. For the consistency argument, we proceed in chunks. We have

$$\mathrm{E}_n\left[X_i X_i'\right] \xrightarrow{p} \mathrm{E}\left[X_i X_i'\right]$$

as $n \to \infty$ by WLLNs. Moreover,

$$
\begin{aligned}
& \mathrm{E}_n\left[\mathrm{E}_n\left[X_i|S_i\right]\mathrm{E}_n\left[X_i'|S_i\right]\right] \\
= & \sum_{s=1}^{k} \mathrm{E}_n\left[\mathbb{1}\{S_i = s\}\right]\mathrm{E}_n\left[X_i|S_i = s\right]\mathrm{E}_n\left[X_i'|S_i = s\right] \\
= & \sum_{s=1}^{k} \mathrm{E}_n\left[\mathbb{1}\{S_i = s\}\right]\frac{\mathrm{E}_n\left[X_i\mathbb{1}\{S_i = s\}\right]}{\mathrm{E}_n\left[\mathbb{1}\{S_i = s\}\right]}\frac{\mathrm{E}_n\left[X_i'\mathbb{1}\{S_i = s\}\right]}{\mathrm{E}_n\left[\mathbb{1}\{S_i = s\}\right]} \\
\xrightarrow{p} & \sum_{s=1}^{k} \mathrm{E}\left[\mathbb{1}\{S_i = s\}\right]\frac{\mathrm{E}\left[X_i\mathbb{1}\{S_i = s\}\right]}{\mathrm{E}\left[\mathbb{1}\{S_i = s\}\right]}\frac{\mathrm{E}\left[X_i'\mathbb{1}\{S_i = s\}\right]}{\mathrm{E}\left[\mathbb{1}\{S_i = s\}\right]} \quad \text{as } n \to \infty \text{ by WLLNs and Slutsky's Theorem} \\
= & \sum_{s=1}^{k} \mathrm{E}\left[\mathbb{1}\{S_i = s\}\right]\mathrm{E}\left[X_i|S_i = s\right]\mathrm{E}\left[X_i'|S_i = s\right] \\
= & \,\mathrm{E}\left[\mathrm{E}\left[X_i|S_i\right]\mathrm{E}\left[X_i'|S_i\right]\right].
\end{aligned}
$$

Combining results using Slutsky's Theorem, we have

$$\begin{aligned}
&\left(\mathrm{E}_n\left[X_i X_i'\right]-\mathrm{E}_n\left[\mathrm{E}_n\left[X_i|S_i\right]\mathrm{E}_n X_i'|S_i\right]\right)^{-1}\\
\xrightarrow{p}\ &\left(\mathrm{E}\left[X_i X_i'\right]-\mathrm{E}\left[\mathrm{E}\left[X_i|S_i\right]\mathrm{E}\left[X_i'|S_i\right]\right]\right)^{-1} \quad \text{as } n\to\infty\\
=\ &\left(\mathrm{E}\left[\mathrm{E}\left[X_i X_i'|S_i\right]-\mathrm{E}\left[X_i|S_i\right]\mathrm{E}\left[X_i'|S_i\right]\right]\right)^{-1}\\
=\ &\left(\mathrm{E}\left[\mathrm{Var}\left(X_i|S_i\right)\right]\right)^{-1}.
\end{aligned}$$

Next, we have

$$\begin{aligned}
&\mathrm{E}_n\left[\frac{\mathrm{E}_n\left[X_i Y_{ai}\mathbb{R}_{ai}|S_i\right]}{\mathrm{E}_n\left[\mathbb{R}_{ai}|S_i\right]}-\mathrm{E}_n\left[X_i|S_i\right]\frac{\mathrm{E}_n\left[Y_{ai}\mathbb{R}_{ai}|S_i\right]}{\mathrm{E}_n\left[\mathbb{R}_{ai}|S_i\right]}\right]\\
=&\sum_{s=1}^{k}\mathrm{E}_n\left[\mathbb{1}\{S_i=s\}\right]\left(\frac{\mathrm{E}_n\left[X_i Y_{ai}\mathbb{R}_{ai}|S_i=s\right]}{\mathrm{E}_n\left[\mathbb{R}_{ai}|S_i=s\right]}-\mathrm{E}_n\left[X_i|S_i=s\right]\frac{\mathrm{E}_n\left[Y_{ai}\mathbb{R}_{ai}|S_i=s\right]}{\mathrm{E}_n\left[\mathbb{R}_{ai}|S_i=s\right]}\right)\\
=&\sum_{s=1}^{k}\mathrm{E}_n\left[\mathbb{1}\{S_i=s\}\right]\left(\frac{\mathrm{E}_n\left[\mathbb{1}\{S_i=s\}X_i Y_{ai}\mathbb{R}_{ai}\right]/\mathrm{E}_n\left[\mathbb{1}\{S_i=s\}\right]}{\mathrm{E}_n\left[\mathbb{R}_{ai}\mathbb{1}\{S_i=s\}\right]/\mathrm{E}_n\left[\mathbb{1}\{S_i=s\}\right]}\right.\\
&\left.-\frac{\mathrm{E}_n\left[X_i\mathbb{1}\{S_i=s\}\right]}{\mathrm{E}_n\left[\mathbb{1}\{S_i=s\}\right]}\cdot\frac{\mathrm{E}_n\left[Y_{ai}\mathbb{R}_{ai}\mathbb{1}\{S_i=s\}\right]/\mathrm{E}_n\left[\mathbb{1}\{S_i=s\}\right]}{\mathrm{E}_n\left[\mathbb{R}_{ai}\mathbb{1}\{S_i=s\}\right]/\mathrm{E}_n\left[\mathbb{1}\{S_i=s\}\right]}\right)\\
\xrightarrow{p}&\sum_{s=1}^{k}\mathrm{E}\left[\mathbb{1}\{S_i=s\}\right]\left(\frac{\mathrm{E}\left[\mathbb{1}\{S_i=s\}X_i Y_{ai}\mathbb{R}_{ai}\right]/\mathrm{E}\left[\mathbb{1}\{S_i=s\}\right]}{\mathrm{E}\left[\mathbb{R}_{ai}\mathbb{1}\{S_i=s\}\right]/\mathrm{E}\left[\mathbb{1}\{S_i=s\}\right]}\right.\\
&\left.-\frac{\mathrm{E}\left[X_i\mathbb{1}\{S_i=s\}\right]}{\mathrm{E}\left[\mathbb{1}\{S_i=s\}\right]}\cdot\frac{\mathrm{E}\left[Y_{ai}\mathbb{R}_{ai}\mathbb{1}\{S_i=s\}\right]/\mathrm{E}\left[\mathbb{1}\{S_i=s\}\right]}{\mathrm{E}\left[\mathbb{R}_{ai}\mathbb{1}\{S_i=s\}\right]/\mathrm{E}\left[\mathbb{1}\{S_i=s\}\right]}\right) \quad \text{by WLLN's + Slutsky's Theorem}\\
=&\sum_{s=1}^{k}\mathrm{E}\left[\mathbb{1}\{S_i=s\}\right]\left(\frac{\mathrm{E}\left[\mathbb{1}\{S_i=s\}X_i Y_{ai}\mathbb{R}_{ai}\right]/\mathrm{E}\left[\mathbb{1}\{S_i=s\}\right]}{\mathrm{E}\left[\mathbb{R}_{ai}\mathbb{1}\{S_i=s\}\right]/\mathrm{E}\left[\mathbb{1}\{S_i=s\}\right]}\right.\\
=&\sum_{s=1}^{k}\mathrm{E}\left[\mathbb{1}\{S_i=s\}\right]\left(\frac{\mathrm{E}\left[X_i Y_{ai}\mathbb{R}_{ai}|S_i=s\right]}{\mathrm{E}\left[\mathbb{R}_{ai}|S_i=s\right]}-\mathrm{E}\left[X_i|S_i=s\right]\frac{\mathrm{E}\left[Y_{ai}\mathbb{R}_{ai}|S_i=s\right]}{\mathrm{E}\left[\mathbb{R}_{ai}|S_i=s\right]}\right)\\
=&\,\mathrm{E}\left[\frac{\mathrm{E}\left[X_i Y_{ai}\mathbb{R}_{ai}|S_i\right]}{\mathrm{E}\left[\mathbb{R}_{ai}|S_i\right]}-\mathrm{E}\left[X_i|S_i\right]\frac{\mathrm{E}\left[Y_{ai}\mathbb{R}_{ai}|S_i\right]}{\mathrm{E}\left[\mathbb{R}_{ai}|S_i\right]}\right]\\
=&\,\mathrm{E}\left[\frac{\mathrm{E}\left[X_i Y_{ai}|S_i\right]\mathrm{E}\left[\mathbb{R}_{ai}|S_i\right]}{\mathrm{E}\left[\mathbb{R}_{ai}|S_i\right]}-\mathrm{E}\left[X_i|S_i\right]\frac{\mathrm{E}\left[Y_{ai}|S_i\right]\mathrm{E}\left[\mathbb{R}_{ai}|S_i\right]}{\mathrm{E}\left[\mathbb{R}_{ai}|S_i\right]}\right]\\
=&\,\mathrm{E}\left[\mathrm{E}\left[X_i Y_{ai}|S_i\right]-\mathrm{E}\left[X_i|S_i\right]\mathrm{E}\left[Y_{ai}|S_i\right]\right]\\
=&\,\mathrm{E}\left[\mathrm{Cov}\left(X_i, Y_{ai}|S_i\right)\right].
\end{aligned}$$

Combining results one more time gives

$$\begin{aligned}
\widehat{\gamma}_{an}^{*} &\xrightarrow{p} \left(\mathrm{E}\left[\mathrm{Var}\left(X_i|S_i\right)\right]\right)^{-1}\mathrm{E}\left[\mathrm{Cov}\left(X_i, Y_{ai}|S_i\right)\right] \quad \text{as } n\to\infty\\
&=\gamma_a^{*}
\end{aligned}$$

as desired. □

A.4 Proofs for Section 4

Proof of Theorem 4.1. By Theorem 3.3, we have

$$\sqrt{n}\left(\mathrm{E}\left[Y_{ai}\right]-\widehat{\mu}_{an}\right) \xrightarrow{d} \mathcal{N}\left(0, V\left(\gamma_a\right)\right).$$

where

$$\begin{aligned} V\left(\gamma_a\right) &= \operatorname{Var}\left(Y_{ai}\right)+\mathrm{E}\left[\left(\frac{1-\pi_{ai}}{\pi_{ai}}\right)\left(Y_{ai}-\gamma_a' X_i\right)^2\right] \\ &= \mathrm{E}\left[Y_{ai}^2\right]-\left(\mathrm{E}\left[Y_{ai}\right]\right)^2+\mathrm{E}\left[\left(\frac{1-\pi_{ai}}{\pi_{ai}}\right)\left(Y_{ai}^2-2\gamma_a' X_i Y_{ai}+\left(\gamma_a' X_i\right)^2\right)\right] \end{aligned}$$

and γ_a is the probability limit of $\widehat{\gamma}_{an}$. We just need to check that

$$\widehat{V}_n\left(\widehat{\gamma}_{an}\right)=\mathrm{E}_n\left[\frac{Y_{ai}^2 \mathbb{R}_{ai}}{\pi_{ai}}\right]-\left(\mathrm{E}_n\left[\frac{Y_{ai} \mathbb{R}_{ai}}{\pi_{ai}}\right]\right)^2+\mathrm{E}_n\left[\left(\frac{1-\pi_{ai}}{\pi_{ai}}\right)\left(\frac{\mathbb{R}_{ai} Y_{ai}^2}{\pi_{ai}}-2\widehat{\gamma}_{an}' X_i \frac{\mathbb{R}_{ai} Y_{ai}}{\pi_{ai}}+\left(\widehat{\gamma}_{an}' X_i\right)^2\right)\right]$$

is a consistent estimator for $V\left(\gamma_a\right)$. We proceed term by term. Note first that

$$\begin{aligned} \mathrm{E}_n\left[\frac{Y_{ai}^2 \mathbb{R}_{ai}}{\pi_{ai}}\right] &\xrightarrow{p} \mathrm{E}\left[\frac{Y_{ai}^2 \mathbb{R}_{ai}}{\pi_{ai}}\right] \quad \text{as } n \rightarrow \infty \text{ by WLLNs} \\ &= \mathrm{E}\left[\mathrm{E}\left[Y_{ai}^2 \mid X_i\right] \frac{\mathbb{R}_{ai}}{\pi_{ai}}\right] \quad \text{by LIE} \\ &= \mathrm{E}\left[Y_{ai}^2\right]. \end{aligned}$$

We will utilize the unbiasedness of similar IPW-style estimators for moments of Y_{ai} throughout the proof. Similarly,

$$\begin{aligned} \left(\mathrm{E}_n\left[\frac{Y_{ai} \mathbb{R}_{ai}}{\pi_{ai}}\right]\right)^2 &\xrightarrow{p} \left(\mathrm{E}\left[\frac{Y_{ai} \mathbb{R}_{ai}}{\pi_{ai}}\right]\right)^2 \quad \text{by WLLNs and Slutsky's Theorem} \\ &= \left(\mathrm{E}\left[Y_{ai}\right]\right)^2. \end{aligned}$$

Now we analyze the third term. We have

$$\begin{aligned} & \mathrm{E}_n\left[\left(\frac{1-\pi_{ai}}{\pi_{ai}}\right)\left(\frac{\mathbb{R}_{ai} Y_{ai}^2}{\pi_{ai}}-2\widehat{\gamma}_{an}' X_i \frac{\mathbb{R}_{ai} Y_{ai}}{\pi_{ai}}+\left(\widehat{\gamma}_{an}' X_i\right)^2\right)\right] \\ = & \mathrm{E}_n\left[\left(\frac{1-\pi_{ai}}{\pi_{ai}}\right)\left(\frac{\mathbb{R}_{ai} Y_{ai}^2}{\pi_{ai}}\right)\right]-2\widehat{\gamma}_{an}' \mathrm{E}_n\left[\left(\frac{1-\pi_{ai}}{\pi_{ai}}\right) \frac{\mathbb{R}_{ai} Y_{ai}}{\pi_{ai}}\right]+\widehat{\gamma}_{an}' \mathrm{E}_n\left[\left(\frac{1-\pi_{ai}}{\pi_{ai}}\right) X_i X_i'\right] \widehat{\gamma}_{an} \\ \xrightarrow{p} & \mathrm{E}\left[\left(\frac{1-\pi_{ai}}{\pi_{ai}}\right)\left(\frac{\mathbb{R}_{ai} Y_{ai}^2}{\pi_{ai}}\right)\right]-2\gamma_a' \mathrm{E}\left[\left(\frac{1-\pi_{ai}}{\pi_{ai}}\right) \frac{\mathbb{R}_{ai} Y_{ai}}{\pi_{ai}}\right]+\gamma_a' \mathrm{E}\left[\left(\frac{1-\pi_{ai}}{\pi_{ai}}\right) X_i X_i'\right] \gamma_a \\ = & \mathrm{E}\left[\left(\frac{1-\pi_{ai}}{\pi_{ai}}\right) Y_{ai}^2\right]-2\gamma_a' \mathrm{E}\left[\left(\frac{1-\pi_{ai}}{\pi_{ai}}\right) Y_{ai}\right]+\gamma_a' \mathrm{E}\left[\left(\frac{1-\pi_{ai}}{\pi_{ai}}\right) X_i X_i'\right] \gamma_a \\ = & \mathrm{E}\left[\left(\frac{1-\pi_{ai}}{\pi_{ai}}\right)\left(Y_{ai}^2-2\gamma_a' X_i Y_{ai}+\left(\gamma_a' X_i\right)^2\right)\right] \end{aligned}$$

The convergence in probability step follows from a combination of WLLNs and Slutsky's Theorem, as well as the consistency of $\widehat{\gamma}_{an}$ for γ_a. Combining the results for each of the three terms gives $\widehat{V}_n\left(\widehat{\gamma}_{an}\right) \xrightarrow{p} V\left(\gamma_a\right)$ as $n \to \infty$. Therefore, an asymptotically valid $1-\alpha$ CI for $\mathrm{E}\left[Y_{ai}\right]$ is

$$\left[\widehat{\mu}_{an} - z_{1-\alpha/2}\sqrt{\frac{\widehat{V}_n\left(\widehat{\gamma}_{an}\right)}{n}},\ \widehat{\mu}_{an} + z_{1-\alpha/2}\sqrt{\frac{\widehat{V}_n\left(\widehat{\gamma}_{an}\right)}{n}}\right].$$

as claimed. □

Proof of Theorem 4.2. From Theorem 3.14, we have

$$\sqrt{n}\left(\mathrm{E}\left[Y_{ai}\right] - \widehat{\mu}_{an}\right) \xrightarrow{d} \mathcal{N}\left(0, V\left(\gamma_a\right)\right)$$

where

$$\begin{aligned} V\left(\gamma_a\right) &= \mathrm{Var}\left(Y_{ai}\right) + \left(\frac{1-\pi_a}{\pi_a}\right)\mathrm{Var}\left(Y_{ai} - \gamma_a' X_i\right) \\ &= \mathrm{Var}\left(Y_{ai}\right) + \left(\frac{1-\pi_a}{\pi_a}\right)\left(\mathrm{Var}\left(Y_{ai}\right) + \gamma_a' \mathrm{Var}\left(X_i\right)\gamma_a + 2\gamma_a' \,\mathrm{Cov}\left(X_i, Y_{ai}\right)\right) \end{aligned}$$

and γ_a is the probability limit of $\widehat{\gamma}_{an}$. We just need to check that

$$\widehat{V}_n\left(\widehat{\gamma}_{an}\right) \xrightarrow{p} V\left(\gamma_a\right)$$

as $n \to \infty$. Recall from the theorem statement that

$$\widehat{V}_n\left(\widehat{\gamma}_{an}\right) = A_n + \left(\frac{1-\pi_a}{\pi_a}\right)\left(A_n + B_n + 2C_n\right)$$

with

$$\begin{aligned} A_n &= \mathrm{E}_n\left[\frac{\mathbb{R}_{ai} Y_{ai}^2}{\pi_a}\right] - \left(\mathrm{E}_n\left[\frac{\mathbb{R}_{ai} Y_{ai}}{\pi_a}\right]\right)^2 \\ B_n &= \widehat{\gamma}_{an}' \,\mathrm{E}_n\left[\left(X_i - \mathrm{E}_n\left[X_i\right]\right)\left(X_i - \mathrm{E}_n\left[X_i\right]\right)'\right]\widehat{\gamma}_{an} \\ C_n &= \widehat{\gamma}_{an}'\left(\mathrm{E}_n\left[X_i \frac{Y_{ai}\mathbb{R}_{ai}}{\pi_a}\right] - \mathrm{E}_n\left[X_i\right]\mathrm{E}_n\left[\frac{Y_{ai}\mathbb{R}_{ai}}{\pi}\right]\right). \end{aligned}$$

First, we have

$$\begin{aligned} A_n &\xrightarrow{p} \mathrm{E}\left[\frac{\mathbb{R}_{ai} Y_{ai}^2}{\pi_a}\right] - \left(\mathrm{E}\left[\frac{\mathbb{R}_{ai} Y_{ai}}{\pi_a}\right]\right)^2 \quad \text{by WLLNs and Slutsky's Theorem} \\ &= \frac{\mathrm{E}\left[\mathbb{R}_{ai}\right]\mathrm{E}\left[Y_{ai}^2\right]}{\pi_a} - \left(\frac{\mathrm{E}\left[\mathbb{R}_{ai}\right]\mathrm{E}\left[Y_{ai}\right]}{\pi_a}\right)^2 \\ &= \mathrm{E}\left[Y_{ai}^2\right] - \left(\mathrm{E}\left[Y_{ai}\right]\right)^2 \\ &= \mathrm{Var}\left(Y_{ai}\right) \end{aligned}$$

as desired. Next, we have

$$\begin{aligned}B_n &= \widehat{\gamma}_{an}' \left(\mathrm{E}_n \left[X_i X_i'\right] - \mathrm{E}_n \left[X_i\right] \mathrm{E}_n \left[X_i'\right]\right) \widehat{\gamma}_{an} \\ &\xrightarrow{p} \gamma_a' \left(\mathrm{E} \left[X_i X_i'\right] - \mathrm{E} \left[X_i\right] \mathrm{E} \left[X_i'\right]\right) \gamma_a \\ &= \gamma_a' \mathrm{Var} \left(X_i\right) \gamma_a\end{aligned}$$

where in the second step we use WLLNs, Slutsky's Theorem, and the consistency of $\widehat{\gamma}_{an}$. Finally, we have

$$\begin{aligned}C_n &\xrightarrow{p} \gamma_a' \left(\mathrm{E}\left[\frac{X_i Y_{ai} \mathbb{R}_{ai}}{\pi_a}\right] - \mathrm{E}\left[X_i\right] \mathrm{E}\left[\frac{Y_{ai} \mathbb{R}_{ai}}{\pi_a}\right]\right) \quad \text{as } n \to \infty \\ &= \gamma_a' \left(\mathrm{E}\left[X_i Y_{ai}\right] - \mathrm{E}\left[X_i\right] \mathrm{E}\left[Y_{ai}\right]\right) \\ &= \gamma_a' \,\mathrm{Cov}\left(X_i, Y_{ai}\right).\end{aligned}$$

Combining results, we have,

$$\begin{aligned}\widehat{V}_n\left(\widehat{\gamma}_{an}\right) &\xrightarrow{p} \mathrm{Var}\left(Y_{ai}\right) + \left(\frac{1-\pi_a}{\pi_a}\right)\left(\mathrm{Var}\left(Y_{ai}\right) + \gamma_a' \mathrm{Var}\left(X_i\right)\gamma_a + 2\gamma_a' \,\mathrm{Cov}\left(X_i, Y_{ai}\right)\right) \\ &= V\left(\gamma_a\right)\end{aligned}$$

as desired. Thus, an asymptotically valid $1-\alpha$ CI for $\mathrm{E}\left[Y_{ai}\right]$ is

$$\left[\widehat{\mu}_{an} - z_{1-\alpha/2}\sqrt{\frac{\widehat{V}_n\left(\widehat{\gamma}_{an}\right)}{n}},\ \widehat{\mu}_{an} + z_{1-\alpha/2}\sqrt{\frac{\widehat{V}_n\left(\widehat{\gamma}_{an}\right)}{n}}\right].$$

as claimed. □

Proof of Theorem 4.3. We will give a proof sketch rather than the complete proof. Recall that

$$\sqrt{n}\left(\widehat{\mu}_{an} - \mathrm{E}\left[Y_{ai}\right]\right) \xrightarrow{d} \mathcal{N}\left(0, V\left(\gamma_a\right)\right)$$

as $n \to \infty$, where

$$\begin{aligned}V(\gamma_a) &= \mathrm{Var}\left(Y_{ai}\right) + \left(\frac{1-\pi_a}{\pi_a}\right)\mathrm{E}\left[\mathrm{Var}\left(Y_{ai} - \gamma_a' X_i | S_i\right)\right] \\ &= \mathrm{Var}\left(Y_{ai}\right) + \left(\frac{1-\pi_a}{\pi_a}\right)\left(\mathrm{E}\left[\mathrm{Var}\left(Y_{ai}|S_i\right)\right] + \gamma_a' \,\mathrm{E}\left[\mathrm{Var}\left(X_i|S_i\right)\right]\gamma_a + 2\gamma_a' \,\mathrm{E}\left[\mathrm{Cov}\left(X_i, Y_{ai}|S_i\right)\right]\right).\end{aligned}$$

Recall in our that statement that

$$\widehat{V}_n\left(\widehat{\gamma}_{an}\right) = A_n + \left(\frac{1-\pi_a}{\pi_a}\right)\left(B_n + C_n + 2D_n\right)$$

with

$$\begin{aligned}
A_n &= \mathrm{E}_n\left[\frac{\mathbb{R}_{ai}Y_{ai}^2}{\pi_a}\right] - \left(\mathrm{E}_n\left[\frac{\mathbb{R}_{ai}Y_{ai}}{\pi_a}\right]\right)^2 \\
B_n &= \mathrm{E}_n\left[\mathrm{E}_n\left[\frac{Y_{ai}^2\mathbb{R}_{ai}}{\pi_a}\middle|S_i\right] - \left(\mathrm{E}_n\left[\frac{Y_{ai}\mathbb{R}_{ai}}{\pi_a}\middle|S_i\right]\right)^2\right] \\
C_n &= \widehat{\gamma}_{an}'\,\mathrm{E}_n\left[\mathrm{E}_n\left[X_iX_i'|S_i\right] - \mathrm{E}_n\left[X_i|S_i\right]\mathrm{E}_n\left[X_i'|S_i\right]\right]\widehat{\gamma}_{an} \\
D_n &= \widehat{\gamma}_{an}\,\mathrm{E}_n\left[\mathrm{E}_n\left[X_i\frac{Y_{ai}\mathbb{R}_{ai}}{\pi_a}\middle|S_i\right] - \mathrm{E}_n\left[X_i|S_i\right]\mathrm{E}_n\left[\frac{Y_{ai}\mathbb{R}_{ai}}{\pi_a}\middle|S_i\right]\right].
\end{aligned}$$

Analogously to the proof of Theorem 4.2, we can show that

$$\begin{aligned}
A_n &\xrightarrow{p} \mathrm{Var}\left(Y_{ai}\right) \\
B_n &\xrightarrow{p} \mathrm{E}\left[\mathrm{Var}\left(Y_{ai}|S_i\right)\right] \\
C_n &\xrightarrow{p} \gamma_a'\,\mathrm{E}\left[\mathrm{Var}\left(X_i|S_i\right)\right]\gamma_a \\
D_n &\xrightarrow{p} \gamma_a'\,\mathrm{E}\left[\mathrm{Cov}\left(X_i, Y_{ai}|S_i\right)\right]
\end{aligned}$$

as $n \to \infty$. Combining results gives $\widehat{V}_n\left(\widehat{\gamma}_{an}\right) \xrightarrow{p} V\left(\gamma_a\right)$ as $n \to \infty$. Thus, an asymptotically valid $1-\alpha$ CI for $\mathrm{E}\left[Y_{ai}\right]$ is

$$\left[\widehat{\mu}_{an} - z_{1-\alpha/2}\sqrt{\frac{\widehat{V}_n\left(\widehat{\gamma}_{an}\right)}{n}},\ \widehat{\mu}_{an} + z_{1-\alpha/2}\sqrt{\frac{\widehat{V}_n\left(\widehat{\gamma}_{an}\right)}{n}}\right]$$

as desired. □

B Remark on Arbitrary Contrast in Section 3.1

We proceed with assumptions from Section 3.1, except that we now introduce some additional notation.

B.1 Additional Notation

We introduce some special notation. Let

$$\begin{aligned}
\mu &= \mathrm{E}[Y_i] \\
&= \mathrm{E}[(Y_{1i}, Y_{2i}, \ldots, Y_{ki})'].
\end{aligned}$$

Then let

$$\boldsymbol{\pi}_i = \begin{pmatrix} \mathrm{E}\left[\mathbb{R}_{1i}|x_i\right] & & & \\ & \mathrm{E}\left[\mathbb{R}_{2i}|x_i\right] & & \\ & & \ddots & \\ & & & \mathrm{E}\left[\mathbb{R}_{ki}|x_i\right] \end{pmatrix}$$

$$= \begin{pmatrix} \pi_{1i} & & & \\ & \pi_{2i} & & \\ & & \ddots & \\ & & & \pi_{ki} \end{pmatrix} \in \mathbf{R}^{k\times k}.$$

We assume that $\boldsymbol{\pi}_i$ is known function of x_i. Moreover, let $\mathbb{R}_i$ be defined as

$$\mathbb{R}_i = \begin{pmatrix} \mathbb{R}_{1i} & & & \\ & \mathbb{R}_{2i} & & \\ & & \ddots & \\ & & & \mathbb{R}_{ki} \end{pmatrix} \in \mathbf{R}^{k\times k}.$$

LIE refers to the law of iterated expectations. $\mathbf{I}_k$ is the $k \times k$ identity matrix.

B.2 Simple Estimator

A simple estimator for μ to consider is

$$\widehat{\mu}_n = n^{-1} \sum_{i=1}^{n} \mathbb{R}_i \boldsymbol{\pi}_i^{-1} Y_i$$

We can check that this estimator is unbiased. Let $\widehat{\mu}_{n_a}$ denote the a-th element of $\widehat{\mu}_n$. Then we have

$$
\begin{aligned}
\mathrm{E}[\widehat{\mu}_{n_a}] &= \mathrm{E}\left[n^{-1}\sum_{i=1}^{n}\frac{Y_{ai}\mathbb{R}_{ai}}{\mathrm{E}\left[\mathbb{R}_{ai}|x_i\right]}\right] \\
&= \mathrm{E}\left[\frac{Y_{ai}\mathbb{R}_{ai}}{\mathrm{E}\left[\mathbb{R}_{ai}|x_i\right]}\right] \quad \text{by identically distributed assumption} \\
&= \mathrm{E}\left[\mathrm{E}\left[\frac{Y_{ai}\mathbb{R}_{ai}}{\mathrm{E}\left[\mathbb{R}_{ai}|x_i\right]}\middle| x_i\right]\right] \quad \text{by LIE} \\
&= \mathrm{E}\left[\frac{\mathrm{E}(Y_{ai}\mathbb{R}_{ai}|x_i)}{\mathrm{E}\left[\mathbb{R}_{ai}|x_i\right]}\right] \\
&= \mathrm{E}\left[\frac{\mathrm{E}\left[Y_{ai}|x_i\right]\mathrm{E}\left[\mathbb{R}_{ai}|x_i\right]}{\mathrm{E}\left[\mathbb{R}_{ai}|x_i\right]}\right] \quad \text{by conditional independence assumption} \\
&= \mathrm{E}\left[\mathrm{E}\left[Y_{ai}|x_i\right]\right] \\
&= \mathrm{E}\left[Y_{ai}\right] \quad \text{by LIE} \\
&= \mathrm{E}\left[Y_{ai}\right]
\end{aligned}
$$

for each $a = 1, \ldots, k$. Thus, $\mathrm{E}\left[\widehat{\mu}_n\right] = \mu$.

Under the assumption that $\Sigma = \mathrm{Var}\left[\mathbb{R}_i \boldsymbol{\pi}_i^{-1} Y_i\right]$ exists, we have by the CLT that

$$
\sqrt{n}(\widehat{\mu}_n - \mu) \xrightarrow{d} \mathcal{N}\left(0, \Sigma\right)
$$

as $n \to \infty$.

B.3 Asymptotic Variance with Covariate Adjustment

A covariate-adjusted estimator to consider is

$$
\widehat{\mu}_n^N = n^{-1}\sum_{i=1}^{n}\left[\mathbb{R}_i \boldsymbol{\pi}_i^{-1} Y_i - \left(\mathbb{R}_i \boldsymbol{\pi}_i^{-1} - \mathbf{I}_k\right)\mathbf{N}x_i\right]
$$

where

$$
\mathbf{N} = \begin{pmatrix} N_{11} & N_{12} & \ldots & N_{1p} \\ N_{21} & N_{22} & \ldots & N_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ N_{k1} & N_{k2} & \ldots & N_{kp} \end{pmatrix} \in \mathbf{R}^{k\times p}
$$

is the covariate adjustment coefficient matrix. The expanded form of $\widehat{\mu}_n^N$ is

$$\widehat{\mu}_n^N = n^{-1}\sum_{i=1}^{n}\begin{bmatrix} \frac{Y_{1i}\mathbb{R}_{1i}}{\pi_{1i}} - \left(\frac{\mathbb{R}_{1i}}{\pi_{1i}} - 1\right)\sum_{j=1}^{p} N_{1j}x_{ji} \\ \frac{Y_{2i}\mathbb{R}_{2i}}{\pi_{2i}} - \left(\frac{\mathbb{R}_{2i}}{\pi_{2i}} - 1\right)\sum_{j=1}^{p} N_{2j}x_{ji} \\ \vdots \\ \frac{Y_{ki}\mathbb{R}_{ki}}{\pi_{ki}} - \left(\frac{\mathbb{R}_{ki}}{\pi_{ki}} - 1\right)\sum_{j=1}^{p} N_{kj}x_{ji} \end{bmatrix}$$

$$= n^{-1}\sum_{i=1}^{n}\begin{bmatrix} \frac{Y_{1i}\mathbb{R}_{1i}}{\pi_{1i}} - N_{11}\left(\frac{x_{1i}\mathbb{R}_{1i}}{\pi_{1i}} - x_{1i}\right) - N_{12}\left(\frac{x_{2i}\mathbb{R}_{1i}}{\pi_{1i}} - x_{2i}\right) - \ldots - N_{1p}\left(\frac{x_{pi}\mathbb{R}_{1i}}{\pi_{1i}} - x_{pi}\right) \\ \frac{Y_{2i}\mathbb{R}_{2i}}{\pi_{2i}} - N_{21}\left(\frac{x_{1i}\mathbb{R}_{2i}}{\pi_{2i}} - x_{1i}\right) - N_{22}\left(\frac{x_{2i}\mathbb{R}_{2i}}{\pi_{2i}} - x_{2i}\right) - \ldots - N_{2p}\left(\frac{x_{pi}\mathbb{R}_{2i}}{\pi_{2i}} - x_{pi}\right) \\ \vdots \\ \frac{Y_{ki}\mathbb{R}_{ki}}{\pi_{ki}} - N_{k1}\left(\frac{x_{1i}\mathbb{R}_{ki}}{\pi_{ki}} - x_{1i}\right) - N_{k2}\left(\frac{x_{2i}\mathbb{R}_{ki}}{\pi_{ki}} - x_{2i}\right) - \ldots - N_{kp}\left(\frac{x_{pi}\mathbb{R}_{ki}}{\pi_{ki}} - x_{pi}\right) \end{bmatrix}$$

Note that for each $a = 1, \ldots, k$ and $j = 1, \ldots, p$,

$$\begin{aligned} \mathrm{E}\left(\frac{x_{ji}\mathbb{R}_{ai}}{\pi_{ai}}\right) &= \mathrm{E}\left(\mathrm{E}\left(\frac{x_{ji}\mathbb{R}_{ai}}{\pi_{ai}}\middle| x_i\right)\right) \quad \text{by LIE} \\ &= \mathrm{E}\left(\frac{x_{ji}\,\mathrm{E}\left(\mathbb{R}_{ai}|x_i\right)}{\mathrm{E}\left(\mathbb{R}_{ai}|x_i\right)}\right) \\ &= \mathrm{E}(x_{ji}). \end{aligned}$$

Thus, we have that

$$\begin{aligned} \mathrm{E}\left(\widehat{\mu}_n^N\right) &= \mathrm{E}\left(\widehat{\mu}_n\right) \\ &= \mu \end{aligned}$$

for any choice of $\mathbf{N}$. That is, the covariate-adjusted estimator is unbiased for μ regardless of our our choice of $\mathbf{N}$.

To find the asymptotic distribution of $\sqrt{n}(\widehat{\mu}_n^N - \mu)$, we apply the CLT. We have

$$\sqrt{n}(\widehat{\mu}_n^N - \mu) \xrightarrow{d} \mathcal{N}(0, \Sigma^N)$$

where

$$\Sigma^N = \mathrm{Var}\left[\mathbb{R}_i\boldsymbol{\pi}_i^{-1}Y_i - \left(\mathbb{R}_i\boldsymbol{\pi}_i^{-1} - \mathbf{I}_k\right)\mathbf{N}x_i\right].$$

B.4 Optimal Covariate Adjustment given Contrast Vector

Suppose that instead of doing inference on μ itself, we want to do inference on $c'\mu$, where $c \in \mathbf{R}^k$. We call c the contrast vector. This kind of situation would arise if we want to, say, compare two treatment arms. We estimate $c'\mu$ using $c'\widehat{\mu}_n^N$. It follows from the result in the previous section that $\sqrt{n}(c'\widehat{\mu}_n^N - c'\mu) \xrightarrow{d} \mathcal{N}(0, c'\Sigma^N c)$. Since we want to minimize the asymptotic variance of our estimator given a choice of contrast vector c, we want to find the choice of N which minimizes $c'\Sigma^N c$.

To solve the aforementioned optimization problem, let's first expand Σ^N:

$$\Sigma^N = \text{Var}\begin{bmatrix} \frac{Y_{1i}\mathbb{R}_{1i}}{\pi_{1i}} - \left(\frac{\mathbb{R}_{1i}}{\pi_{1i}} - 1\right)\sum_{j=1}^p N_{1j}x_{ji} \\ \frac{Y_{2i}\mathbb{R}_{2i}}{\pi_{2i}} - \left(\frac{\mathbb{R}_{2i}}{\pi_{2i}} - 1\right)\sum_{j=1}^p N_{2j}x_{ji} \\ \vdots \\ \frac{Y_{ki}\mathbb{R}_{ki}}{\pi_{ki}} - \left(\frac{\mathbb{R}_{ki}}{\pi_{ki}} - 1\right)\sum_{j=1}^p N_{kj}x_{ji} \end{bmatrix}$$

It follows that

$$\begin{aligned}
\Sigma_{r,s}^N &= \text{Cov}\left(\frac{Y_{ri}\mathbb{R}_{ri}}{\pi_{ri}} - \left(\frac{\mathbb{R}_{ri}}{\pi_{ri}} - 1\right)\sum_{j=1}^p N_{rj}x_{ji},\ \frac{Y_{si}\mathbb{R}_{si}}{\pi_{si}} - \left(\frac{\mathbb{R}_{si}}{\pi_{si}} - 1\right)\sum_{j=1}^p N_{sj}x_{ji}\right) \\
&= \text{Cov}\left(\frac{Y_{ri}\mathbb{R}_{ri}}{\pi_{ri}}, \frac{Y_{si}\mathbb{R}_{si}}{\pi_{si}}\right) - \text{Cov}\left(\frac{Y_{ri}\mathbb{R}_{ri}}{\pi_{ri}},\ \left(\frac{\mathbb{R}_{si}}{\pi_{si}} - 1\right)\sum_{j=1}^p N_{sj}x_{ji}\right) \\
&\quad - \text{Cov}\left(\left(\frac{\mathbb{R}_{ri}}{\pi_{ri}} - 1\right)\sum_{j=1}^p N_{rj}x_{ji},\ \frac{Y_{si}\mathbb{R}_{si}}{\pi_{si}}\right) \\
&\quad + \text{Cov}\left(\left(\frac{\mathbb{R}_{ri}}{\pi_{ri}} - 1\right)\sum_{j=1}^p N_{rj}x_{ji}, \left(\frac{\mathbb{R}_{si}}{\pi_{si}} - 1\right)\sum_{j=1}^p N_{sj}x_{ji}\right) \\
&= \text{Cov}\left(\frac{Y_{ri}\mathbb{R}_{ri}}{\pi_{ri}}, \frac{Y_{si}\mathbb{R}_{si}}{\pi_{si}}\right) - \sum_{j=1}^p N_{sj}\,\text{Cov}\left(\frac{Y_{ri}\mathbb{R}_{ri}}{\pi_{ri}}, \left(\frac{\mathbb{R}_{si}}{\pi_{si}} - 1\right)x_{ji}\right) \\
&\quad - \sum_{j=1}^p N_{rj}\,\text{Cov}\left(\frac{Y_{si}\mathbb{R}_{si}}{\pi_{si}}, \left(\frac{\mathbb{R}_{ri}}{\pi_{ri}} - 1\right)x_{ji}\right) \\
&\quad + \sum_{j=1}^p\sum_{q=1}^p N_{rj}N_{sq}\,\text{Cov}\left(\left(\frac{\mathbb{R}_{ri}}{\pi_{ri}} - 1\right)x_{ji}, \left(\frac{\mathbb{R}_{si}}{\pi_{si}} - 1\right)x_{qi}\right).
\end{aligned}$$

Then

$$
\begin{aligned}
c'\Sigma^N c &= \sum_{r=1}^{k}\sum_{s=1}^{k} c_r c_s \Sigma^N_{r,s} \\
&= \sum_{r=1}^{k}\sum_{s=1}^{k} c_r c_s \operatorname{Cov}\left(\frac{Y_{ri}\mathbb{R}_{ri}}{\pi_{ri}}, \frac{Y_{si}\mathbb{R}_{si}}{\pi_{si}}\right) - 2\sum_{r=1}^{k}\sum_{s=1}^{k} c_r c_s \sum_{j=1}^{p} N_{sj} \operatorname{Cov}\left(\frac{Y_{ri}\mathbb{R}_{ri}}{\pi_{ri}}, \left(\frac{\mathbb{R}_{si}}{\pi_{si}} - 1\right) x_{ji}\right) \\
&\quad + \sum_{r=1}^{k}\sum_{s=1}^{k} c_r c_s \sum_{j=1}^{p}\sum_{q=1}^{p} N_{rj} N_{sq} \operatorname{Cov}\left(\left(\frac{\mathbb{R}_{ri}}{\pi_{ri}} - 1\right) x_{ji}, \left(\frac{\mathbb{R}_{si}}{\pi_{si}} - 1\right) x_{qi}\right)
\end{aligned}
$$

We need to find N^* such that

$$
\frac{\partial}{\partial N} \sum_{r=1}^{k}\sum_{s=1}^{k} c_r c_s \Sigma^N_{r,s}
$$

evaluated at N^* is equal to $0_{p\times k}$. Note that for fixed choices $a \in \{1, 2, \ldots, k\}$ and $t \in \{1, 2, \ldots, p\}$, we have

$$
\begin{aligned}
\frac{\partial}{\partial N_{at}} \sum_{r=1}^{k}\sum_{s=1}^{k} c_r c_s \Sigma^N_{r,s} &= 0 - 2\sum_{r=1}^{k} c_r c_a \operatorname{Cov}\left(\frac{Y_{ri}\mathbb{R}_{ri}}{\pi_{ri}}, \left(\frac{\mathbb{R}_{ai}}{\pi_{ai}} - 1\right) x_{ti}\right) \\
&\quad + 2c_a^2 N_{at} \operatorname{Cov}\left(\left(\frac{\mathbb{R}_{ai}}{\pi_{ai}} - 1\right) x_{ti}, \left(\frac{\mathbb{R}_{ai}}{\pi_{ai}} - 1\right) x_{ti}\right) \\
&\quad + \sum_{s=1}^{k}\sum_{q=1}^{p}{}_{(s,q)\neq(a,t)} c_a c_s N_{sq} \operatorname{Cov}\left(\left(\frac{\mathbb{R}_{ai}}{\pi_{ai}} - 1\right) x_{ti}, \left(\frac{\mathbb{R}_{si}}{\pi_{si}} - 1\right) x_{qi}\right) \\
&\quad + \sum_{r=1}^{k}\sum_{j=1}^{p}{}_{(r,j)\neq(a,t)} c_r c_a N_{rj} \operatorname{Cov}\left(\left(\frac{\mathbb{R}_{ri}}{\pi_{ri}} - 1\right) x_{ji}, \left(\frac{\mathbb{R}_{ai}}{\pi_{ai}} - 1\right) x_{ti}\right) \\
&= -2\sum_{r=1}^{k} c_r c_a \operatorname{Cov}\left(\frac{Y_{ri}\mathbb{R}_{ri}}{\pi_{ri}}, \left(\frac{\mathbb{R}_{ai}}{\pi_{ai}} - 1\right) x_{ti}\right) \\
&\quad + 2\sum_{r=1}^{k}\sum_{j=1}^{p} c_r c_a N_{rj} \operatorname{Cov}\left(\left(\frac{\mathbb{R}_{ri}}{\pi_{ri}} - 1\right) x_{ji}, \left(\frac{\mathbb{R}_{ai}}{\pi_{ai}} - 1\right) x_{ti}\right)
\end{aligned}
$$

Therefore, the FOC corresponding to N_{at}, where $a \in \{1, \ldots, k\}$ and $t \in \{1, \ldots, p\}$, is

$$
\sum_{r=1}^{k}\sum_{j=1}^{p} c_r c_a N_{rj} \operatorname{Cov}\left(\left(\frac{\mathbb{R}_{ri}}{\pi_{ri}} - 1\right) x_{ji}, \left(\frac{\mathbb{R}_{ai}}{\pi_{ai}} - 1\right) x_{ti}\right) = \sum_{r=1}^{k} c_r c_a \operatorname{Cov}\left(\frac{Y_{ri}\mathbb{R}_{ri}}{\pi_{ri}}, \left(\frac{\mathbb{R}_{ai}}{\pi_{ai}} - 1\right) x_{ti}\right).
$$

Let's simplify the covariance expressions. Noting that

$$\begin{aligned}
\mathrm{E}\left[\left(\frac{\mathbb{R}_{ri}}{\pi_{ri}}-1\right)x_{ji}\right] &= \mathrm{E}\left[\mathrm{E}\left[\left(\frac{\mathbb{R}_{ri}}{\pi_{ri}}-1\right)x_{ji}\middle|x_i\right]\right] \\
&= \mathrm{E}\left[x_{ji}\,\mathrm{E}\left[\frac{\mathbb{R}_{ri}}{\pi_{ri}}-1\middle|x_i\right]\right] \\
&= \mathrm{E}\left[x_{ji}\left(\frac{\pi_{ri}}{\pi_{ri}}-1\right)\right] \\
&= 0,
\end{aligned}$$

we have

$$\begin{aligned}
\mathrm{Cov}\left(\left(\frac{\mathbb{R}_{ri}}{\pi_{ri}}-1\right)x_{ji},\left(\frac{\mathbb{R}_{ai}}{\pi_{ai}}-1\right)x_{ti}\right) &= \mathrm{E}\left[\left(\frac{\mathbb{R}_{ri}}{\pi_{ri}}-1\right)\left(\frac{\mathbb{R}_{ai}}{\pi_{ai}}-1\right)x_{ji}x_{ti}\right] \\
&= \mathrm{E}\left[\left(\frac{\mathbb{R}_{ri}}{\pi_{ri}}\right)\left(\frac{\mathbb{R}_{ai}}{\pi_{ai}}\right)x_{ji}x_{ti}\right] - \mathrm{E}\left[\left(\frac{\mathbb{R}_{ri}}{\pi_{ri}}-1\right)x_{ji}x_{ti}\right] \\
&\quad - \mathrm{E}\left[\left(\frac{\mathbb{R}_{ai}}{\pi_{ai}}-1\right)x_{ji}x_{ti}\right] + \mathrm{E}\left[x_{ji}x_{ti}\right] \\
&= \mathrm{E}\left[\left(\frac{\mathbb{R}_{ri}}{\pi_{ri}}\right)\left(\frac{\mathbb{R}_{ai}}{\pi_{ai}}\right)x_{ji}x_{ti}\right] + \mathrm{E}\left[x_{ji}x_{ti}\right].
\end{aligned}$$

When $r \neq a$,

$$\mathrm{E}\left[\left(\frac{\mathbb{R}_{ri}}{\pi_{ri}}\right)\left(\frac{\mathbb{R}_{ai}}{\pi_{ai}}\right)x_{ji}x_{ti}\right] = 0$$

because each unit is assigned to exactly one treatment by assumption and thus $\mathbb{R}_{ri}\mathbb{R}_{ai} = 0$. When $r = a$, we have

$$\begin{aligned}
\mathrm{E}\left[\left(\frac{\mathbb{R}_{ri}}{\pi_{ri}}\right)\left(\frac{\mathbb{R}_{ai}}{\pi_{ai}}\right)x_{ji}x_{ti}\right] &= \mathrm{E}\left[\frac{\mathbb{R}_{ri}^2}{\pi_{ri}^2}x_{ji}x_{ti}\right] \\
&= \mathrm{E}\left[\frac{\mathbb{R}_{ri}}{\pi_{ri}^2}x_{ji}x_{ti}\right] \quad \text{because } \mathbb{R}_{ri} \in \{0,1\} \\
&= \mathrm{E}\left[\mathrm{E}\left[\frac{\mathbb{R}_{ri}}{\pi_{ri}^2}x_{ji}x_{ti}\middle|x_i\right]\right] \\
&= \mathrm{E}\left[\frac{1}{\pi_{ri}}x_{ji}x_{ti}\right] = \mathrm{E}\left[\frac{1}{\pi_{ai}}x_{ji}x_{ti}\right]
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\sum_{r=1}^{k}\sum_{j=1}^{p} c_r c_a N_{rj}\,\mathrm{Cov}\left(\left(\frac{\mathbb{R}_{ri}}{\pi_{ri}}-1\right)x_{ji},\left(\frac{\mathbb{R}_{ai}}{\pi_{ai}}-1\right)x_{ti}\right) \\
&= \sum_{j=1}^{p} c_a^2 N_{aj}\,\mathrm{E}\left[\frac{1}{\pi_{ai}}x_{ji}x_{ti}\right] + \sum_{r=1}^{k}\sum_{j=1}^{p} c_r c_a N_{rj}\,\mathrm{E}\left[x_{ji}x_{ti}\right].
\end{aligned}$$

Rewriting the FOC corresponding to N_{at}, we have

$$\sum_{j=1}^{p} c_a^2 N_{aj} \operatorname{E}\left[\frac{1}{\pi_{ai}} x_{ji} x_{ti}\right] + \sum_{r=1}^{k}\sum_{j=1}^{p} c_r c_a N_{rj} \operatorname{E}\left[x_{ji} x_{ti}\right] = \sum_{r=1}^{k} c_r c_a \operatorname{Cov}\left(\frac{Y_{ri}}{\pi_{ri}}, \left(\frac{\mathbb{R}_{ai}}{\pi_{ai}} - 1\right) x_{ti}\right).$$

Let's try simplifying the right side. We have

$$\begin{aligned}
\operatorname{Cov}\left(\frac{Y_{ri}\mathbb{R}_{ri}}{\pi_{ri}}, \left(\frac{\mathbb{R}_{ai}}{\pi_{ai}} - 1\right) x_{ti}\right) &= \operatorname{E}\left[\left(\frac{Y_{ri}\mathbb{R}_{ri}}{\pi_{ri}}\right)\left(\frac{\mathbb{R}_{ai}}{\pi_{ai}} - 1\right) x_{ti}\right] \\
&= \operatorname{E}\left[\left(\frac{Y_{ri}\mathbb{R}_{ri}}{\pi_{ri}}\right)\left(\frac{\mathbb{R}_{ai}}{\pi_{ai}}\right) x_{ti}\right] - \operatorname{E}\left[\left(\frac{Y_{ri}\mathbb{R}_{ri}}{\pi_{ri}}\right) x_{ti}\right].
\end{aligned}$$

For the first term, note that the expectation equals zero when $r \neq a$. When $r = a$, we have

$$\begin{aligned}
\operatorname{E}\left[\left(\frac{Y_{ai}\mathbb{R}_{ai}}{\pi_{ai}}\right)\left(\frac{\mathbb{R}_{ai}}{\pi_{ai}}\right) x_{ti}\right] &= \operatorname{E}\left[\frac{Y_{ai}\mathbb{R}_{ai}}{\pi_{ai}^2} x_{ti}\right] \quad \text{because } \mathbb{R}_{ai} \in \{0,1\} \\
&= \operatorname{E}\left[\operatorname{E}\left[\frac{Y_{ai}\mathbb{R}_{ai}}{\pi_{ai}^2} x_{ti} \middle| x_i\right]\right] \quad \text{by LIE} \\
&= \operatorname{E}\left[\operatorname{E}\left[Y_{ai}|x_i\right] \operatorname{E}\left[\mathbb{R}_{ai}|x_i\right] \frac{x_{ti}}{\pi_{ai}^2}\right] \quad \text{by conditional independence} \\
&= \operatorname{E}\left[\frac{\operatorname{E}\left[Y_{ai}|x_i\right] x_{ti}}{\pi_{ai}}\right] \\
&= \operatorname{E}\left[\frac{Y_{ai} x_{ti}}{\pi_{ai}}\right] \quad \text{by LIE.}
\end{aligned}$$

For the second term, we have

$$\begin{aligned}
-\operatorname{E}\left[\left(\frac{Y_{ri}\mathbb{R}_{ri}}{\pi_{ri}}\right) x_{ti}\right] &= -\operatorname{E}\left[\operatorname{E}\left[\left(\frac{Y_{ri}\mathbb{R}_{ri}}{\pi_{ri}}\right) x_{ti} \middle| x_i\right]\right] \\
&= -\operatorname{E}\left[\operatorname{E}\left[Y_{ri}\mathbb{R}_{ri} \middle| x_i\right] \frac{x_{ti}}{\pi_{ri}}\right] \\
&= -\operatorname{E}\left[\operatorname{E}\left[Y_{ri}|x_i\right] \operatorname{E}\left[\mathbb{R}_{ri}|x_i\right] \frac{x_{ti}}{\pi_{ri}}\right] \\
&= -\operatorname{E}\left[\operatorname{E}\left[Y_{ri}|x_i\right] x_{ti}\right] \\
&= -\operatorname{E}\left[Y_{ri} x_{ti}\right] \quad \text{by LIE.}
\end{aligned}$$

Therefore,

$$\sum_{r=1}^{k} c_r c_a \operatorname{Cov}\left(\frac{Y_{ri}\mathbb{R}_{ri}}{\pi_{ri}}, \left(\frac{\mathbb{R}_{ai}}{\pi_{ai}} - 1\right) x_{ti}\right) = c_a^2 \operatorname{E}\left[\frac{Y_{ai} x_{ti}}{\pi_{ai}}\right] - \sum_{r=1}^{k} c_r c_a \operatorname{E}\left[Y_{ri} x_{ti}\right]$$

Rewriting the FOC corresponding to N_{at} one more time, we have

$$\sum_{j=1}^{p} c_a^2 N_{aj} \operatorname{E}\left[\frac{1}{\pi_{ai}} x_{ji} x_{ti}\right] + \sum_{r=1}^{k}\sum_{j=1}^{p} c_r c_a N_{rj} \operatorname{E}\left[x_{ji} x_{ti}\right] = c_a^2 \operatorname{E}\left[\frac{Y_{ai} x_{ti}}{\pi_{ai}}\right] - \sum_{r=1}^{k} c_r c_a \operatorname{E}\left[Y_{ri} x_{ti}\right]$$

In general, the above equation has many solutions, so the optimal adjustment coefficient is not unique.