# Quantifying Treatment Effects: Estimating Risk Ratios via Observational Studies

**Ahmed Boughdiri**
INRIA Sophia-Antipolis

**Julie Josse**
INRIA Sophia-Antipolis

**Erwan Scornet**
Sorbonne Université and
Université Paris Cité

## Abstract

Randomized Controlled Trials (RCT) are the current gold standards to empirically measure the effect of a new drug. However, resorting to complementary non-randomized data (observational data), which are larger and more diverse, is promising. Medical guidelines recommend reporting the Risk Difference (RD) and the Risk Ratio (RR), which may provide a different comprehension of the effect of the same drug. Contrary to RD, only few methods exist to estimate the RR for observational data. In this paper, we analyze the well-known RR estimator used in RCT and propose several RR estimators in observational data. For all estimators, we establish their asymptotic normality and derive asymptotic confidence intervals. We compare the empirical performances of the different estimators in a simulation study. We also analyze the coverage and the length of the proposed confidence intervals.

## 1 INTRODUCTION

**Treatment effect estimation in trials.** Modern evidence-based medicine prioritizes Randomized Controlled Trials (RCTs) as the cornerstone of clinical evidence. Randomization in RCTs allows for the quantification of the average treatment effect (ATE) by removing confounding influences from extraneous or undesirable factors. The medical guideline CONSORT (Moher et al., 2010) recommends reporting the treatment effect with relative measures like the Risk Ratio (RR) along with absolute measures like the Risk Difference (RD) to provide a more comprehensive understanding of the effect and its implications, as neither measure alone offers a complete picture. Indeed, selecting one measure over another carries several implications. Naylor et al. (1992) and Forrow et al. (1992) demonstrated that physicians' inclination to treat patients, based on their perception of therapeutic impact, is influenced by the scale utilized to present clinical effects. In addition, the treatment effect may be heterogeneous in one scale, i.e. the treatment effect varies according to patient characteristics, but homogeneous in another scale (Rothman, 2011), which significantly disrupts interpretation. Colnet et al. (2024) discusses causal measure properties with a focus on generalization of the treatment effect from a trial to a target population.

Consequently, both RD and RR measures are widely used in the analysis of clinical trial data as explained in Malenka et al. (1993), Sinclair and Bracken (1994) and Nakayama et al. (1998). The Risk Ratio is particularly relevant in scenarios where outcomes are always either positive or negative (Malenka et al., 1993) and in cases where the two proportions being compared are small, as it is more stable and interpretable than the RD. Moreover, when probabilities are low, the RR closely approximates the Odds Ratio (OR), further enhancing its utility in clinical analyses Schechtman (2002). Barratt et al. (2004) recommend using the RR from clinical trials with an estimation of the individual patient baseline to provide the right treatment.

**Treatment effect estimation in observational data.** Despite being the gold standard to assess treatment effects, RCTs may face limitations due to stringent eligibility criteria, unrealistic real-world compliance, short study durations, and limited sample sizes. Medical journals such as JAMA (Bibbins-Domingo, 2024) and others (Hernan and Robins, 2016) have advocated the use of real-world data, often referred to as observational data, to provide additional sources of evidence. These data sets are typically less expensive to collect, more representative of the target population, and usually encompass large sample sizes.

Proceedings of the 28[th] International Conference on Artificial Intelligence and Statistics (AISTATS) 2025, Mai Khao, Thailand. PMLR: Volume 258. 

In the context of observational studies, various estimators exist to measure the treatment effect, mainly on an absolute scale. Different methods such as re-weighting using Inverse Probability Weighting (IPW) (Hirano et al., 2003), outcome modeling with the G-formula, and doubly robust approaches like Augmented Inverse Probability Weighting (AIPW) (Robins, 1986) aim to estimate the RD while minimizing confounding effects.

However, to the best of our knowledge, there exist no work proposing or using estimators of the ATE measured with RR in observational studies based on (non-)parametric estimation (G-formula or AIPW approaches), nor derivations of their theoretical properties. There is a clear gap and a need for robust estimators and comprehensive analyses of their properties to offer better assessments of treatment effects and follow medical recommendations even in observational studies.

**Contributions.** In this paper, we propose and analyze different estimators of the Risk Ratio in observational studies. Considering first the well studied RCT setting in Section 2, we analyze the first RR estimator introduced by Cornfield (1956), establishing a Central Limit Theorem and asymptotic confidence intervals. As for the RD, we prove that in RCT adjusting for covariates as well as estimating the probability of being treated rather than using the true probability is better in terms of estimator variance. As the probability of being treated varies across individuals in observational studies, the above estimator is no longer valid. In Section 3, we detail different estimators that can be used for estimating the RR in observational studies: Inverse Propensity Weighting (RR-IPW), G-formula (RR-G) or doubly robust estimators (RR-OS and RR-AIPW). For the first two, we prove their asymptotic normality when the true surface responses and propensity scores are known. For the last two, using influence function theory (see, e.g., Kennedy, 2022), we prove that they are asymptotically unbiased and have the minimal variance among all asymptotically unbiased estimators. Surprisingly, the RR-AIPW estimator turns out to be a plug-in version of AIPW estimators for both the numerators and denominators, which requires weaker assumptions than RR-OS to be asymptotically normal. All Central Limit Theorems allow us to build asymptotic confidence intervals.

Compared to the Risk Difference, studying the Risk Ratio induces additional technical difficulties, due to its non-linear nature. Dedicated mathematical tools are used based on semi-parametric theory. In Section 4, we evaluate all estimators on observational data, and study the empirical properties in terms of coverage and confidence intervals lengths of all asymptotic confidence intervals established in Section 2 and Section 3.

**Related work - Estimation of RR.** To the best of our knowledge, Cornfield (1956) was the first to propose an estimate of the RR, together with exact and asymptotic confidence intervals, for binary responses, in a RCT scenario. Following this seminal work, Kupper et al. (1975); Katz et al. (1978); Bailey (1987); Morris and Gardner (1988); Sato (1992) also propose asymptotic confidence intervals for the risk ratio for binary outcomes. Considering a logistic model, Schouten et al. (1993) propose a RR estimator. Later on, exact confidence intervals were derived by Wang and Shan (2015). Recently, Inverse Propensity Weighting schemes have been used in different study designs to estimate the Odds Ratios, quantities close to the Risk Ratio in some scenario (Staus et al., 2022). Besides, pseudo-Poisson and pseudo normal distribution have been proposed with IPW strategies to estimate RD and RR in clinical trials (Noma et al., 2023).

In observational studies, one can mention the work of Richardson et al. (2017), Yadlowsky et al. (2021) and Shirvaikar and Holmes (2023) who focus on estimators of the conditional average treatment effect (CATE) for the RR. Curth et al. (2020) introduces an "IF-learning" approach with pseudo outcome regression and derive the influence function for the CATE of the RR. Unfortunately, CATE estimations do not directly yield ATE for the RR. This departs from the RD, for which the ATE is simply the expectation of the CATE.

## 2 A WELL-KNOWN RISK RATIO ESTIMATOR IN RCT

**Problem setting** Following the potential outcome framework (see Rubin, 1974; Splawa-Neyman et al., 1990), we consider the random variables $(X, T, Y^{(0)}, Y^{(1)})$, where $X \in \mathbb{R}^p$ denotes covariates describing a patient, $T$ is the treatment assignment ($T = 1$ when the treatment is given to an individual, $T = 0$ otherwise) and $Y^{(0)}$ (resp. $Y^{(1)}$) is the outcome of interest, describing the status of a patient without treatment (and with treatment respectively). In practice, we do not have access simultaneously to $Y^{(1)}$ and $Y^{(0)}$, and we only observe

$$Y = TY^{(1)} + (1 - T)Y^{(0)}.$$

Causal effect measures are functions of the joint distribution of potential outcomes (see Pearl, 2009). In particular, the Risk Difference (RD) and the Risk Ratio (RR) contrast the two states as followed

$$\tau_{\mathrm{RD}} = \mathbb{E}[Y^{(1)}] - \mathbb{E}[Y^{(0)}] \quad \text{and} \quad \tau_{\mathrm{RR}} = \frac{\mathbb{E}[Y^{(1)}]}{\mathbb{E}[Y^{(0)}]}. \quad (1)$$

The aim of this paper is to propose and study estimators of $\tau_{\mathrm{RR}}$. To estimate this quantity, we assume to be given an i.i.d dataset $(X_1, T_1, Y_1), \ldots, (X_n, T_n, Y_n)$.

The most simple estimator of the risk ratio consists in replacing expectations by empirical means in $\tau_{\mathrm{RR}}$ (Equation 1). Such an estimator has already been proposed outside the causal inference framework, with confidence intervals for the binary case (where $Y \in \{0,1\}$, see Katz et al., 1978; Bailey, 1987). In the potential outcome framework, inspired by the Neyman estimator of the Risk Difference (Splawa-Neyman et al., 1990), we call this estimator the Risk Ratio Neyman estimator.

**Definition 1** (**Risk Ratio Neyman estimator**). *Let $N_1 = \sum_{i=1}^n T_i$ and $N_0 = n - n_1$. The Risk Ratio Neyman estimator, denoted $\hat{\tau}_{RR,N,n}$, is defined as*

$$\hat{\tau}_{RR,N,n} = \frac{\frac{1}{N_1}\sum_{i=1}^n T_i Y_i}{\frac{1}{N_0}\sum_{i=1}^n (1-T_i)Y_i}, \tag{2}$$

*if the denominator is nonzero and* 0 *otherwise.*

With our notation, the 95% confidence interval for $\tau_{\mathrm{RR}}$ proposed by Katz et al. (1978) for binary outcomes takes the form

$$\left[\hat{\tau}_{\mathrm{RR,N,n}} e^{-z_{1-\alpha/2}\hat{\sigma}_n}, \hat{\tau}_{\mathrm{RR,N,n}} e^{z_{1-\alpha/2}\hat{\sigma}_n}\right] \tag{3}$$

where $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of a standard Gaussian $\mathcal{N}(0,1)$ and

$$\hat{\sigma}_n = \sqrt{\frac{1}{\sum_{i=1}^n T_i Y_i} - \frac{1}{N_1} + \frac{1}{\sum_{i=1}^n (1-T_i)Y_i} - \frac{1}{N_0}}. \tag{4}$$

In the sequel, we establish under which theoretical assumptions the RR-N is an accurate estimator of the Risk Ratio in Randomized Clinical Trials.

Randomized Controlled Trials (RCT) randomly assign treatment to patients in order to evaluate treatment effects. We focus on the Bernoulli design, one of the most widely used RCT designs (Rubin, 1974; Imbens and Rubin, 2015), where each participant has the same probability $e \in (0,1)$ of being treated, independently of the treatments of others. In this section, we use the following assumptions.

**Assumption 1** (Bernoulli Trial). *We consider the following assumptions:*

1. ***Ignorability*** *or Exchangeability, that is:* $T \perp\!\!\!\perp (Y^{(0)}, Y^{(1)})$.
2. ***SUTVA*** *(Stable Unit Treatment Value Assumption):* $Y = TY^{(1)} + (1-T)Y^{(0)}$.
3. ***i.i.d.*** *The data set is i.i.d.* $(X_i, T_i, Y_i^{(0)}, Y_i^{(1)})_{i=1,\dots,n} \overset{i.i.d.}{\sim} \mathcal{P}$. *In particular, the treatment assignment of one participant does not influence that of another, that is, for all* $i \neq j$, $T_i \perp\!\!\!\perp T_j$.
4. ***Trial positivity****: Each participant $i$ has a fixed probability $e \in (0,1)$ of being assigned to the intervention group $\mathbb{P}[T_i = 1] = e$.*

To ensure our estimates are valid, we need to guarantee the existence of the ratio we aim to estimate.

**Assumption 2** (**Outcome positivity**). *We suppose that both $Y^{(0)}$ and $Y^{(1)}$ are squared integrable and that $\mathbb{E}\left[Y^{(0)}|X\right], \mathbb{E}\left[Y^{(1)}|X\right] > 0$.*

**Proposition 1** (**Asymptotic normality of $\hat{\tau}_{\mathbf{RR,N,n}}$**). *Grant Assumption 1 and Assumption 2, the Risk Ratio Neyman estimator is asymptotically unbiased and satisfies*

$$\sqrt{n}\left(\hat{\tau}_{RR,N,n} - \tau_{RR}\right) \xrightarrow{d} \mathcal{N}\left(0, V_{RR,N}\right) \tag{5}$$

*where*

$$V_{RR,N} = \tau_{RR}^2 \left(\frac{\mathrm{Var}\left(Y^{(1)}\right)}{e\mathbb{E}[Y^{(1)}]^2} + \frac{\mathrm{Var}\left(Y^{(0)}\right)}{(1-e)\mathbb{E}[Y^{(0)}]^2}\right).$$

Proposition 1 establishes the asymptotic normality of the RR-N estimator, a simple ratio of mean estimates, which leads to asymptotic confidence intervals (CI). Indeed, according to Proposition 1, for all $\alpha \in (0,1)$, a $(1-\alpha)$ asymptotic confidence interval for $\tau_{\mathrm{RR}}$ is given by

$$\left[\hat{\tau}_{\mathrm{RR,N,n}} \pm z_{1-\alpha/2}\sqrt{\widehat{V_{\mathrm{RR,N}}}/n}\right]. \tag{6}$$

Throughout this paper, based on the Central Limit Theorems we establish, we will consider such CI. The properties of the different CI are studied in Section 4.3.

Contrary to Katz et al. (1978), Proposition 1 is valid for both continuous and binary outcomes. However, considering binary outcomes in Proposition 1 leads to an asymptotic confidence interval equivalent to that presented in Katz et al. (1978) (see Section 6.2.3). Deriving a Central Limit Theorem for $\log(\hat{\tau}_{\mathrm{RR,N,n}})$ instead of $\hat{\tau}_{\mathrm{RR,N,n}}$ would lead to the exact same CI (see Section 6.2.4).

**Probability of receiving treatment** As the probability of treatment $e$ is known in an RCT, one could be tempted to consider what we call the Risk Ratio Horvitz-Thomson estimator (in reference of the Risk Difference Horvitz-Thomson estimator of the RD, see Horvitz and Thompson, 1952) defined as

$$\hat{\tau}_{\mathrm{RR,HT,n}} = \frac{\sum_{i=1}^n \frac{T_i Y_i}{e}}{\sum_{i=1}^n \frac{(1-T_i)Y_i}{1-e}} \tag{7}$$

if $\sum_{i=1}^n T_i < n$ and 0 otherwise. Indeed, the frequency of treatments assignments in the sample may be different from the actual probability of receiving treatment

$e$. Similarly to what Hirano et al. (2003); Hahn (1998); Robins et al. (1992) noticed for the RD, we prove in 6.2.2 that opting for $\hat{e}$ over $e$ in the Risk Ratio estimator (thereby employing the RR-N instead of the RR-HT) results in a reduced asymptotic variance, with a larger reduction when $e$ is close to zero or one. More precisely, letting $V_{\mathrm{RR,HT}}$ the asymptotic variance of $\hat{\tau}_{\mathrm{RR,HT,n}}$, we have

$$V_{\mathrm{RR,N}} = V_{\mathrm{RR,HT}} - \tau_{\mathrm{RR}}^2/e(1-e). \tag{8}$$

# 3 RISK RATIO ESTIMATORS IN OBSERVATIONAL STUDIES

Observational studies reveal the complexities of real-world scenarios, which may be missed by the controlled designs of RCTs. A key difference between RCTs and observational studies is the handling of confounding variables. If not properly addressed, these can distort the true causal association between exposure and outcome due to their correlation with both. Therefore, estimating the Risk Ratio in observational studies is more complex than in RCTs, as randomization assumptions do not apply (i.e. the propensity score now depends on the covariates $X$).

**Assumption 3 (Observational study identifiability assumptions).** *We have*

1. ***Unconfoundedness*** *or Conditional Exchangeability:* $(Y(0), Y(1)) \perp\!\!\!\perp T \mid X$.
2. ***Overlap*** *or Positivity,* $\exists \eta \in (0, 1/2]$ *such that, almost surely,* $\eta \leq \mathbb{P}[T = 1|X] \leq 1 - \eta$.
3. ***SUTVA*** *(Stable Unit Treatment Value Assumption) holds:* $Y = TY^{(1)} + (1-T)Y^{(0)}$.
4. ***i.i.d.*** *We still assume that the data set is i.i.d.* $(X_i, T_i, Y_i^{(0)}, Y_i^{(1)})_{i=1,\dots,n} \overset{i.i.d.}{\sim} \mathcal{P}$.

Unconfoundedness means that after accounting for known confounding variables, no hidden factors affect both treatment assignment and outcomes. It is a relaxed form of exchangeability.

The RR-N estimator (or the RR-HT estimator) cannot be used in the context of observational studies, since they are built on the assumption of a constant propensity score. However, the RR-N estimator can be extended to observational studies as follows.

## 3.1 Risk Ratio Inverse Propensity Weighting (RR-IPW)

Treatment effect in observational studies can be estimated via reweighting individuals by the inverse of their propensity score, thus giving more weights to people who are very likely/unlikely to be treated. Such a method, called Inverse Propensity Weighting (IPW, see Hirano et al., 2003) for estimating the Risk Difference, can be straightforwardly extended to build Risk Ratio estimators.

**Definition 2 (RR-IPW).** *Grant Assumption 2 and Assumption 3. Given an estimator* $0 < \hat{e}(\cdot) < 1$ *of the propensity score* $e(x) = \mathbb{P}[T = 1|X = x]$*, the Risk Ratio IPW, denoted by* $\hat{\tau}_{RR,IPW,n}$*, is defined as*

$$\hat{\tau}_{RR,IPW,n} = \frac{\sum_{i=1}^n \frac{T_i Y_i}{\hat{e}(X_i)}}{\sum_{i=1}^n \frac{(1-T_i)Y_i}{1-\hat{e}(X_i)}}.$$

Proposition 2 demonstrates the asymptotic normality of the Oracle Ratio IPW estimator, defined as the RR-IPW but where $\hat{e}(\cdot)$ is replaced by the oracle propensity score $e(\cdot)$.

**Proposition 2 (RR-IPW asymptotic normality).** *Grant Assumption 2 and Assumption 3. Then the Oracle Risk Ratio IPW defined above is asymptotically unbiased and satisfies*

$$\sqrt{n}\left(\tau^{\star}_{RR,IPW} - \tau_{RR}\right) \xrightarrow{d} \mathcal{N}\left(0, V_{RR,IPW}\right) \tag{9}$$

*where* $V_{RR,IPW} = \tau_{RR}^2 \left( \frac{\mathbb{E}\left[\frac{(Y^{(1)})^2}{e(X)}\right]}{\mathbb{E}\left[Y^{(1)}\right]^2} + \frac{\mathbb{E}\left[\frac{(Y^{(0)})^2}{1-e(X)}\right]}{\mathbb{E}\left[Y^{(0)}\right]^2} \right)$.

Note that when the propensity score is constant, one can retrieve the variance of the RR-HT as expected. Note also that the asymptotic variance may be large, due to strata on which the propensity score is close to zero or one. In other words, a correct estimation is difficult when some subpopulations are unlikely to be treated (or untreated).

## 3.2 Risk Ratio G-formula estimator (RR-G)

For all $x \in \mathbb{R}^p$, let $\mu_{(0)}(x) = \mathbb{E}\left[Y^{(0)}|X = x\right]$ and $\mu_{(1)}(x) = \mathbb{E}\left[Y^{(1)}|X = x\right]$ be the surface responses of the potential outcomes. Assume that we have at our disposal two estimators $\hat{\mu}_{(0)}(\cdot)$ and $\hat{\mu}_{(1)}(\cdot)$ which respectively estimate $\mu_{(0)}(\cdot)$ and $\mu_{(1)}(\cdot)$. We then employ the ratio of these two potential outcome estimations to compute the Risk Ratio. This method, termed the plug-in G-formula or outcome-based modeling, was first introduced by Robins (1986) for the Risk Difference.

**Definition 3 (Ratio plug-in G-formula).** *Given two estimators* $\hat{\mu}_{(0)}(\cdot)$ *and* $\hat{\mu}_{(1)}(\cdot)$*, the Risk Ratio G-formula estimator, denoted* $\hat{\tau}_{RR,G,n}$*, is defined as*

$$\hat{\tau}_{RR,G,n} = \frac{\sum_{i=1}^n \hat{\mu}_{(1)}(X_i)}{\sum_{i=1}^n \hat{\mu}_{(0)}(X_i)}, \tag{10}$$

*if* $\sum_{i=1}^n \hat{\mu}_{(0)}(X_i) \neq 0$ *and zero otherwise.*

The properties of RR-G depend on the estimators $\hat{\mu}_{(0)}$ and $\hat{\mu}_{(1)}$. We analyze in the following the behavior of Oracle Risk Ratio G-formula estimator defined as $\tau^\star_{\mathrm{RR,G}} = (\sum_{i=1}^n \mu_{(1)}(X_i))/(\sum_{i=1}^n \mu_{(0)}(X_i))$.

**Proposition 3** (**Asymptotic Normality of $\tau^\star_{\mathbf{RR,G}}$**). *Grant Assumption 1 and Assumption 2. Then, the Oracle Risk Ratio G-formula estimator, $\tau^\star_{RR,G}$, is asymptotically unbiased and satisfies*

$$\sqrt{n}\left(\tau^\star_{RR,G} - \tau_{RR}\right) \xrightarrow{d} \mathcal{N}\left(0, V_{RR,G}\right), \qquad (11)$$

*where* $V_{RR,G} = \tau_{RR}^2 \operatorname{Var}\left(\frac{\mu_{(1)}(X)}{\mathbb{E}\left[Y^{(1)}\right]} - \frac{\mu_{(0)}(X)}{\mathbb{E}\left[Y^{(0)}\right]}\right)$.

Proposition 3 establishes that the Oracle Risk Ratio G-formula estimator is asymptotically normal. Surprisingly, in the case where there is no effect (i.e. $\tau_{\mathrm{RR}} = 1$), the asymptotic variance is driven by the variance of the Risk Difference on each strata determined by $X$, namely $\operatorname{Var}(\mu_{(1)}(X) - \mu_{(0)}(X))$. By considering the Oracle RR-G instead of RR-G, we remove the additional randomness related to the estimation of the surface responses. It is thus likely that the true variance of RR-G is larger than that of Oracle RR-G.

If we assume a linear model for the $Y^{(t)}$ and estimate both response surfaces $\hat{\mu}_{(0)}$ and $\hat{\mu}_{(1)}$ using ordinary least squares, the variance of the RR-G can be derived,

**Assumption 4** (Linear model). *For all $t \in \{0,1\}$,*

$$Y^{(t)} = c_{(t)} + X^\top \beta_{(t)} + \varepsilon_{(t)} \quad \mathbb{E}[X] = \mu$$
$$\mathbb{E}[\varepsilon_{(t)}|X] = 0 \quad Var[\varepsilon_{(t)}|X] = \sigma^2,$$

*where we assume that $Y^{(t)} \geq c > 0$ for some $c$.*

For any positive semi-definite matrix $A$ and any vector $X$, let $\|X\|_A = \sqrt{X^\top A X}$.

**Proposition 4** (Asymptotic normality of $\hat{\tau}_{\mathrm{RR,OLS}}$). *Grant Assumption 4. Then, the Risk Ratio G-formula estimator $\hat{\tau}_{RR,OLS}$ that uses linear regression to estimate $\mu_{(t)}$ satisfies*

$$\sqrt{n}(\hat{\tau}_{RR,OLS} - \tau_{RR}) \xrightarrow{d} \mathcal{N}\left(0, V_{RR\text{-}OLS}\right)$$

*where, letting $\nu_t = \mathbb{E}[X|T=t]$ and $\Sigma_t = \operatorname{Var}(X|T=t)$,*

$$\frac{V_{RR\text{-}OLS}}{\tau_{RR}^2} = \left\| \frac{\beta_{(1)}}{\mathbb{E}\left[Y^{(1)}\right]} - \frac{\beta_{(0)}}{\mathbb{E}\left[Y^{(0)}\right]} \right\|_\Sigma^2 + \sigma^2 \times \left( \frac{1 + (1-e)^2 \|\nu_1 - \nu_0\|^2_{\Sigma_1^{-1}}}{e\mathbb{E}\left[Y^{(1)}\right]^2} + \frac{1 + e^2 \|\nu_1 - \nu_0\|^2_{\Sigma_0^{-1}}}{(1-e)\mathbb{E}\left[Y^{(0)}\right]^2} \right).$$

The variance of RR-OLS can be decomposed in two terms, one which corresponds to the oracle variance of RR-G and another term which is due to the fact that the response surfaces are estimated by OLS estimators. Note, that if we use RR-G in a Bernoulli Trial under a linear model then one can show that even in an RCT setting, adjusting for covariates is beneficial as the variance of the RR-G is smaller than the variance of RR-N. These results are provided in Appendix 6.3.3.

### 3.3 Risk Ratio One-step estimator (RR-OS)

A popular estimator for the RD is the augmented inverse probability weighted estimator (AIPW, see Robins et al., 1992). AIPW combines the properties of G-formula and IPW estimator and is *doubly robust* in the sense that it is consistent as soon as either the propensity or outcome models to be correctly specified. By calculating the influence function of the statistical estimand $\psi_{\mathrm{RD}} = \mathbb{E}\left[\mathbb{E}\left[Y|T=1,X\right] - \mathbb{E}\left[Y|T=0,X\right]\right]$ we obtain an efficient estimator, since it has no asymptotic bias and the minimal asymptotic variance.

(Kennedy, 2022). Therefore, to estimate the Risk Ratio (RR), a natural approach is to derive an efficient estimator using semi-parametric theory (Tsiatis, 2006). Considering the statistical estimand $\psi_{\mathrm{RR}} = \frac{\mathbb{E}[\mathbb{E}[Y|T=1,X]]}{\mathbb{E}[\mathbb{E}[Y|T=0,X]]}$, we obtain an estimator RR-OS presented below, which is efficient using non-parametric estimation of the nuisance components combined with cross-fitting (see Chernozhukov et al., 2017, for cross-fitting).

**Definition 4** (**Crossfitted RR-OS**). *For all $t \in \{0,1\}$ and all $x$, let $\mu_{(t)}(x) = \mathbb{E}\left[Y^{(0)}|X=x\right]$ and $e_{(t)}(x) = \mathbb{P}\left[T=t|X=x\right]$. We denote $\mathcal{I} = \{1,\dots,n\}$, let $\mathcal{I}_1, \mathcal{I}_2, ..., \mathcal{I}_K$ be a partition of $\mathcal{I}$. For all $t \in \{0,1\}$, let*

$$\hat{\tau}_{AIPW,t} = \sum_{k=1}^K \sum_{i \in \mathcal{I}_k} \left( \hat{\mu}_{(t)}^{\mathcal{I}_{-k}}(X_i) + \frac{Y_i - \hat{\mu}_{(t)}^{\mathcal{I}_{-k}}(X_i)}{\hat{e}_{(t)}^{\mathcal{I}_{-k}}(X_i)} \mathbb{1}_{T_i = t} \right) \qquad (12)$$

$$\text{and} \quad \hat{\tau}_{G,t} = \sum_{k=1}^K \sum_{i \in \mathcal{I}_k} \hat{\mu}_{(t)}^{\mathcal{I}_{-k}}(X_i), \qquad (13)$$

*where $\hat{\mu}_{(t)}^{\mathcal{I}_{-k}}(X)$ and $\hat{e}_{(t)}^{\mathcal{I}_{-k}}(X)$ are estimators of $\mu_{(t)}$ and $e_{(t)}$ built on the sample $\mathcal{I}_{-k} = \mathcal{I} \backslash \mathcal{I}_k$. The crossfitted Risk Ratio One-Step (RR-OS) estimator $\hat{\tau}_{RR\text{-}OS}$ is defined as*

$$\hat{\tau}_{RR\text{-}OS} = \frac{\hat{\tau}_{G,1}}{\hat{\tau}_{G,0}} \left(1 - \frac{\hat{\tau}_{AIPW,0}}{\hat{\tau}_{G,0}}\right) + \frac{\hat{\tau}_{AIPW,1}}{\hat{\tau}_{G,0}}.$$

**Proposition 5** (**Asymptotic normality of $\hat{\tau}_{\mathbf{RR\text{-}OS}}$**). *Grant Assumption 2 and Assumption 3. Assume that for all $1 \leq k \leq K$, and for all $t \in \{0,1\}$,*

$$\mathbb{E}\left[\left(\hat{\mu}_{(t)}^{\mathcal{I}_{-k}}(X_i) - \mu_{(t)}(X)\right)^2\right] \mathbb{E}\left[\left(\hat{e}^{\mathcal{I}_{-k}}(X) - e(X)\right)^2\right] = o\left(n^{-1}\right) \quad (14)$$

$$\mathbb{E}\left[\hat{\mu}_{(0)}^{\mathcal{I}_{-k}}(X)\right] - \mathbb{E}\left[\mu_{(0)}^{\mathcal{I}_{-k}}(X))\right] = o\left(n^{-1/4}\right) \qquad (15)$$

$$\mathbb{E}\Big[(\hat{\mu}_{(0)}^{\mathcal{I}_{-k}}(X)-\mu_{(0)}^{\mathcal{I}_{-k}}(X))^2\Big]\mathbb{E}\Big[(\hat{\mu}_{(1)}^{\mathcal{I}_{-k}}(X)-\mu_{(1)}^{\mathcal{I}_{-k}}(X))^2\Big]=o\big(n^{-1}\big), \tag{16}$$

*with* $\eta \leq \hat{e}^{\mathcal{I}_{-k}}(\cdot) \leq 1-\eta$ *(see Positivity in Assumption 3). Then the One-Step estimator is asymptotically unbiased and satisfies*

$$\sqrt{n}\left(\hat{\tau}_{RR\text{-}OS} - \tau_{RR}\right) \xrightarrow{d} \mathcal{N}\left(0, V_{RR,OS}\right),$$

*where*

$$\frac{V_{RR,OS}}{\tau_{RR}^2} = \mathrm{Var}\left(\frac{\mu_{(1)}(X)}{\mathbb{E}\left[Y^{(1)}\right]} - \frac{\mu_{(0)}(X)}{\mathbb{E}\left[Y^{(0)}\right]}\right) + \mathbb{E}\left[\frac{\mathrm{Var}\left(Y^{(1)}|X\right)}{e(X)\mathbb{E}\left[Y^{(1)}\right]^2}\right] + \mathbb{E}\left[\frac{\mathrm{Var}\left(Y^{(0)}|X\right)}{(1-e(X))\mathbb{E}\left[Y^{(0)}\right]^2}\right].$$

Similar to the Risk Difference AIPW, this estimator is efficient: its asymptotic variance is optimal. The semi parametric theory develops efficient estimators by compensating for the first-order bias (Kennedy, 2022), this can be achieved either by estimating and subtracting the first-order bias, leading to the RR-OS estimator or by finding values for the target parameter and nuisance parameters that solve the estimating equation (see A.Schuler, 2024, for the RD case) and eliminate the first-order bias (as detailed in Section 6.3.5), resulting in RR-AIPW presented below.

### 3.4 Risk Ratio Augmented Inverse Propensity Weighting (RR-AIPW)

**Definition 5** (**Crossfitted RR-AIPW**). *The Risk Ratio AIPW crossfitted is defined as*

$$\hat{\tau}_{RR,AIPW} := \frac{\hat{\tau}_{AIPW,1}}{\hat{\tau}_{AIPW,0}},$$

*where* $\hat{\tau}_{AIPW,0}$*, and* $\hat{\tau}_{AIPW,1}$ *are defined in 4.*

The RR-AIPW is simply the ratio of two one-step estimators, one for $\mathbb{E}\left[Y^{(1)}\right]$ and one for $\mathbb{E}\left[Y^{(0)}\right]$. This method may seem simplistic at first glance, since approximating both the numerator and denominator usually results in a non-zero asymptotic bias. However, RR-AIPW is derived via the estimating equation method using influence function theory, which results in an efficient (asymptotically unbiased) estimator. Note that in the case of the Risk Difference (RD), both approaches (One-step bias correction and estimating equation) yield the same AIPW estimator. However, because our statistical estimand for the Risk Ratio is nonlinear, the resulting estimators differ. It remains that they are both efficient, as shown below.

**Proposition 6** (**Risk Ratio AIPW asymptotic normality**). *Grant Assumption 2 and Assumption 3. Assume that Equation (14) holds and that, for all* $1 \leq k \leq K$ *and all* $t \in \{0,1\}$*,*

$$\mathbb{E}\Big[\Big(\hat{\mu}_{(t)}^{\mathcal{I}_{-k}}(X)-\mu_{(t)}(X)\Big)^2\Big]=o(1), \quad \mathbb{E}\Big[\Big(\hat{e}^{\mathcal{I}_{-k}}(X)-e(X)\Big)^2\Big]=o(1), \tag{17}$$

*with* $\eta \leq \hat{e}^{\mathcal{I}_{-k}}(\cdot) \leq 1-\eta$*. Then, the crossfitted Risk Ratio AIPW estimator is asymptotically unbiased and satisfies*

$$\sqrt{n}\left(\hat{\tau}_{RR,AIPW} - \tau_{RR}\right) \xrightarrow{d} \mathcal{N}\left(0, V_{RR,OS}\right),$$

*where* $V_{RR,OS}$ *is defined in Proposition 5.*

Assumptions in Proposition 6 are the same as those used in the Risk Difference AIPW estimator (Wager, 2020) to achieve double robustness. Additionally, Assumption 14, often referred to as risk decay, holds when either surfaces responses or propensity score achieve a parametric rate, while the other is only consistent. This departs from RR-OS where asymptotic normality is not achieved when only the propensity score has a parametric convergence rate. Consequently, we recommend RR-AIPW over RR-OS as they have the same asymptotic properties, with weaker assumptions for RR-AIPW.

## 4 SIMULATION

Simulations for randomized controlled trials are provided in Appendix 7. For observational studies, we generate datasets $(X, T, Y^{(0)}, Y^{(1)})$ according to the following model

$$\begin{aligned} Y^{(1)} &= m(X) + b(X) + \varepsilon_{(1)} & \mathbb{P}\left[T=1|X\right] &= e(X), \\ Y^{(0)} &= b(X) + \varepsilon_{(0)} & \text{with } \varepsilon_{(t)} &\sim \mathcal{N}\left(0, \sigma^2\right). \end{aligned}$$

Each of the following setups has specificities regarding the $m(.)$, $b(.)$, and $e(.)$ functions, which respectively correspond to the treatment effect, the baseline and propensity score. The true Risk Ratio can be expressed as $\tau_{\mathrm{RR}} = \mathbb{E}\left[Y^{(1)}\right]/\mathbb{E}\left[Y^{(0)}\right] = \mathbb{E}\left[m(X)\right]/\mathbb{E}\left[b(X)\right] + 1$. We compare the performances of all estimators defined in Section 3 where nuisance components (regression surfaces and propensity score) are estimated via parametric (linear/logistic regression) or non-parametric methods (random forests). Each simulation is repeated 3000 times. More details are provided in Appendix 7.

### 4.1 Linear and Logistic DGP

The first observational data generating process (DGP) is a parametric setup introduced in Lunceford and Davidian (2004), composed of linear outcome models (linear treatment effect and baseline) and a logistic

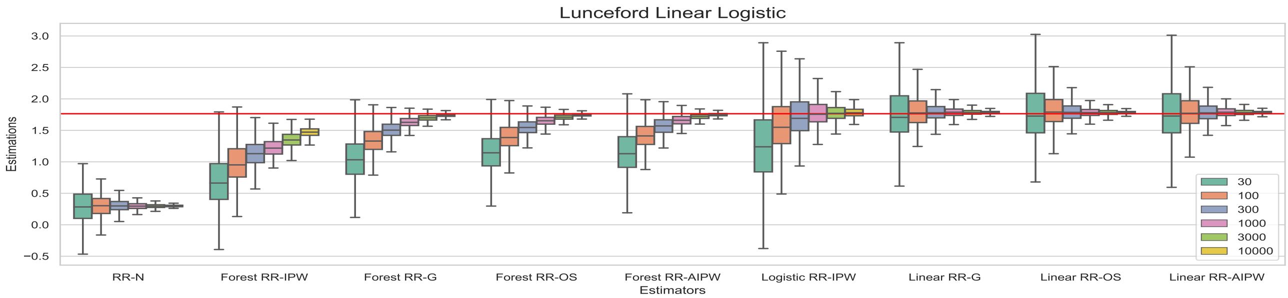


Figure 1: Risk Ratio estimators computed for a Linear/Logistic DGP, with 3000 repetitions.

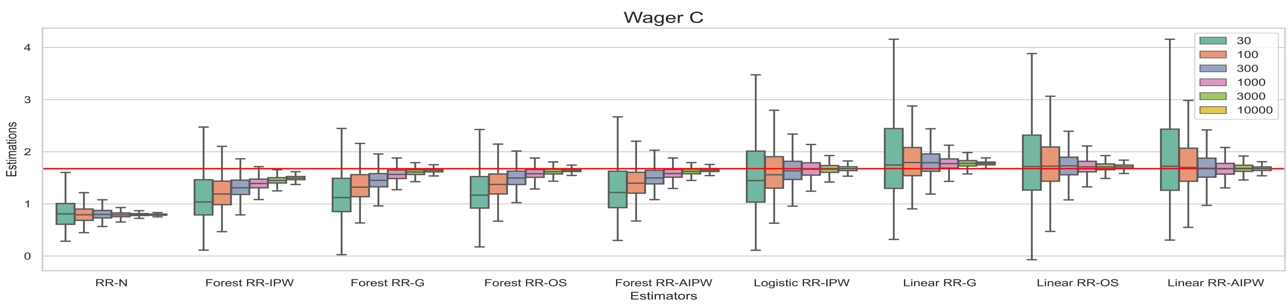


Figure 2: Risk Ratio estimators computed for a non-Linear-Logistic DGP, with 3000 repetitions.

propensity score, that is $m(X, V) = 2$ and :

$$b(X, V) = \beta_0^\top [X, V], \quad e(X) = (1 + \exp(-\beta_e^\top X))^{-1}$$
$$\beta_0 = (-1, 1, -1, -1, 1, 1) \quad \beta_e = (-0.6, 0.6, -0.6),$$

The covariates $X = (X_1, X_2, X_3)^{\mathrm{T}}$ are associated with both treatment exposure and potential response while $V = (V_1, V_2, V_3)^{\mathrm{T}}$ are associated with the potential response but not directly related to treatment exposure. $[X, V]$ follow a joint distribution by taking $X_3 \sim$ Bernoulli $(0.2)$ and then generating $V_3$ as Bernoulli with $P(V_3 = 1 \mid X_3) = 0.75 X_3 + 0.25(1 - X_3)$. Conditional on $X_3, (X_1, V_1, X_2, V_2)^{\mathrm{T}}$ was then generated as multivariate normal $\mathcal{N}(\lambda_{X_3}, \Sigma)$, where

$$\lambda_1 = -\lambda_0 = (1, 1, -1, -1)$$

and

$$\Sigma = \begin{pmatrix} 1 & 0.5 & -0.5 & -0.5 \\ 0.5 & 1 & -0.5 & -0.5 \\ -0.5 & -0.5 & 1 & 0.5 \\ -0.5 & -0.5 & 0.5 & 1 \end{pmatrix}.$$

Results are depicted in Figure 1. Only confounding variables are used as inputs in the different estimators. As expected, since the generative process is linear, methods that use parametric estimators (logistic/linear regression) outperform those using non-parametric approaches (random forests) in finite-sample settings. While all methods (except maybe Forest RR-IPW) converge to the correct RR, methods based on parametric estimators exhibit a faster rate of convergence and are unbiased (except for Logistic RR-IPW) even for small sample sizes. Indeed, random forests are not suited for linear generative process and require here more than 10000 samples to estimate correctly the RR.

All in all, when the outcome modelling and the propensity scores are linear, the two doubly robust estimators (RR-AIPW and RR-OS) and the RR-G, all based on linear estimators, achieve the best performances: they are unbiased, even for small sample sizes, and converge quickly to the true RR.

Furthermore, both Linear RR-OS and RR-AIPW estimators give very similar results.

### 4.2 Non-Linear and Logistic DGP

We use a semi-parametric setup (see Nie and Wager, 2020) with non-linear baseline models, a constant treatment effect and a logistic propensity score:

$$b(X) = 2 \log\left(1 + e^{X_1 + X_2 + X_3}\right),$$
$$e(X) = 1 / \left(1 + e^{X_2 + X_3}\right) \quad \text{and} \quad m(X) = 1,$$

where $X \sim \mathcal{N}(0, I_{d \times d})$. Results are presented in Figure 2. The Forest IPW and Linear G-formula estimators yield poor RR estimates for the largest sample

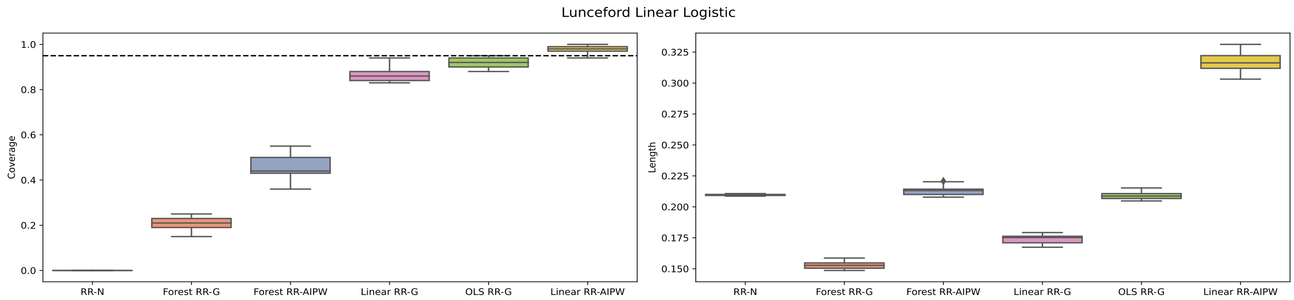


Figure 3: Coverage (left) and Length (right) of asymptotic CI derived from Section 2 and Section 3 for different estimators with $n = 1000$ and 300 repetitions.

size. The forest IPW uses random forests to estimate a logistic model, which may still converge, but at a slower rate than other methods. The Linear G-formula employs linear regressions to estimate the response surfaces, potentially leading to an irreducible asymptotic bias.

The Forest RR-G, Forest RR-AIPW, and Forest RR-OS estimators converge slowly to the true RR. This simulation highlights the doubly robust properties of the Linear RR-AIPW and Linear RR-OS estimators: they target the true RR even at small sample sizes, as they have at least one well-specified model.

### 4.3 Confidence intervals (CI)

In the Linear/Logistic DGP presented in Section 4.1, we build asymptotic 95% CI for the RR-AIPW, RR-G and RR-N estimators based on their asymptotic normality. Variances were estimated following the protocol described in equation 132, 33, 47, 48 and 30. We generate 300 datasets and present, in Figure 3, the distribution of the length and coverage (probability that the CI contains the risk ratio) for each estimator. The IPW was excluded due to poor performances (too large CI). As expected, RR-N CI has nearly zero coverage, since we are not in a RCT setting. The Forest RR-G and RR-AIPW confidence intervals also exhibit poor coverage, which is in agreement with the linear and logistic DGP context. In contrast, Linear RR-G and RR-AIPW demonstrate good coverage. Note that the CI for OLS RR-G, built based on Proposition 4, has a better coverage than the Linear RR-G method, as it takes into account the additional randomness due to linear estimations. Although only the RR AIPW has coverage above 95%, the OLS RR-G has a shorter average predicted length compared to the Linear RR AIPW. Results for the Non-Linear/Logistic and Non-Linear/Non-Logistic DGP can be found in Section 7.2.

## 5 CONCLUSION

Quantifying treatment effects presents challenges, since different measures may lead to different understanding of the same phenomenon. In our study, we focus on one of these measures, the Risk Ratio and introduced several estimators, valid in RCT or observational studies. Using dedicated mathematical tools (influence function theory, M-estimation), we establish their asymptotic normality, limiting variance and derive asymptotic confidence intervals. Empirical evaluations show that RR-N and RR-IPW have poor performances. Either Linear or Forest RR-AIPW (or RR-OS) show similar (good) behaviors to estimate the Risk Ratio, with the best theoretical guarantees among all studied estimators. Since RR-AIPW requires fewer assumption and is simpler to compute, we would recommend to use RR-AIPW. As for the doubly robust approaches, G-formula is competitive, with performances that depend on the setting and the estimation method used for the nuisance components.

Identifying guidelines establishing when linear nuisance components should be used instead of non-parametric ones still remains an open problem. In practice, observational studies may be used to generalize the treatment effect from a RCT population to the general population of interest. Our work is a first step toward proposing procedures to generalize the Risk Ratio to general populations.

### References


A.Schuler, M. v. d. L. (2024). Introduction to modern causal inference. Technical report, https://alejandroschuler.github.io/mci/index.html.

Bailey, B. (1987). Confidence limits to the risk ratio. *Biometrics*, pages 201–205.

Barratt, A., Wyer, P. C., Hatala, R., McGinn, T., Dans, A. L., Keitz, S., Moyer, V., et al. (2004). Tips for learners of evidence-based medicine: 1. relative risk reduction, absolute risk reduction and number needed to treat. *Cmaj*, 171(4):353–358.

Bibbins-Domingo, K. (2024). Causal inference about the effects of interventions from observational studies in medical journals. *JAMA*.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2017). Double/debiased machine learning for treatment and causal parameters.

Colnet, B., Josse, J., Varoquaux, G., and Scornet, E. (2024). Risk ratio, odds ratio, risk difference... which causal measure is easier to generalize? *arXiv preprint arXiv:2303.16008*.

Cornfield, J. (1956). A statistical problem arising from retrospective studies. In *Proceedings of the third Berkeley symposium on mathematical statistics and probability*, volume 4, pages 135–148.

Curth, A., Alaa, A. M., and van der Schaar, M. (2020). Estimating structural target functions using machine learning and influence functions. *arXiv preprint arXiv:2008.06461*.

Forrow, L., Taylor, W. C., and Arnold, R. M. (1992). Absolutely relative: How research results are summarized can affect treatment decisions. *The American Journal of Medicine*, 92(2):121–124.

Hahn, J. (1998). On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica*, 66(2):315–332.

Hernan, M. and Robins, J. (2016). Using big data to emulate a target trial when a randomized trial is not available: Table 1. *American Journal of Epidemiology*, 183:kwv254.

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

Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260):663–685.

Imbens, G. W. and Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press.

Katz, D., Baptista, J., Azen, S., and Pike, M. (1978). Obtaining confidence intervals for the risk ratio in cohort studies. *Biometrics*, pages 469–474.

Kennedy, E. (2015). Semiparametric theory and empirical processes in causal inference.

Kennedy, E. H. (2022). Semiparametric doubly robust targeted double machine learning: a review. *arXiv preprint arXiv:2203.06469*.

Kupper, L., McMichael, A., and Spirtas, R. (1975). A hybrid epidemiologic study design useful in estimating relative risk. *Journal of the American Statistical Association*, 70(351a):524–528.

Lunceford, J. and Davidian, M. (2004). Stratification and weighting via the propensity score in estimation of causal treatment effects: A comparative study. *Statistics in medicine*, 23:2937–60.

Malenka, D. J., Baron, J. A., Johansen, S., Wahrenberger, J. W., and Ross, J. M. (1993). The framing effect of relative and absolute risk. *Journal of general internal medicine*, 8:543–548.

Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gøtzsche, P. C., Devereaux, P. J., Elbourne, D., Egger, M., and Altman, D. G. (2010). Consort 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ*, 340.

Morris, J. A. and Gardner, M. J. (1988). Statistics in medicine: Calculating confidence intervals for relative risks (odds ratios) and standardised ratios and rates. *British medical journal (Clinical research ed.)*, 296(6632):1313.

Nakayama, T., Zaman, M. M., and Tanaka, H. (1998). Reporting of attributable and relative risks, 1966–97. *The Lancet*, 351(9110):1179.

Naylor, C. D., Chen, E., and Strauss, B. (1992). Measured enthusiasm: does the method of reporting trial results alter perceptions of therapeutic effectiveness? *Ann. Intern. Med.*, 117(11):916–921.

Nie, X. and Wager, S. (2020). Quasi-oracle estimation of heterogeneous treatment effects.

Noma, H., Misumi, M., and Tanaka, S. (2023). Risk ratio and risk difference estimation in case-cohort studies. *Journal of Epidemiology*, 33(10):508–513.

Pearl, J. (2009). *Causality*. Cambridge University Press, 2 edition.

Richardson, T. S., Robins, J. M., and Wang, L. (2017). On modeling and estimation for the relative risk and risk difference. *Journal of the American Statistical Association*, 112(519):1121–1130.

Robins, J. (1986). A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical Modelling*, 7(9):1393–1512.

Robins, J. M., Mark, S. D., and Newey, W. (1992). Estimating exposure effects by modelling the expectation of exposure conditional on confounders. *Biometrics*, 48 2:479–95.

Rothman, K. J. (2011). *Epidemiology: an introduction*. Oxford University Press, 2 edition.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational psychology*, 66(5):688–701.

Sato, T. (1992). Maximum likelihood estimation of the risk ratio in case-cohort studies. *Biometrics*, pages 1215–1221.

Schechtman, E. (2002). Odds ratio, relative risk, absolute risk reduction, and the number needed to treat—which of these should we use? *Value in health*, 5(5):431–436.

Schouten, E. G., Dekker, J. M., Kok, F. J., Cessie, S. L., Van Houwelingen, H. C., Pool, J., and Vandnbroucke, J. P. (1993). Risk ratio and rate ratio estimation in case-cohort designs: Hypertension and cardiovascular mortality. *Statistics in medicine*, 12(18):1733–1745.

Seber, G. A. and Lee, A. J. (2012). *Linear regression analysis*. John Wiley & Sons.

Shirvaikar, V. and Holmes, C. (2023). Targeting relative risk heterogeneity with causal forests.

Sinclair, J. C. and Bracken, M. B. (1994). Clinically useful measures of effect in binary analyses of randomized trials. *Journal of clinical epidemiology*, 47(8):881–889.

Splawa-Neyman, J., Dabrowska, D. M., and Speed, T. P. (1990). On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9. *Statistical Science*, 5(4):465 – 472.

Staus, P., von Cube, M., Hazard, D., Doerken, S., Ershova, K., Balmford, J., and Wolkewitz, M. (2022). Inverse probability weighting enhances absolute risk estimation in three common study designs of nosocomial infections. *Clinical Epidemiology*, pages 1053–1064.

Stefanski, L. A. and Boos, D. D. (2002). The calculus of m-estimation. *The American Statistician*, 56(1):29–38.

Tsiatis, A. (2006). Semiparametric theory and missing data. 73.

Vaart, A. W. v. d. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Wager, S. (2020). Stats 361: Causal inference. Technical report, Technical report, Stanford University, 2020. URL: https://web. stanford. edu ....

Wang, W. and Shan, G. (2015). Exact confidence intervals for the relative risk and the odds ratio. *Biometrics*, 71(4):985–995.

Yadlowsky, S., Pellegrini, F., Lionetto, F., Braune, S., and Tian, L. (2021). Estimation and validation of ratio-based conditional average treatment effects using observational data. *Journal of the American Statistical Association*, 116(533):335–352.

# Supplementary Materials

## Contents

# 6 Proofs

## 6.1 Preliminary results

Since we are studying the asymptotic properties of the risk ratio, we cannot directly apply a central limit theorem as in Wager (2020). We will therefore rely on Theorem 1 to prove most of our asymptotic results.

**Theorem 1 (Asymptotic normality of the ratio of two estimators).** *Let $(Z_1, \dots, Z_n)$ be $n$ i.i.d. random variables, $g_0$ and $g_1$ two functions square integrable such that $\mathbb{E}[g_0(Z_i)] = \tau_0$ and $\mathbb{E}[g_1(Z_i)] = \tau_1$, where $\tau_0 \neq 0$. Then, we have that*

$$\sqrt{n}\left(\frac{\sum_{i=1}^n g_1(Z_i)}{\sum_{i=1}^n g_0(Z_i)} - \frac{\tau_1}{\tau_0}\right) \xrightarrow{d} \mathcal{N}\left(0, V_{RR}^{\star}\right),$$

*where*

$$V_{RR}^{\star} = \left(\frac{\tau_1}{\tau_0}\right)^2 \operatorname{Var}\left(\frac{g_1(Z)}{\tau_1} - \frac{g_0(Z)}{\tau_0}\right).$$

*Proof.* We rely on M-estimation theory to prove Theorem 1. Let

$$\hat{\boldsymbol{\theta}}_n = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n g_0(Z_i) \\ \frac{1}{n}\sum_{i=1}^n g_1(Z_i) \\ \frac{\sum_{i=1}^n g_1(Z_i)}{\sum_{i=1}^n g_0(Z_i)} \end{pmatrix} \quad \text{and} \quad \psi(Z, \boldsymbol{\theta}) = \begin{pmatrix} g_0(Z) - \theta_0 \\ g_1(Z) - \theta_1 \\ \theta_1 - \theta_2\theta_0 \end{pmatrix}, \tag{18}$$

where $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2)$. We have that

$$\sum_{i=1}^n \left(g_0(Z_i) - \frac{1}{n}\sum_{j=1}^n g_0(Z_j)\right) = \sum_{i=1}^n g_0(Z_i) - \sum_{j=1}^n g_0(Z_j) = 0,$$

and similarly

$$\sum_{i=1}^n \left(g_1(Z_i) - \frac{1}{n}\sum_{j=1}^n g_1(Z_j)\right) = \sum_{i=1}^n g_1(Z_i) - \sum_{j=1}^n g_1(Z_j) = 0.$$

Besides,

$$\sum_{i=1}^n \left(\frac{1}{n}\sum_{j=1}^n g_1(Z_j) - \frac{\sum_{j=1}^n g_1(Z_j)}{\sum_{j=1}^n g_0(Z_j)}\frac{1}{n}\sum_{j=1}^n g_0(Z_j)\right) = 0.$$

Gathering the three previous equalities, we obtain

$$\sum_{i=1}^n \psi(Z_i, \boldsymbol{\theta}_n) = 0, \tag{19}$$

which proves that $\hat{\boldsymbol{\theta}}_n$ is an M-estimator of type $\psi$ (see Stefanski and Boos, 2002). Furthermore, letting $\boldsymbol{\theta}_\infty = (\tau_0, \tau_1, \tau_1/\tau_0)$, simple calculations show that

$$\mathbb{E}\left[\psi(Z, \boldsymbol{\theta}_\infty)\right] = 0. \tag{20}$$

Since the first two components of $\psi$ are linear with respect to $\theta_0$ and $\theta_1$ and since the third component is linear with respect to $\theta_2$, $\boldsymbol{\theta}_\infty$ defined above is the only value satisfying (20). Define

$$A\left(\theta_\infty\right) = \mathbb{E}\left[\frac{\partial \psi}{\partial \theta}|_{\theta=\theta_\infty}\right] \quad \text{and} \quad B(\theta_\infty) = \mathbb{E}\left[\psi(Z,\theta_\infty)\psi(Z,\theta_\infty)^T\right]. \tag{21}$$

We now check the conditions of Theorem 7.2 in Stefanski and Boos (2002). First, let us compute $A\left(\theta_\infty\right)$ and $B\left(\theta_\infty\right)$. Since

$$\frac{\partial \psi}{\partial \theta}(Z,\theta) = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ -\theta_2 & 1 & -\theta_0 \end{pmatrix}, \tag{22}$$

we obtain

$$A\left(\boldsymbol{\theta_\infty}\right) = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ -\frac{\tau_1}{\tau_0} & 1 & -\tau_0 \end{pmatrix}, \tag{23}$$

which leads to

$$A^{-1}\left(\boldsymbol{\theta_\infty}\right) = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ \frac{\tau_1}{\tau_0^2} & -\frac{1}{\tau_0} & -\frac{1}{\tau_0} \end{pmatrix}. \tag{24}$$

Regarding $B\left(\theta_\infty\right)$, elementary calculations show that

$$\psi(Z,\theta_\infty)\psi(Z,\theta_\infty)^T = \begin{pmatrix} \left(g_0(Z)-\tau_0\right)^2 & \left(g_0(Z)-\tau_0\right)\left(g_1(Z)-\tau_1\right) & 0 \\ \left(g_0(Z)-\tau_0\right)\left(g_1(Z)-\tau_1\right) & \left(g_1(Z)-\tau_1\right)^2 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

which leads to

$$B(\theta_\infty) = \begin{pmatrix} \operatorname{Var}\left[g_0(Z)\right] & \operatorname{Cov}\left(g_0(Z), g_1(Z)\right) & 0 \\ \operatorname{Cov}\left(g_0(Z), g_1(Z)\right) & \operatorname{Var}\left[g_1(Z)\right] & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Based on the previous calculations, we have

- $\psi(z, \boldsymbol{\theta})$ and its first two partial derivatives with respect to $\boldsymbol{\theta}$ exist for all $z$ and for all $\boldsymbol{\theta}$ in the neighborhood of $\boldsymbol{\theta_\infty}$.
- For each $\boldsymbol{\theta}$ in the neighborhood of $\boldsymbol{\theta_\infty}$, we have for all $i, j, k \in \{0, 2\}$:

$$\left|\frac{\partial^2}{\partial\theta_i \partial\theta_j}\psi_k(z,\boldsymbol{\theta})\right| \leq 1$$

  and 1 is integrable.
- $A(\theta_\infty)$ exists and is nonsingular.
- $B(\theta_\infty)$ exists and is finite.

Since we have:

$$\sum_{i=1}^{n} \psi(T_i, Y_i, \hat{\boldsymbol{\theta}}_n) = 0 \quad \text{and} \quad \hat{\boldsymbol{\theta}}_n \xrightarrow{p} \theta_\infty.$$

Then the conditions of Theorem 7.2 in Stefanski and Boos (2002) are satisfied, we have

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \theta_\infty\right) \xrightarrow{d} \mathcal{N}\left(0, A(\theta_\infty)^{-1}B(\theta_\infty)(A(\theta_\infty)^{-1})^\top\right),$$

where

$$\begin{aligned} &A(\theta_\infty)^{-1}B(\theta_\infty)(A(\theta_\infty)^{-1})^\top \\ &= \begin{bmatrix} \operatorname{Var}\left[g_0(Z)\right] & \operatorname{Cov}\left(g_0(Z), g_1(Z)\right) & \frac{\operatorname{Cov}(g_0(Z),g_1(Z))}{\tau_0} - \frac{\tau_1 \operatorname{Var}[g_0(Z)]}{\tau_0^2} \\ \operatorname{Cov}\left(g_0(Z), g_1(Z)\right) & \operatorname{Var}\left[g_1(Z)\right] & -\frac{\operatorname{Cov}(g_0(Z),g_1(Z))\tau_1}{\tau_0^2} + \frac{\operatorname{Var}[g_1(Z)]}{\tau_0} \\ \frac{\operatorname{Cov}(g_0(Z),g_1(Z))}{\tau_0} - \frac{\tau_1 \operatorname{Var}[g_0(Z)]}{\tau_0^2} & -\frac{\operatorname{Cov}(g_0(Z),g_1(Z))\tau_1}{\tau_0^2} + \frac{\operatorname{Var}[g_1(Z)]}{\tau_0} & V^\star_{\mathrm{RR}} \end{bmatrix}, \end{aligned} \tag{25}$$

with

$$V^\star_{\mathrm{RR}} = \left(\frac{\tau_1}{\tau_0}\right)^2 \operatorname{Var}\left(\frac{g_1(Z)}{\tau_1} - \frac{g_0(Z)}{\tau_0}\right). \tag{26}$$

In particular,

$$\sqrt{n}\left(\frac{\sum_{i=1}^n g_1(Z_i)}{\sum_{i=1}^n g_0(Z_i)} - \frac{\tau_1}{\tau_0}\right) \xrightarrow{d} \mathcal{N}\left(0, V^\star_{\mathrm{RR}}\right). \tag{27}$$

□

**Theorem 2** (**Finite sample bias and variance of the ratio of two estimators**). *Let* $T_1(\boldsymbol{Z})$ *and* $T_0(\boldsymbol{Z})$ *be two unbiased estimators of* $\tau_1$ *and* $\tau_0 > 0$ *where* $\boldsymbol{Z} = (Z_1, \ldots, Z_n)$ *be* $n$ *i.i.d. random variables. We assume that* $M_0 \geq T_0(\boldsymbol{Z}) \geq m_0 > 0$, $|T_1(\boldsymbol{Z})| \leq M_1$. *We also assume that* $\operatorname{Var}(T_1(\boldsymbol{Z})) = O_p\left(\frac{1}{n}\right)$ *and* $\operatorname{Var}(T_0(\boldsymbol{Z})) = O_p\left(\frac{1}{n}\right)$ *Then, we have that*

$$\operatorname{Bias}\left(\frac{T_1(\boldsymbol{Z})}{T_0(\boldsymbol{Z})}, \frac{\tau_1}{\tau_0}\right) = \left|\mathbb{E}\left[\frac{T_1(\boldsymbol{Z})}{T_0(\boldsymbol{Z})}\right] - \frac{\tau_1}{\tau_0}\right| \leq \frac{M_1 M_0}{n m_0^2}\left(\frac{M_0}{m_0} + 1\right),$$

*and*

$$\left|\operatorname{Var}\left(\frac{T_1(\boldsymbol{Z})}{T_0(\boldsymbol{Z})}\right) - \left(\frac{\tau_1}{\tau_0}\right)^2 \operatorname{Var}\left(\frac{T_1(\boldsymbol{Z})}{\tau_1} - \frac{T_0(\boldsymbol{Z})}{\tau_0}\right)\right| \leq \frac{2M_0M_1}{nm_0^4}\left(\frac{M_0M_1}{m_0^2} + 1\right)$$

.

*Proof.* We rely on the multivariate version of Taylor's theorem to prove Theorem 2. We first introduce the multi-index notation:

$$|\alpha| = \alpha_1 + \cdots + \alpha_n, \quad \alpha! = \alpha_1! \cdots \alpha_n!, \quad \boldsymbol{x}^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$$

and

$$D^\alpha f = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}, \qquad |\alpha| \leq k$$

Let $f$ be the ratio function

$$\begin{aligned} f \colon \mathbb{R}_+^* \times \mathbb{R}_+^* &\longrightarrow \mathbb{R} \\ (x_1, x_2) &\longmapsto x_1/x_2. \end{aligned}$$

Since $f$ is two times continuously differentiable then one can derive an exact formula for the remainder in terms of 2nd order partial derivatives of $f$. Namely, if we define $\boldsymbol{x} = (x_1, x_2)$ and for $\boldsymbol{a} \in \mathbb{R}_+^* \times \mathbb{R}_+^*$

$$f(\boldsymbol{x}) = \sum_{|\alpha| \leq 1} \frac{D^\alpha f(\boldsymbol{a})}{\alpha!}(\boldsymbol{x} - \boldsymbol{a})^\alpha + R_{k+1}(\boldsymbol{x}), \tag{28}$$

with

$$R_{k+1}(\boldsymbol{x}) = \sum_{|\beta|=k+1} (\boldsymbol{x}-\boldsymbol{a})^{\beta} \frac{|\beta|}{\beta!} \int_0^1 (1-t)^{|\beta|-1} D^{\beta} f(\boldsymbol{a}+t(\boldsymbol{x}-\boldsymbol{a}))\, dt.$$

**Bias:**
Computing 28 for the ratio function with $\boldsymbol{x} = (T_1(\boldsymbol{Z}), T_0(\boldsymbol{Z}))$, $\boldsymbol{a} = (\tau_1, \tau_0)$ and taking the expectation gives us:

$$\begin{aligned}
&\mathbb{E}\left[f(T_1(\boldsymbol{Z}), T_0(\boldsymbol{Z}))\right] \\
&= \mathbb{E}\left[f(\tau_1,\tau_0) + \frac{\partial f(\tau_1,\tau_0)}{\partial T_1(\boldsymbol{Z})}(T_1(\boldsymbol{Z})-\tau_1) + \frac{\partial f(\tau_1,\tau_0)}{\partial T_0(\boldsymbol{Z})}(T_0(\boldsymbol{Z})-\tau_0) + R_2(T_1(\boldsymbol{Z}), T_0(\boldsymbol{Z}))\right] \\
&= \mathbb{E}\left[f(\tau_1,\tau_0)\right] + \frac{\partial f(\tau_1,\tau_0)}{\partial T_1(\boldsymbol{Z})}\mathbb{E}\left[(T_1(\boldsymbol{Z})-\tau_1)\right] \\
&+ \frac{\partial f(\tau_1,\tau_0)}{\partial T_0(\boldsymbol{Z})}\mathbb{E}\left[(T_0(\boldsymbol{Z})-\tau_0)\right] + \mathbb{E}\left[R_2(T_1(\boldsymbol{Z}), T_0(\boldsymbol{Z}))\right] \\
&= \frac{\tau_1}{\tau_0} + \mathbb{E}\left[R_2(T_1(\boldsymbol{Z}), T_0(\boldsymbol{Z}))\right]
\end{aligned}$$

In order to produce Theorem 2, we just need to show that $\mathbb{E}\left[R_2(T_1(\boldsymbol{Z}), T_0(\boldsymbol{Z}))\right] = O_p\left(\frac{1}{n}\right)$. To do so, we first compute $R_2(T_1(\boldsymbol{Z}), T_0(\boldsymbol{Z}))$

$$\begin{aligned}
R_2(T_1(\boldsymbol{Z}), T_0(\boldsymbol{Z})) &= \underbrace{2(T_0(\boldsymbol{Z})-\tau_0)^2 \int_0^1 \frac{(1-t)(\tau_1+t(T_1(\boldsymbol{Z})-\tau_1))}{(\tau_0+t(T_0(\boldsymbol{Z})-\tau_0))^3}\, dt}_{R_2^1(T_1(\boldsymbol{Z}),T_0(\boldsymbol{Z}))} \\
&\quad \underbrace{-\, 2(T_0(\boldsymbol{Z})-\tau_0)(T_1(\boldsymbol{Z})-\tau_1)\int_0^1 \frac{1-t}{(\tau_0+t(T_0(\boldsymbol{Z})-\tau_0))^2}\, dt}_{R_2^2(T_1(\boldsymbol{Z}),T_0(\boldsymbol{Z}))}
\end{aligned}$$

Since we assume that $T_0(\boldsymbol{Z}) \geq m_0 > 0$ and that $|T_1(\boldsymbol{Z})| \leq M_1$ we have:

$$\begin{aligned}
\left|R_2^1(T_1(\boldsymbol{Z}), T_0(\boldsymbol{Z}))\right| &= \left|2(T_0(\boldsymbol{Z})-\tau_0)^2 \int_0^1 \frac{(1-t)(\tau_1+t(T_1(\boldsymbol{Z})-\tau_1))}{(\tau_0+t(T_0(\boldsymbol{Z})-\tau_0))^3}\, dt\right| \\
&\leq 2(T_0(\boldsymbol{Z})-\tau_0)^2 \int_0^1 \left|\frac{(1-t)\max(\tau_1, M_1)}{\min(m_0,\tau_0)^3}\right| dt \\
&= (T_0(\boldsymbol{Z})-\tau_0)^2 \underbrace{\frac{M_1}{m_0^3}}_{C_1}
\end{aligned}$$

Similarly, we have:

$$\begin{aligned}
\left|R_2^2(T_1(\boldsymbol{Z}), T_0(\boldsymbol{Z}))\right| &= \left|2(T_0(\boldsymbol{Z})-\tau_0)(T_1(\boldsymbol{Z})-\tau_1)\int_0^1 \frac{1-t}{(\tau_0+t(T_0(\boldsymbol{Z})-\tau_0))^2}\, dt\right| \\
&= 2\left|(T_0(\boldsymbol{Z})-\tau_0)(T_1(\boldsymbol{Z})-\tau_1)\right| \left|\int_0^1 \frac{1-t}{(\tau_0+t(T_0(\boldsymbol{Z})-\tau_0))^2}\, dt\right| \\
&\leq \left|(T_0(\boldsymbol{Z})-\tau_0)(T_1(\boldsymbol{Z})-\tau_1)\right| \underbrace{\frac{1}{m_0^2}}_{C_2}
\end{aligned}$$

Finally we get that:

$$\begin{aligned}
\left|\mathbb{E}\left[R_2(T_1(\boldsymbol{Z}), T_0(\boldsymbol{Z}))\right]\right| &\leq \mathbb{E}\left[\left|R_2^1(T_1(\boldsymbol{Z}), T_0(\boldsymbol{Z}))\right| + \left|R_2^2(T_1(\boldsymbol{Z}), T_0(\boldsymbol{Z}))\right|\right] \\
&\leq C_1 \operatorname{Var}(T_0(\boldsymbol{Z})) + C_2 \mathbb{E}\left[\left|(T_0(\boldsymbol{Z})-\tau_0)(T_1(\boldsymbol{Z})-\tau_1)\right|\right] \\
&\leq C_1 \operatorname{Var}(T_0(\boldsymbol{Z})) + C_2 \sqrt{\operatorname{Var}(T_0(\boldsymbol{Z}))\operatorname{Var}(T_1(\boldsymbol{Z}))} \\
&\leq C_1 M_0^2 + C_2 M_0 M_1
\end{aligned}$$

Since we have that $\mathrm{Var}(T_1(\boldsymbol{Z})) = O_p\left(\frac{1}{n}\right)$ and $\mathrm{Var}(T_0(\boldsymbol{Z})) = O_p\left(\frac{1}{n}\right)$, we can conclude by:

$$\mathrm{Bias}\left(\frac{T_1(\boldsymbol{Z})}{T_0(\boldsymbol{Z})}, \frac{\tau_1}{\tau_0}\right) = \left|\mathbb{E}\left[\frac{T_1(\boldsymbol{Z})}{T_0(\boldsymbol{Z})}\right] - \frac{\tau_1}{\tau_0}\right| \lesssim \frac{M_1 M_0}{n m_0^2}\left(\frac{M_0}{m_0} + 1\right)$$

**Variance:**
Let us begin by expanding the variance of the function $f$:

$$\mathrm{Var}\, f(T_1(\boldsymbol{Z}), T_0(\boldsymbol{Z})) = \mathbb{E}\left[(f(T_1(\boldsymbol{Z}), T_0(\boldsymbol{Z})) - \mathbb{E}\left[f(T_1(\boldsymbol{Z}), T_0(\boldsymbol{Z}))\right])^2\right]$$

Next, apply Taylor's expansion around the means $\tau_1$ and $\tau_0$:

$$\begin{aligned} &= \mathbb{E}\Bigg[(f(\tau_1, \tau_0) + \frac{\partial f(\tau_1, \tau_0)}{\partial T_1(\boldsymbol{Z})}(T_1(\boldsymbol{Z}) - \tau_1) + \frac{\partial f(\tau_1, \tau_0)}{\partial T_0(\boldsymbol{Z})}(T_0(\boldsymbol{Z}) - \tau_0) \\ &\quad + R_2(T_1(\boldsymbol{Z}), T_0(\boldsymbol{Z})) - \mathbb{E}\left[f(T_1(\boldsymbol{Z}), T_0(\boldsymbol{Z}))\right])^2\Bigg] \end{aligned}$$

Simplify by focusing on the first-order derivatives and residual terms:

$$\begin{aligned} &= \mathbb{E}\Bigg[(\frac{\partial f(\tau_1, \tau_0)}{\partial T_1(\boldsymbol{Z})}(T_1(\boldsymbol{Z}) - \tau_1) + \frac{\partial f(\tau_1, \tau_0)}{\partial T_0(\boldsymbol{Z})}(T_0(\boldsymbol{Z}) - \tau_0) \\ &\quad + R_2(T_1(\boldsymbol{Z}), T_0(\boldsymbol{Z})) - \mathbb{E}\left[R_2(T_1(\boldsymbol{Z}), T_0(\boldsymbol{Z}))\right])^2\Bigg] \end{aligned}$$

Decompose the variance into linear, cross-term, and residual contributions:

$$\begin{aligned} &= \mathbb{E}\left[(\frac{\partial f(\tau_1, \tau_0)}{\partial T_1(\boldsymbol{Z})}(T_1(\boldsymbol{Z}) - \tau_1) + \frac{\partial f(\tau_1, \tau_0)}{\partial T_0(\boldsymbol{Z})}(T_0(\boldsymbol{Z}) - \tau_0))^2\right] \\ &+ 2\mathbb{E}\Bigg[\frac{\partial f(\tau_1, \tau_0)}{\partial T_1(\boldsymbol{Z})}(T_1(\boldsymbol{Z}) - \tau_1) \\ &\quad + \frac{\partial f(\tau_1, \tau_0)}{\partial T_0(\boldsymbol{Z})}(T_0(\boldsymbol{Z}) - \tau_0)\Bigg]\mathbb{E}\left[R_2(T_1(\boldsymbol{Z}), T_0(\boldsymbol{Z})) - \mathbb{E}\left[R_2(T_1(\boldsymbol{Z}), T_0(\boldsymbol{Z}))\right]\right] \\ &+ \mathrm{Var}(R_2(T_1(\boldsymbol{Z}), T_0(\boldsymbol{Z}))) \end{aligned}$$

Finally, re-express the result using a simplified ratio of variances:

$$= \left(\frac{\tau_1}{\tau_0}\right)^2 \mathrm{Var}\left(\frac{T_1(\boldsymbol{Z})}{\tau_1} - \frac{T_0(\boldsymbol{Z})}{\tau_0}\right) + \mathrm{Var}(R_2(T_1(\boldsymbol{Z}), T_0(\boldsymbol{Z}))$$

We now focus on $\mathrm{Var}(R_2(T_1(\boldsymbol{Z}), T_0(\boldsymbol{Z})))$:

$$\begin{aligned} \mathrm{Var}(R_2(T_1(\boldsymbol{Z}), T_0(\boldsymbol{Z}))) &\leq \mathbb{E}\left[(R_2(T_1(\boldsymbol{Z}), T_0(\boldsymbol{Z})))^2\right] \\ &\leq 2\mathbb{E}\left[|R_2^1(T_1(\boldsymbol{Z}), T_0(\boldsymbol{Z}))|^2\right] + 2\mathbb{E}\left[|R_2^2(T_1(\boldsymbol{Z}), T_0(\boldsymbol{Z}))|^2\right] \end{aligned}$$

We first focus on the first term:

$$\begin{aligned} \mathbb{E}\left[|R_2^1(T_1(\boldsymbol{Z}), T_0(\boldsymbol{Z}))|^2\right] &\leq \mathbb{E}\left[\left(C_{02}(T_0(\boldsymbol{Z}) - \tau_0)^2\right)^2\right] \\ &\leq C_1^2 \mathbb{E}\left[(T_0(\boldsymbol{Z}) - \tau_0)^4\right] \\ &\leq C_1^2 M_0^2 \mathbb{E}\left[(T_0(\boldsymbol{Z}) - \tau_0)^2\right] \quad T_0(\boldsymbol{Z}) \leq M_0 \\ &\leq C_1^2 M_0^2\, \mathrm{Var}(T_0(\boldsymbol{Z})) \end{aligned}$$

For the second term we have:

$$\begin{aligned} \mathbb{E}\left[|R_2^2(T_1(\boldsymbol{Z}), T_0(\boldsymbol{Z}))|^2\right] &= C_2^2 \mathbb{E}\left[(T_0(\boldsymbol{Z}) - \tau_0)^2 (T_1(\boldsymbol{Z}) - \tau_1)^2\right] \\ &\leq C_2^2 \sqrt{\mathbb{E}\left[(T_0(\boldsymbol{Z}) - \tau_0)^4\right] \mathbb{E}\left[(T_1(\boldsymbol{Z}) - \tau_1)^4\right]} \\ T_1(\boldsymbol{Z}), T_0(\boldsymbol{Z}) \text{ bounded} \quad &\leq C_2^2 M_0 M_1 \sqrt{\mathbb{E}\left[(T_0(\boldsymbol{Z}) - \tau_0)^2\right] \mathbb{E}\left[(T_1(\boldsymbol{Z}) - \tau_1)^2\right]} \\ &\leq C_2^2 M_0 M_1 \sqrt{\mathrm{Var}(T_0(\boldsymbol{Z}))\, \mathrm{Var}(T_1(\boldsymbol{Z}))} \end{aligned}$$

Hence we get that:

$$\mathrm{Var}(R_2(T_1(\boldsymbol{Z}), T_0(\boldsymbol{Z}))) \lesssim \frac{2M_0M_1}{nm_0^4}\left(\frac{M_0M_1}{m_0^2} + 1\right)$$

□

### 6.2 Proofs of Section 2

#### 6.2.1 Risk Ratio Neyman estimator

*Proof of Proposition 1.*
**Asymptotic Bias and Variance:** we proceed with M-estimations to prove asymptotic bias and variance of the Ratio Neyman estimator, we first define the following:

$$\hat{\boldsymbol{\theta}}_n = \begin{pmatrix} \frac{1}{n_0}\sum_{T_i=0} Y_i \\ \frac{1}{n_1}\sum_{T_i=1} Y_i \\ \hat{\tau}_{R-N,n} \end{pmatrix} \quad \text{and} \quad \psi(T, Y, \boldsymbol{\theta}) = \begin{pmatrix} \psi_0(\boldsymbol{\theta}) \\ \psi_1(\boldsymbol{\theta}) \\ \psi_2(\boldsymbol{\theta}) \end{pmatrix} =: \begin{pmatrix} (1-T)\,(Y-\theta_0) \\ T\,(Y-\theta_1) \\ \theta_1 - \theta_2\theta_0 \end{pmatrix}, \tag{29}$$

where $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2)$.

Next, we verify that for $\hat{\boldsymbol{\theta}}_n = (\frac{1}{n_0}\sum_{T_i=0} Y_i, \frac{1}{n_1}\sum_{T_i=1} Y_i, \hat{\tau}_{R-N,n})$, we have:

$$\sum_{i=1}^{n} \psi(T_i, Y_i, \hat{\boldsymbol{\theta}}_n) = 0.$$

We begin by demonstrating this for $\psi_1$:

$$\begin{aligned} \sum_{i=1}^{n} \psi_1(T_i, Y_i, \hat{\boldsymbol{\theta}}_n) &= \sum_{i=1}^{n} T_i \left(Y_i - \frac{1}{n_1}\sum_{T_j=1} Y_j\right) \\ &= \sum_{i=1}^{n} T_i \left(Y_i - \frac{1}{n_1}\sum_{j=1}^{n} T_j Y_j\right) \\ &= \sum_{i=1}^{n} T_i Y_i - \frac{1}{n_1}\sum_{i=1}^{n} T_i \sum_{j=1}^{n} T_j Y_j \\ &= \sum_{i=1}^{n} T_i Y_i - \sum_{j=1}^{n} T_j Y_j \\ &= 0. \end{aligned}$$

Similarly, we can show:

$$\sum_{i=1}^{n} \psi_0(T_i, Y_i, \hat{\boldsymbol{\theta}}_n) = 0.$$

Moreover, by construction:

$$\sum_{i=1}^{n} \psi_2(T_i, Y_i, \hat{\boldsymbol{\theta}}_n) = 0.$$

Thus, we have established that $\hat{\boldsymbol{\theta}}_n$ is an M-estimator of type $\psi$ (see Stefanski and Boos, 2002). Given that we are in a Bernoulli Trial, we now demonstrate that $\mathbb{E}\left[\psi(T, Y, \theta_\infty)\right] = 0$ where $\theta_\infty = (\mathbb{E}[Y^{(0)}], \mathbb{E}[Y^{(1)}], \tau_{RR})$. Therefore, we have:

$$\begin{aligned}
\mathbb{E}\left[\psi_1(\theta_\infty)\right] &= \mathbb{E}\left[T\left(Y - \mathbb{E}[Y^{(1)}]\right)\right] \\
&= \mathbb{E}\left[T\left(Y^{(1)} - \mathbb{E}[Y^{(1)}]\right)\right] && \text{(by SUTVA)} \\
&= \mathbb{E}\left[T\right]\mathbb{E}\left[Y^{(1)} - \mathbb{E}[Y^{(1)}]\right] && \text{(by ignorability)} \\
&= 0.
\end{aligned}$$

Similarly, we can show:

$$\mathbb{E}\left[\psi_0(\theta_\infty)\right] = 0.$$

Furthermore, we have:

$$\mathbb{E}\left[\psi_2(\theta_\infty)\right] = \mathbb{E}[Y^{(1)}] - \tau_{RR}\mathbb{E}[Y^{(0)}] = 0.$$

At this point, we note that $\theta_\infty$ is the only value of $\boldsymbol{\theta}$ such that $\mathbb{E}\left[\psi(T, Y, \boldsymbol{\theta})\right] = 0$. We proceed by defining:

$$A\left(\theta_\infty\right) = \mathbb{E}\left[\frac{\partial \psi}{\partial \theta}|_{\theta=\theta_\infty}\right] \quad \text{and} \quad B(\theta_\infty) = \mathbb{E}\left[\psi(Z, \theta_\infty)\psi(Z, \theta_\infty)^T\right].$$

Next, we check the conditions of Theorem 7.2 in Stefanski and Boos (2002). First, we compute $A\left(\theta_\infty\right)$ and $B\left(\theta_\infty\right)$. Since:

$$\frac{\partial \psi}{\partial \theta}(Z, \theta) = \begin{pmatrix} -(1-T) & 0 & 0 \\ 0 & -T & 0 \\ -\theta_2 & 1 & -\theta_0 \end{pmatrix},$$

we obtain:

$$A\left(\boldsymbol{\theta_\infty}\right) = \begin{pmatrix} -(1-e) & 0 & 0 \\ 0 & -e & 0 \\ -\tau_{RR} & 1 & -\mathbb{E}[Y^{(0)}] \end{pmatrix},$$

which leads to:

$$A^{-1}\left(\boldsymbol{\theta_\infty}\right) = \begin{pmatrix} \frac{1}{e-1} & 0 & 0 \\ 0 & -\frac{1}{e} & 0 \\ \tau_{RR}\frac{1}{\mathbb{E}[Y^{(0)}](1-e)} & -\frac{1}{e\mathbb{E}[Y^{(0)}]} & -\frac{1}{\mathbb{E}[Y^{(0)}]} \end{pmatrix}.$$

Regarding $B\left(\theta_\infty\right)$, elementary calculations show that:

$$\begin{aligned}
&\psi(Z, \theta_\infty)\psi(Z, \theta_\infty)^T \\
&= \begin{pmatrix} \left((1-T)(Y - \mathbb{E}[Y^{(0)}])\right)^2 & (1-T)(Y - \mathbb{E}[Y^{(0)}])T(Y - \mathbb{E}[Y^{(1)}]) & 0 \\ (1-T)(Y - \mathbb{E}[Y^{(0)}])T(Y - \mathbb{E}[Y^{(1)}]) & \left(T(Y - \mathbb{E}[Y^{(1)}])\right)^2 & 0 \\ 0 & 0 & 0 \end{pmatrix},
\end{aligned}$$

which leads to:

$$B(\theta_\infty) = \begin{pmatrix} (1-e)\operatorname{Var}\left[Y^{(0)}\right] & 0 & 0 \\ 0 & e\operatorname{Var}\left[Y^{(1)}\right] & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Based on the previous calculations, we have:

- $\psi(z, \boldsymbol{\theta})$ and its first two partial derivatives with respect to $\boldsymbol{\theta}$ exist for all $z$ and for all $\boldsymbol{\theta}$ in the neighborhood of $\boldsymbol{\theta_\infty}$.

- For each $\boldsymbol{\theta}$ in the neighborhood of $\boldsymbol{\theta_\infty}$, we have for all $i, j, k \in \{0, 2\}$:

$$\left|\frac{\partial^2}{\partial\theta_i\partial\theta_j}\psi_k(z,\boldsymbol{\theta})\right| \leq 1$$

  and 1 is integrable.

- $A(\theta_\infty)$ exists and is nonsingular.
- $B(\theta_\infty)$ exists and is finite.

Since we have:

$$\sum_{i=1}^{n}\psi(T_i, Y_i, \hat{\boldsymbol{\theta}}_n) = 0 \quad \text{and} \quad \hat{\boldsymbol{\theta}}_n \xrightarrow{p} \theta_\infty.$$

Then, the conditions of Theorem 7.2 in Stefanski and Boos (2002) are satisfied, we have:

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \theta_\infty\right) \xrightarrow{d} \mathcal{N}\left(0, A(\theta_\infty)^{-1}B(\theta_\infty)(A(\theta_\infty)^{-1})^\top\right),$$

where:

$$A(\theta_\infty)^{-1}B(\theta_\infty)(A(\theta_\infty)^{-1})^\top = \begin{bmatrix} \frac{\mathrm{Var}\left[Y^{(0)}\right]}{(1-e)} & 0 & -\frac{\tau\,\mathrm{Var}\left[Y^{(0)}\right]}{\tau_0(1-e)} \\ 0 & \frac{\mathrm{Var}\left[Y^{(1)}\right]}{e} & \frac{\mathrm{Var}\left[Y^{(1)}\right]}{e\tau_0} \\ -\frac{\tau\,\mathrm{Var}\left[Y^{(0)}\right]}{\tau_0(1-e)} & \frac{\mathrm{Var}\left[Y^{(1)}\right]}{e\tau_0} & V_{R-N} \end{bmatrix},$$

with:

$$V_{R-N} = \tau_{RR}^2\left(\frac{\mathrm{Var}\left(Y^{(1)}\right)}{e\mathbb{E}[Y^{(1)}]^2} + \frac{\mathrm{Var}\left(Y^{(0)}\right)}{(1-e)\mathbb{E}[Y^{(0)}]^2}\right).$$

In particular, we obtain:

$$\sqrt{n}\left(\hat{\tau}_{\mathrm{RR,N,n}} - \tau_{\mathrm{RR}}\right) \xrightarrow{d} \mathcal{N}\left(0, V_{\mathrm{RR,N}}\right).$$

Finally, note that:

$$\begin{aligned} V_{R-N} &= \tau_{RR}^2\left(\frac{\mathrm{Var}\left(Y^{(1)}\right)}{e\mathbb{E}[Y^{(1)}]^2} + \frac{\mathrm{Var}\left(Y^{(0)}\right)}{(1-e)\mathbb{E}[Y^{(0)}]^2}\right) \\ &= \tau_{RR}^2\left(\frac{\mathbb{E}[(Y^{(1)})^2] - \mathbb{E}[Y^{(1)}]^2}{e\mathbb{E}[Y^{(1)}]^2} + \frac{\mathbb{E}[(Y^{(0)})^2] - \mathbb{E}[Y^{(0)}]^2}{(1-e)\mathbb{E}[Y^{(0)}]^2}\right) \\ &= V_{R-HT} - \frac{\tau_{RR}^2}{e(1-e)}. \end{aligned}$$

As a consequence an estimator $\hat{V}_{R-N}$ can be derived :

$$\hat{V}_{R-N} = \hat{\tau}_{\mathrm{RR,N,n}}^2\left(\frac{\frac{1}{n}\sum_{T_i=1}\left(Y_i - \frac{1}{n}\sum_{T_i=1}Y_i\right)^2}{\hat{e}\left(\frac{1}{n}\sum_{T_i=1}Y_i\right)^2} + \frac{\frac{1}{n}\sum_{T_i=0}\left(Y_i - \frac{1}{n}\sum_{T_i=0}Y_i\right)^2}{(1-\hat{e})\left(\frac{1}{n}\sum_{T_i=0}Y_i\right)^2}\right) \tag{30}$$

**Optimal choice of $e$:** the optimal value of $e_{opt}$ is the one that minimizes the variance of the Ratio Neyman estimator. Therefore, we need to solve:

$$\inf_{e\in(0,1)} \tau_{\mathrm{RR}}^2\left(\frac{\mathrm{Var}\left(Y^{(1)}\right)}{e\mathbb{E}\left[Y^{(1)}\right]^2} + \frac{\mathrm{Var}\left(Y^{(0)}\right)}{(1-e)\mathbb{E}\left[Y^{(0)}\right]^2}\right)$$

Noting that the variance we want to minimize is convex in $e$, we can derive the variance and set it to 0 to find $e_{opt}$. We have:

$$\frac{C_1}{e_{opt}^2} = \frac{C_0}{(1-e_{opt})^2}$$

where $C_1 = \frac{\mathrm{Var}\left(Y^{(1)}\right)}{\mathbb{E}\left[Y^{(1)}\right]^2}$ and $C_0 = \frac{\mathrm{Var}\left(Y^{(0)}\right)}{\mathbb{E}\left[Y^{(0)}\right]^2}$.

- If $\frac{\mathrm{Var}\left(Y^{(1)}\right)}{\mathbb{E}[Y^{(1)}]^2} = \frac{\mathrm{Var}\left(Y^{(0)}\right)}{\mathbb{E}[Y^{(0)}]^2}$:

$$e_{opt} = 0.5$$

- otherwise:

$$e_{opt} = \frac{C_1 - \sqrt{C_1 C_0}}{C_1 - C_0} \in (0,1)$$

□

#### 6.2.2 Risk Ratio Horvitz-Thomson estimator

**Definition 6** (**Risk Ratio Horvitz-Thomson estimator**). *Grant Assumption 1 and Assumption 2. The Risk Ratio Horvitz-Thomson estimator denoted $\hat{\tau}_{RR,HT,n}$ is defined as,*

$$\hat{\tau}_{RR,HT,n} = \frac{\sum_{i=1}^{n} \frac{T_i Y_i}{e}}{\sum_{i=1}^{n} \frac{(1-T_i)Y_i}{1-e}} \tag{31}$$

*if $\sum_{i=1}^{n} T_i < n$ and 0 otherwise.*

Within the context of a Bernoulli trial, Proposition 7 proves that the Risk Ratio Horvitz-Thompson estimator is asymptotically unbiased and normally distributed.

**Proposition 7** (**Asymptotic normality of $\hat{\tau}_{\mathbf{RR,HT,n}}$**). *Under Assumption 1 and Assumption 2, the Risk Ratio Horvitz-Thompson estimator is asymptotically unbiased and satisfies*

$$\sqrt{n}\left(\hat{\tau}_{RR,HT,n} - \tau_{RR}\right) \xrightarrow{d} \mathcal{N}\left(0, V_{RR,HT}\right) \tag{32}$$

*where* $V_{RR,HT} = \tau_{RR}^2 \left( \frac{\mathbb{E}\left[\left(Y^{(1)}\right)^2\right]}{e\mathbb{E}\left[Y^{(1)}\right]^2} + \frac{\mathbb{E}\left[\left(Y^{(0)}\right)^2\right]}{(1-e)\mathbb{E}\left[Y^{(0)}\right]^2} \right)$.

*If we assume that for all $i$, $M \geq Y_i \geq m > 0$ and $0 < \sum_{i=1}^{n} T_i < n$, we also have:*

$$|Bias(\hat{\tau}_{RR,\ HT,\ n}| \leq \frac{2M^3(1-e)^3}{nm^3e^3}$$

$$|\mathrm{Var}(\hat{\tau}_{RR,\ HT,\ n})| \leq \frac{4M^4(1-e)^6}{nm^6e^4}$$

*Proof of Proposition 7.*

**Asymptotic Bias and Variance.** Let $Z_i := (T_i, Y_i)$ and define $g_0(Z_i) = \frac{(1-T_i)Y_i}{1-e}$ and $g_1(Z_i) = \frac{T_i Y_i}{e}$. First, we evaluate the expectation of $g_1(Z_i)$:

$$\begin{aligned}
\mathbb{E}\left[g_1(Z_i)\right] &= \mathbb{E}\left[\frac{T_i Y_i}{e}\right] && \text{(by i.i.d)} \\
&= \mathbb{E}\left[\frac{T_i Y_i^{(1)}}{e}\right] && \text{(by SUTVA)} \\
&= \mathbb{E}\left[\frac{T_i}{e}\right] \mathbb{E}\left[Y_i^{(1)}\right] && \text{(by ignorability)} \\
&= \mathbb{E}\left[Y_i^{(1)}\right] && \text{(by Trial positivity)}
\end{aligned}$$

Similarly, we can find the expectation of $g_0(Z_i)$:

$$\mathbb{E}\left[g_0(Z_i)\right] = \mathbb{E}\left[Y^{(0)}\right] > 0.$$

Thus, according to Theorem 1, we have $\sqrt{n}\left(\hat{\tau}_{\text{RR-HT, n}} - \tau_{\text{RR}}\right) \xrightarrow{d} \mathcal{N}\left(0, V_{\text{RR-HT}}\right)$, with

$$\begin{aligned} V_{\text{RR-HT}} &= \left(\frac{\tau_1}{\tau_0}\right)^2 \operatorname{Var}\left(\frac{g_1(Z)}{\tau_1} - \frac{g_0(Z)}{\tau_0}\right) \\ &= \tau_{\text{RR}}^2 \operatorname{Var}\left(\frac{TY}{e\mathbb{E}\left[Y^{(1)}\right]} - \frac{(1-T)Y}{(1-e)\mathbb{E}\left[Y^{(0)}\right]}\right). \end{aligned}$$

Next, we evaluate the variance terms separately:

$$\begin{aligned} \operatorname{Var}\left(\frac{TY}{e\mathbb{E}\left[Y^{(1)}\right]}\right) &= \frac{1}{\mathbb{E}\left[Y^{(1)}\right]^2 e^2} \operatorname{Var}\left(TY\right) \\ &= \frac{1}{\mathbb{E}\left[Y^{(1)}\right]^2 e^2} \left(\mathbb{E}\left[(TY)^2\right] - \mathbb{E}\left[TY\right]^2\right) \\ &= \frac{1}{\mathbb{E}\left[Y^{(1)}\right]^2 e^2} \left(\mathbb{E}\left[T\left(Y\right)^2\right] - \mathbb{E}\left[TY\right]^2\right) && (T \text{ is binary}) \\ &= \frac{1}{\mathbb{E}\left[Y^{(1)}\right]^2 e^2} \left(\mathbb{E}\left[T\left(Y^{(1)}\right)^2\right] - \mathbb{E}\left[TY^{(1)}\right]^2\right) && (\text{by SUTVA}) \\ &= \frac{1}{\mathbb{E}\left[Y^{(1)}\right]^2 e^2} \left(e\mathbb{E}\left[\left(Y^{(1)}\right)^2\right] - e^2\mathbb{E}\left[Y^{(1)}\right]^2\right) && (\text{by ignorability}) \\ &= \frac{\mathbb{E}\left[\left(Y^{(1)}\right)^2\right]}{e\mathbb{E}\left[Y^{(1)}\right]^2} - 1. \end{aligned}$$

Similarly, we find the variance of the second term:

$$\operatorname{Var}\left(\frac{(1-T)Y}{(1-e)\mathbb{E}\left[Y^{(0)}\right]}\right) = \frac{\mathbb{E}\left[\left(Y^{(0)}\right)^2\right]}{(1-e)\mathbb{E}\left[Y^{(0)}\right]^2} - 1.$$

Finally, we compute the covariance between the two terms:

$$\begin{aligned} \operatorname{Cov}\left(\frac{TY}{e\mathbb{E}\left[Y^{(1)}\right]}, \frac{(1-T)Y}{(1-e)\mathbb{E}\left[Y^{(0)}\right]}\right) &= \frac{\operatorname{Cov}(TY, (1-T)Y)}{e\mathbb{E}\left[Y^{(1)}\right](1-e)\mathbb{E}\left[Y^{(0)}\right]} \\ &= \frac{\left(\mathbb{E}\left[T(1-T)Y^2\right] - \mathbb{E}\left[TY\right]\mathbb{E}\left[(1-T)Y\right]\right)}{e\mathbb{E}\left[Y^{(1)}\right](1-e)\mathbb{E}\left[Y^{(0)}\right]} \\ &= \frac{-\mathbb{E}\left[TY\right]\mathbb{E}\left[(1-T)Y\right]}{e\mathbb{E}\left[Y^{(1)}\right](1-e)\mathbb{E}\left[Y^{(0)}\right]} \\ &= -1. \end{aligned}$$

Using Bienayme's identity, we finally obtain:

$$V_{\text{RR-HT}} = \tau_{\text{RR}}^2 \left(\frac{\mathbb{E}\left[\left(Y^{(1)}\right)^2\right]}{e\mathbb{E}\left[Y^{(1)}\right]^2} + \frac{\mathbb{E}\left[\left(Y^{(0)}\right)^2\right]}{(1-e)\mathbb{E}\left[Y^{(0)}\right]^2}\right).$$

As a consequence an estimator $\hat{V}_{RR-HT}$ can be derived:

$$\hat{V}_{RR-HT} = \hat{\tau}^2_{\text{RR,HT,n}} \left( \frac{\frac{1}{n}\sum_{T_i=1} Y_i^2}{\hat{e}\left(\frac{1}{n}\sum_{T_i=1} Y_i\right)^2} + \frac{\frac{1}{n}\sum_{T_i=0} Y_i^2}{(1-\hat{e})\left(\frac{1}{n}\sum_{T_i=0} Y_i\right)^2} \right) \tag{33}$$

**Finite sample Bias and Variance.** Let $T_1(\boldsymbol{Z}) = \frac{1}{n}\sum_{i=1}^n \frac{T_i Y_i}{e}$ and $T_0(\boldsymbol{Z}) = \frac{1}{n}\sum_{i=1}^n \frac{(1-T_i)Y_i}{1-e}$ where $\boldsymbol{Z} = (Z_1, \dots, Z_n)$. First, consider the variance of $T_1(\boldsymbol{Z})$:

$$\begin{aligned}
\operatorname{Var}(T_1(\boldsymbol{Z})) &= \frac{1}{ne^2}\operatorname{Var}(T_iY_i) && \text{(by i.i.d)}\\
&= \frac{1}{ne^2}\left(\mathbb{E}\left[(T_iY_i)^2\right] - \mathbb{E}\left[T_iY_i\right]^2\right)\\
&= \frac{\mathbb{E}\left[\left(Y^{(1)}\right)^2\right] - e\mathbb{E}\left[Y^{(1)}\right]^2}{ne}.
\end{aligned}$$

Thus $\operatorname{Var}(T_1(\boldsymbol{Z})) = O_p(1/n)$ and similarly $\operatorname{Var}(T_0(\boldsymbol{Z})) = O_p(1/n)$. Next, we show that $T_0(\boldsymbol{Z})$ is bounded:

$$\begin{aligned}
T_0(\boldsymbol{Z}) &= \frac{1}{n}\sum_{i=1}^n \frac{(1-T_i)Y_i}{1-e}\\
&= \frac{1}{n(1-e)}\sum_{i=1}^n (1-T_i)Y_i\\
&\geq \frac{m}{(1-e)}\sum_{i=1}^n (1-T_i) && \text{(since } Y_i \geq m > 0\text{)}\\
&\geq \frac{m}{(1-e)} && \text{(as } \sum_{i=1}^n T_i < n\text{).}
\end{aligned}$$

Similarly, we also have the upper bound

$$\begin{aligned}
T_0(\boldsymbol{Z}) &= \frac{1}{n}\sum_{i=1}^n \frac{(1-T_i)Y_i}{1-e}\\
&= \frac{1}{ne}\sum_{i=1}^n (1-T_i)Y_i\\
&\leq \frac{1}{ne}\sum_{i=1}^n Y_i && \text{(since } T \text{ is binary)}\\
&\leq \frac{M}{e} && \text{(since } Y_i \leq M\text{).}
\end{aligned}$$

Similarly, we have $T_1(\boldsymbol{Z}) \leq \frac{M}{e}$. Therefore, we have shown that $T_1(\boldsymbol{Z})$ and $T_0(\boldsymbol{Z})$ are unbiased estimators of $\mathbb{E}\left[Y^{(1)}\right]$ and $\mathbb{E}\left[Y^{(0)}\right] > 0$, respectively. We also established that $M/e \geq T_0(\boldsymbol{Z}) \geq m/(1-e) > 0$ and $|T_1(\boldsymbol{Z})| \leq M/e$. Furthermore, we pointed out that $\operatorname{Var}(T_1(\boldsymbol{Z})) = O_p\left(\frac{1}{n}\right)$ and $\operatorname{Var}(T_0(\boldsymbol{Z})) = O_p\left(\frac{1}{n}\right)$. Applying Theorem 2, we obtain:

$$|\mathbb{E}\left[\hat{\tau}_{\text{RR, HT, n}}\right] - \tau_{\text{RR}}| \leq \frac{M^2(1-e)^2}{ne^2m^2}\left(\frac{M(1-e)}{me} + 1\right) \leq \frac{2M^3(1-e)^3}{nm^3e^3},$$

and

$$|\operatorname{Var}(\hat{\tau}_{\text{RR, HT, n}}) - V_{\text{RR, HT}}| \leq \frac{2M^2(1-e)^4}{nm^4e^2}\left(\frac{M^2(1-e)^2}{m^2e^2} + 1\right) \leq \frac{4M^4(1-e)^6}{nm^6e^4}.$$

**Optimal choice of** $e$ The optimal value of $e_{opt}$ is the one that minimizes the variance of the Ratio Horvitz-Thomson estimator. Therefore, we need to solve:

$$\inf_{e\in(0,1)} \tau_{\mathrm{RR}}^2 \left( \frac{\mathbb{E}\left[\left(Y^{(1)}\right)^2\right]}{e\mathbb{E}\left[Y^{(1)}\right]^2} + \frac{\mathbb{E}\left[\left(Y^{(0)}\right)^2\right]}{(1-e)\mathbb{E}\left[Y^{(0)}\right]^2} \right)$$

Noting that the variance we want to minimize is convex in $e$, we can derive the variance and set it to 0 to find $e_{opt}$. We have:

$$\frac{C_1}{e_{opt}^2} = \frac{C_0}{(1-e_{opt})^2}$$

where $C_1 = \frac{\mathbb{E}\left[\left(Y^{(1)}\right)^2\right]}{\mathbb{E}\left[Y^{(1)}\right]^2}$ and $C_0 = \frac{\mathbb{E}\left[\left(Y^{(0)}\right)^2\right]}{\mathbb{E}\left[Y^{(0)}\right]^2}$.

- If $\frac{\mathrm{Var}\left(Y^{(1)}\right)}{\mathbb{E}[Y^{(1)}]^2} = \frac{\mathrm{Var}\left(Y^{(0)}\right)}{\mathbb{E}[Y^{(0)}]^2}$:

$$e_{opt} = 0.5$$

- otherwise:

$$e_{opt} = \frac{\mathbb{E}\left[\left(Y^{(1)}\right)^2\right]\mathbb{E}\left[Y^{(0)}\right]^2 - \sqrt{\mathbb{E}\left[\left(Y^{(1)}\right)^2\right]\mathbb{E}\left[\left(Y^{(0)}\right)^2\right]}\mathbb{E}\left[Y^{(1)}\right]\mathbb{E}\left[Y^{(0)}\right]}{\mathbb{E}\left[\left(Y^{(1)}\right)^2\right]\mathbb{E}\left[Y^{(0)}\right]^2 - \mathbb{E}\left[\left(Y^{(0)}\right)^2\right]\mathbb{E}\left[Y^{(1)}\right]^2} \in (0,1)$$

□

### 6.2.3 Link with existing asymptotic confidence intervals

According to Proposition 1, a $(1-\alpha)$ asymptotic confidence interval for $\tau_{\mathrm{RR}}$ is given by

$$\left[\hat{\tau}_{\mathrm{RR,N,n}} \pm \frac{\sqrt{\widehat{V_{\mathrm{RR,N}}}}z_{1-\alpha/2}}{n}\right] \tag{34}$$

with $\widehat{V_{\mathrm{RR,N}}}$ an estimator of

$$V_{\mathrm{RR,N}} = \tau_{\mathrm{RR}}^2 \left( \frac{\mathrm{Var}\left(Y^{(1)}\right)}{e\mathbb{E}[Y^{(1)}]^2} + \frac{\mathrm{Var}\left(Y^{(0)}\right)}{(1-e)\mathbb{E}[Y^{(0)}]^2} \right).$$

Now, assume that $Y^{(0)}, Y^{(1)} \in \{0,1\}$ with associated probabilities $\mathbb{P}[Y^{(0)}=1] = p_0$ and $\mathbb{P}[Y^{(1)}=1] = p_1$. In this setting, the variance $V_{\mathrm{RR,N}}$ takes the form

$$\begin{aligned}
\frac{V_{\mathrm{RR,N}}}{n} &= \frac{\tau_{\mathrm{RR}}^2}{n} \left( \frac{\mathrm{Var}\left(Y^{(1)}\right)}{e\mathbb{E}[Y^{(1)}]^2} + \frac{\mathrm{Var}\left(Y^{(0)}\right)}{(1-e)\mathbb{E}[Y^{(0)}]^2} \right) \\
&= \tau_{\mathrm{RR}}^2 \left( \frac{p_1(1-p_1)}{N_1 p_1^2} + \frac{p_0(1-p_0)}{N_0 p_0^2} \right) \\
&= \tau_{\mathrm{RR}}^2 \left( \frac{1-p_1}{N_1 p_1} + \frac{1-p_0}{N_0 p_0} \right) \\
&= \tau_{\mathrm{RR}}^2 \left( \frac{1}{N_1 p_1} - \frac{1}{N_1} + \frac{1}{N_0 p_0} - \frac{1}{N_0} \right) \\
&= \tau_{\mathrm{RR}}^2 \left( \frac{1}{N_1 p_1} - \frac{1}{N_1} + \frac{1}{N_0 p_0} - \frac{1}{N_0} \right).
\end{aligned}$$

An estimation of such a quantity can be constructed by replacing $p_1$ (resp. $p_0$) by $(1/N_1)\sum_{i=1}^n T_iY_i$ (resp. $(1/N_0)\sum_{i=1}^n (1-T_i)Y_i$), which leads to

$$\frac{\widehat{V_{\mathrm{RR,N}}}}{n} = \hat{\tau}_{\mathrm{RR}}^2\left(\frac{1}{\sum_{i=1}^n T_iY_i} - \frac{1}{N_1} + \frac{1}{\sum_{i=1}^n (1-T_i)Y_i} - \frac{1}{N_0}\right). \tag{35}$$

Thus, a $(1-\alpha)$ asymptotic confidence interval for $\tau_{\mathrm{RR}}$ is given by

$$\left[\hat{\tau}_{\mathrm{RR,N,n}} \pm z_{1-\alpha/2}\hat{\tau}_{\mathrm{RR,N,n}}\sqrt{\left(\frac{1}{\sum_{i=1}^n T_iY_i} - \frac{1}{N_1} + \frac{1}{\sum_{i=1}^n (1-T_i)Y_i} - \frac{1}{N_0}\right)}\right] \tag{36}$$

$$= \left[\hat{\tau}_{\mathrm{RR,N,n}}\left(1 \pm z_{1-\alpha/2}\sqrt{\left(\frac{1}{\sum_{i=1}^n T_iY_i} - \frac{1}{N_1} + \frac{1}{\sum_{i=1}^n (1-T_i)Y_i} - \frac{1}{N_0}\right)}\right)\right]. \tag{37}$$

Finally, since $e^x$ is equivalent to $1+x$ near $x=0$, the above interval is equivalent to that given by (3), which concludes the proof.

#### 6.2.4 Delta method with log function

According to Proposition 1, we know that

$$\sqrt{n}\left(\hat{\tau}_{\mathrm{RR,N,n}} - \tau_{\mathrm{RR}}\right) \xrightarrow{d} \mathcal{N}\left(0, V_{\mathrm{RR,N}}\right), \tag{38}$$

where

$$V_{\mathrm{RR,N}} = \tau_{\mathrm{RR}}^2\left(\frac{\mathrm{Var}\left(Y^{(1)}\right)}{e\mathbb{E}[Y^{(1)}]^2} + \frac{\mathrm{Var}\left(Y^{(0)}\right)}{(1-e)\mathbb{E}[Y^{(0)}]^2}\right).$$

Using the Delta method, with the function $\theta \mapsto \log(\theta)$, we obtain

$$\sqrt{n}\left(\log(\hat{\tau}_{\mathrm{RR,N,n}}) - \log(\tau_{\mathrm{RR}})\right) \xrightarrow{d} \mathcal{N}\left(0, (1/\tau_{\mathrm{RR}})^2 V_{\mathrm{RR,N}}\right). \tag{39}$$

Thus, a $(1-\alpha)$ asymptotic confidence interval for $\log(\tau_{\mathrm{RR}})$ is given by

$$\left[\log(\hat{\tau}_{\mathrm{RR,N,n}}) \pm z_{1-\alpha/2}\sqrt{\frac{V_{\mathrm{RR,N}}}{n\tau_{\mathrm{RR}}^2}}\right]. \tag{40}$$

Letting $V_{\log\ \mathrm{RR,N}} = V_{\mathrm{RR,N}}/\tau_{\mathrm{RR}}^2$, a $(1-\alpha)$ asymptotic confidence interval for $\tau_{\mathrm{RR}}$ is

$$\left[\hat{\tau}_{\mathrm{RR,N,n}}\exp\left(\pm z_{1-\alpha/2}\sqrt{\frac{V_{\log\ \mathrm{RR,N}}}{n}}\right)\right]. \tag{41}$$

Now, note that, if $Y^{(0)}, Y^{(1)} \in \{0,1\}$ with $\mathbb{P}[Y^{(t)}=1] = p_t$, we have

$$V_{\log\ \mathrm{RR,N}} = \frac{\mathrm{Var}\left(Y^{(1)}\right)}{e\mathbb{E}[Y^{(1)}]^2} + \frac{\mathrm{Var}\left(Y^{(0)}\right)}{(1-e)\mathbb{E}[Y^{(0)}]^2} \tag{42}$$

$$= \frac{p_1(1-p_1)}{ep_1^2} + \frac{p_0(1-p_0)}{(1-e)p_0^2} \tag{43}$$

$$= \frac{1}{ep_1} - \frac{1}{e} + \frac{1}{ep_0} - \frac{1}{1-e}. \tag{44}$$

Hence,

$$\frac{V_{\log\ \mathrm{RR,N}}}{n} = \frac{1}{enp_1} - \frac{1}{en} + \frac{1}{enp_0} - \frac{1}{n(1-e)}, \tag{45}$$

which can be estimated replacing $ne$ (resp. $n(1-e)$) by $N_1 = \sum_{i=1}^n T_i$ (resp. $N_0 = n - N_1$ and $enp_1$ (resp. $enp_0$) by $\sum_{i=1}^n Y_iT_i$ (resp. $\sum_{i=1}^n Y_i(1-T_i)$). Replacing $V_{\log\ \mathrm{RR,N}}/n$ by such an estimate in the asymptotic confidence interval (41) leads to the well-known formula presented in Equation (3).

## 6.3 Proofs of Section 3

### 6.3.1 Risk Ratio Inverse Propensity Weighting

*Proof of Proposition 2.*

**Asymptotic bias and variance of the oracle Risk Ratio IPW estimator** Recall that the oracle Risk Ratio IPW is defined as

$$\tau^{\star}_{\mathrm{RR,IPW}} = \left(\sum_{i=1}^{n} \frac{T_i Y_i}{e(X_i)}\right) \Big/ \left(\sum_{i=1}^{n} \frac{(1-T_i)Y_i}{1-e(X_i)}\right),$$

where the propensity score $e$ is assumed to be known. Let us define $g_1(Z) = TY/e(X)$ and $g_0(Z) = (1-T)Y/(1-e(X))$ with $Z = (X, T, Y)$. Since

$$\frac{m}{1-\eta} \leq g_1(Z) \leq \frac{M}{\eta} \quad \text{and} \quad g_0(Z) \leq \frac{M}{\eta},$$

the function $g_0$ and $g_1$ are bounded from above and below and thus square integrable. Besides, $\mathbb{E}\left[g_0(Z_i)\right] = \mathbb{E}\left[Y^{(0)}\right]$ and $\mathbb{E}\left[g_1(Z_i)\right] = \mathbb{E}\left[Y^{(1)}\right]$. We can therefore apply Theorem 2 and conclude that

$$\sqrt{n}(\tau^{\star}_{\mathrm{RR,IPW}} - \tau_{\mathrm{RR}}) \to \mathcal{N}(0, V_{\mathrm{RR,IPW}}),$$

where

$$V_{\mathrm{RR,IPW}} = \tau_{\mathrm{RR}}^2 \operatorname{Var}\left(\frac{\frac{T_iY_i}{e(X_i)}}{\mathbb{E}\left[Y^{(1)}\right]} - \frac{\frac{(1-T_i)Y_i}{1-e(X_i)}}{\mathbb{E}\left[Y^{(1)}\right]}\right). \tag{46}$$

Moreover,

$$\begin{aligned}
\operatorname{Var}\left(\frac{TY}{e(X)}\right) &= \mathbb{E}\left[\left(\frac{TY}{e(X)}\right)^2\right] - \mathbb{E}\left[\frac{TY}{e(X)}\right]^2 \\
&= \mathbb{E}\left[\frac{TY^2}{e(X)^2}\right] - \mathbb{E}\left[Y^{(1)}\right]^2 \\
&= \mathbb{E}\left[\frac{1}{e(X)^2}\mathbb{E}\left[T(Y^{(1)})^2|X\right]\right] - \mathbb{E}\left[Y^{(1)}\right]^2 \\
&= \mathbb{E}\left[\frac{1}{e(X)}\mathbb{E}\left[(Y^{(1)})^2|X\right]\right] - \mathbb{E}\left[Y^{(1)}\right]^2 \\
&= \mathbb{E}\left[\frac{1}{e(X)}\mathbb{E}\left[(Y^{(1)})^2|X\right]\right] - \mathbb{E}\left[Y^{(1)}\right]^2 \\
&= \mathbb{E}\left[\frac{(Y^{(1)})^2}{e(X)}\right] - \mathbb{E}\left[Y^{(1)}\right]^2.
\end{aligned}$$

Similarly

$$\operatorname{Var}\left(\frac{(1-T)Y}{1-e(X)}\right) = \mathbb{E}\left[\frac{(Y^{(0)})^2}{1-e(X)}\right] - \mathbb{E}\left[Y^{(0)}\right]^2.$$

Additionally, the covariance satisfies

$$\begin{aligned}
\operatorname{Cov}\left(\frac{TY}{e(X)}, \frac{(1-T)Y}{1-e(X)}\right) &= \mathbb{E}\left[\left(\frac{TY}{e(X)} - \mathbb{E}\left[Y^{(1)}\right]\right)\left(\frac{(1-T)Y}{1-e(X)} - \mathbb{E}\left[Y^{(0)}\right]\right)\right] \\
&= \mathbb{E}\left[\frac{TY}{e(X)}\frac{(1-T)Y}{1-e(X)}\right] - \mathbb{E}\left[Y^{(1)}\right]\mathbb{E}\left[\frac{(1-T)Y}{1-e(X)}\right] \\
&\quad - \mathbb{E}\left[Y^{(0)}\right]\mathbb{E}\left[\frac{TY}{e(X)}\right] + \mathbb{E}\left[Y^{(1)}\right]\mathbb{E}\left[Y^{(0)}\right] \\
&= -\mathbb{E}\left[Y^{(1)}\right]\mathbb{E}\left[Y^{(0)}\right].
\end{aligned}$$

Therefore, we get that

$$V_{\text{RR,IPW}} = \tau_{\text{RR}}^2 \left( \frac{\mathbb{E}\left[\frac{(Y^{(1)})^2}{e(X)}\right]}{\mathbb{E}\left[Y^{(1)}\right]^2} + \frac{\mathbb{E}\left[\frac{(Y^{(0)})^2}{1-e(X)}\right]}{\mathbb{E}\left[Y^{(0)}\right]^2} \right).$$

As a consequence an estimator $\hat{V}_{RR-IPW}$ can be derived:

$$\hat{V}_{RR-IPW} = \hat{\tau}_{\text{RR,IPW,n}}^2 \left( \frac{\frac{1}{n}\sum_{T_i=1}\left(\frac{Y_i}{\hat{e}(x_i)}\right)^2}{\left(\frac{1}{n}\sum_{T_i=1} Y_i\right)^2} + \frac{\frac{1}{n}\sum_{T_i=0}\left(\frac{Y_i}{1-\hat{e}(x_i)}\right)^2}{\left(\frac{1}{n}\sum_{T_i=0} Y_i\right)^2} \right) \tag{47}$$

Since we have $\mathbb{E}\left[\left(\frac{TY}{e(X)}\right)^2\right] = \mathbb{E}\left[\frac{(Y^{(1)})^2}{e(X)}\right]$.

**Finite sample bias and variance of the oracle Risk Ratio IPW estimator** Let $T_1(\boldsymbol{Z}) = \frac{1}{n}\sum_{i=1}^n \frac{T_i Y_i}{e(X_i)}$ and $T_0(\boldsymbol{Z}) = \frac{1}{n}\sum_{i=1}^n \frac{(1-T_i)Y_i}{1-e(X_i)}$ where $\boldsymbol{Z} = (Z_1, \ldots, Z_n)$. We first show that $\text{Var}(T_1(\boldsymbol{Z})) = O_p\left(\frac{1}{n}\right)$ and $\text{Var}(T_0(\boldsymbol{Z})) = O_p\left(\frac{1}{n}\right)$:

$$\begin{aligned}
\text{Var}(T_1(\boldsymbol{Z})) &= \frac{1}{n^2}\,\text{Var}\left(\sum_{i=1}^n \frac{T_i Y_i}{e(X_i)}\right) \\
&= \frac{1}{n^2}\sum_{i=1}^n \text{Var}\left(\frac{T_i Y_i}{e(X_i)}\right) && \text{(by i.i.d.)} \\
&= \frac{1}{n}\left(\mathbb{E}\left[\left(\frac{T_i Y_i}{e(X_i)}\right)^2\right] - \mathbb{E}\left[\frac{T_i Y_i}{e(X_i)}\right]^2\right) && \text{(by law of total expectation)} \\
&= \frac{\mathbb{E}\left[\frac{(Y^{(1)})^2}{e(X_i)}\right] - \mathbb{E}\left[Y^{(1)}\right]^2}{n} \\
&= O_p\left(\frac{1}{n}\right)
\end{aligned}$$

Similarly, $\text{Var}(T_0(\boldsymbol{Z})) = O_p\left(\frac{1}{n}\right)$. And we also have:

$$\mathbb{E}\left[T_1(\boldsymbol{Z})\right] = \mathbb{E}\left[\frac{T_i Y_i}{e(X_i)}\right] = \mathbb{E}\left[Y^{(1)}\right]$$

$$\mathbb{E}\left[T_0(\boldsymbol{Z})\right] = \mathbb{E}\left[\frac{(1-T_i)Y_i}{1-e(X_i)}\right] = \mathbb{E}\left[Y^{(0)}\right]$$

Therefore, we showed that $T_1(\boldsymbol{Z})$ and $T_0(\boldsymbol{Z})$ are respectively unbiased estimators of $\mathbb{E}\left[Y^{(1)}\right]$ and $\mathbb{E}\left[Y^{(0)}\right] > 0$ such that $\text{Var}(T_1(\boldsymbol{Z})) = O_p\left(\frac{1}{n}\right)$ and $\text{Var}(T_0(\boldsymbol{Z})) = O_p\left(\frac{1}{n}\right)$. By assumption,

$$\frac{m}{1-\eta} \leq T_0(\boldsymbol{Z}) \leq \frac{M}{\eta} \quad \text{and} \quad T_1(\boldsymbol{Z}) \leq \frac{M}{\eta},$$

thus $T_0(\boldsymbol{Z})$ and $T_1(\boldsymbol{Z})$ are bounded. Applying Theorem 2, we obtain

$$\left|\mathbb{E}\left[\hat{\tau}_{\text{RR, HT, n}}\right] - \tau_{\text{RR}}\right| \leq \frac{2M^3(1-\eta)^3}{nm^3\eta^3},$$

and

$$|\mathrm{Var}(\hat{\tau}_{\mathrm{RR,\ HT,\ n}}) - V_{\mathrm{RR,\ HT}}| \leq \frac{4M^4(1-\eta)^6}{nm^6\eta^4}.$$

□

### 6.3.2 Risk Ratio G formula estimator

*Proof of Proposition 3.*

**Asymptotic bias and variance of the oracle risk ratio G formula estimator** Recall that the oracle risk ratio G formula is defined as

$$\tau^{\star}_{\mathrm{RR,G,n}} = \frac{\sum_{i=1}^{n} \mu_{(1)}(X_i)}{\sum_{i=1}^{n} \mu_{(0)}(X_i)},$$

where the response surfaces $\mu_{(0)}$ and $\mu_{(1)}$ are assumed to be known. Let us define $g_1(Z) = \mu_{(1)}(X_i)$ and $g_0(Z) = \mu_{(0)}(X_i)$ with $Z = X$. Since $g_1(Z)$ and $g_0(Z)$ are bounded, they are square integrable. We also have that $\mathbb{E}\left[g_0(Z_i)\right] = \mathbb{E}\left[Y^{(0)}\right]$ and $\mathbb{E}\left[g_1(Z_i)\right] = \mathbb{E}\left[Y^{(1)}\right]$. We can therefore apply Theorem 2 and conclude that

$$\sqrt{n}(\tau^{\star}_{\mathrm{RR,G,n}} - \tau_{\mathrm{RR}}) \to \mathcal{N}(0, V_{\mathrm{RR,G}}),$$

where $V_{\mathrm{RR,G}} = \tau^2_{\mathrm{RR}} \,\mathrm{Var}\left(\frac{\mu_1^{\star}(X)}{\mathbb{E}\left[Y^{(1)}\right]} - \frac{\mu_0^{\star}(X)}{\mathbb{E}\left[Y^{(0)}\right]}\right)$. As a consequence an estimator $\hat{V}_{RR,G}$ can be derived:

$$\hat{V}_{RR,G} = \frac{\hat{\tau}^2_{\mathrm{RR,G,n}}}{n} \sum_{i=1}^{n} \left( \frac{\hat{\mu}_1(X_i)}{\frac{1}{n}\sum_{T_i=1} Y_i} - \frac{\hat{\mu}_0(X_i)}{\frac{1}{n}\sum_{T_i=0} Y_i} - \frac{1}{n}\sum_{i=1}^{n} \frac{\hat{\mu}_1(X_i)}{\frac{1}{n}\sum_{T_i=1} Y_i} - \frac{\hat{\mu}_0(X_i)}{\frac{1}{n}\sum_{T_i=0} Y_i} \right)^2 \tag{48}$$

**Finite sample bias and variance of the oracle ratio G formula estimator** Let $T_1(\boldsymbol{Z}) = \frac{1}{n}\sum_{i=1}^{n} \mu_{(1)}(X_i)$ and $T_0(\boldsymbol{Z}) = \frac{1}{n}\sum_{i=1}^{n} \mu_{(0)}(X_i)$ where $\boldsymbol{Z} = (X_1, \ldots, X_n)$. We first show that $\mathrm{Var}(T_1(\boldsymbol{Z})) = O_p\left(\frac{1}{n}\right)$ and $\mathrm{Var}(T_0(\boldsymbol{Z})) = O_p\left(\frac{1}{n}\right)$:

$$\begin{aligned}
\mathrm{Var}(T_1(\boldsymbol{Z})) &= \frac{1}{n^2}\,\mathrm{Var}\left(\sum_{i=1}^{n} \mu_{(1)}(X_i)\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}(\mu_{(1)}(X_i)) && \text{(by i.i.d.)} \\
&= \frac{1}{n}\left(\mathbb{E}\left[(\mu_{(1)}(X_i))^2\right] - \mathbb{E}\left[Y^{(1)}\right]^2\right) && \text{(by law of total expectation)} \\
&\leq \frac{M^2 - \mathbb{E}\left[Y^{(1)}\right]^2}{n} && (\mu_{(1)}(X_i)) \leq M) \\
&= O_p\left(\frac{1}{n}\right).
\end{aligned}$$

Similarly, $\mathrm{Var}(T_0(\boldsymbol{Z})) = O_p\left(1/n\right)$. Since we also have that

$$\mathbb{E}\left[T_1(\boldsymbol{Z})\right] = \mathbb{E}\left[Y^{(1)}\right] \quad \mathbb{E}\left[T_0(\boldsymbol{Z})\right] = \mathbb{E}\left[Y^{(0)}\right]$$

Therefore, we showed that $T_1(\boldsymbol{Z})$ and $T_0(\boldsymbol{Z})$ are unbiased estimators of $\mathbb{E}\left[Y^{(1)}\right]$ and $\mathbb{E}\left[Y^{(0)}\right] > 0$ such that $\mathrm{Var}(T_1(\boldsymbol{Z})) = O_p\left(1/n\right)$ and $\mathrm{Var}(T_0(\boldsymbol{Z})) = O_p\left(1/n\right)$. We also have that $T_0(\boldsymbol{Z})$ and $T_1(\boldsymbol{Z})$ are bounded:

$$m_0 \leq T_0(\boldsymbol{Z}) \leq M_0 \quad \text{and} \quad T_1(\boldsymbol{Z}) \leq M_1$$

Applying Theorem 2, we obtain:

$$|\mathbb{E}\left[\hat{\tau}_{\mathrm{RR,\ HT,\ n}}\right] - \tau_{\mathrm{RR}}| \leq \frac{2M_1M_0^2}{nm_0^3} \quad \text{and} \quad |\mathrm{Var}(\hat{\tau}_{\mathrm{RR,\ HT,\ n}}) - V_{\mathrm{RR,\ HT}}| \leq \frac{2M_0^2M_1(M_1+M_0)}{m_0^6}$$

□

### 6.3.3 Risk Ratio G-formula in linear models

**Lemma 1** (see, e.g. Seber and Lee (2012)). *Grant Assumption 4 linear model. Let $\gamma_{(t)} = (c_{(t)}, \beta_{(t)}) \in \mathbb{R}^{d+1}$ and $Z = (1, X)$. We rearrange the $Y_i$ and $Z_i$ so that the first $n_1$ observations correspond to $T = 1$. We then define $\mathbf{Y}_1 = (Y_1, \ldots, Y_{n_1})^\top$ and $\mathbf{Y}_0 = (Y_{n_1+1}, \ldots, Y_n)^\top$, as well as $\mathbf{Z}_1 = (Z_1, \ldots, Z_{n_1})^\top$ and $\mathbf{Z}_0 = (Z_{n_1+1}, \ldots, Z_n)^\top$. Then for $t \in \{0, 1\}$, the linear model can be formulated as:*

$$Y^{(t)} = Z^\top \gamma_{(t)} + \varepsilon_{(t)}, \quad \mathbb{E}[\varepsilon_{(t)}|Z] = 0, \quad Var[\varepsilon_{(t)}|Z] = \sigma^2,$$

*and the least square estimator is given as*

$$\hat{\gamma}_{(t)} = \left(\frac{1}{n_t}\mathbf{Z}_t^\top \mathbf{Z}_t\right)^{-1} \frac{1}{n_t}\mathbf{Z}_t^\top \mathbf{Y}_t$$

**Proposition 8.** *Grant Assumption 4. Let $\hat{e} = (\sum_{i=1}^n T_i)/n$ and for all $t \in \{0, 1\}$,*

$$\bar{Z}_{(t)} = \frac{1}{\sum_{i=1}^n \mathbb{1}_{T_i=t}} \sum_{i=1}^n \mathbb{1}_{T_i=t} Z_i. \tag{49}$$

*Defining $\nu_t = \mathbb{E}[X|T = t]$ and $\Sigma_t = Var(X|T = t)$, we have*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_\infty) \xrightarrow{d} \mathcal{N}(0, \Sigma),$$

*where*

$$\boldsymbol{\theta}_n = \begin{pmatrix} \bar{Z}_{(0)} \\ \bar{Z}_{(1)} \\ \hat{\gamma}_{(0)} \\ \hat{\gamma}_{(1)} \\ \hat{e} \end{pmatrix}, \quad \boldsymbol{\theta}_\infty = \begin{pmatrix} E[Z|T=0] \\ E[Z|T=1] \\ \gamma_{(0)} \\ \gamma_{(1)} \\ e \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \frac{\mathrm{Var}[Z|T=0]}{(1-e)} & 0 & 0 & 0 & 0 \\ 0 & \frac{\mathrm{Var}[Z|T=1]}{e} & 0 & 0 & 0 \\ 0 & 0 & \frac{\sigma^2 Q_0^{-1}}{1-e} & 0 & 0 \\ 0 & 0 & 0 & \frac{\sigma^2 Q_1^{-1}}{e} & 0 \\ 0 & 0 & 0 & 0 & e(1-e) \end{pmatrix},$$

*with* $Q_t^{-1} = \begin{bmatrix} 1 + \nu_t^T \Sigma_t^{-1} \nu_t & -\nu_t^T \Sigma_t^{-1} \\ -\Sigma_t^{-1} \nu_t & \Sigma_t^{-1} \end{bmatrix}$.

*Proof.* Using M-estimation theory to prove asymptotic normality of the $\theta_n$, we first define the following:

$$\psi(T, Z, \boldsymbol{\theta}) = \begin{pmatrix} \psi_0(T, Z, \boldsymbol{\theta}) \\ \psi_1(T, Z, \boldsymbol{\theta}) \\ \psi_2(T, Z, \boldsymbol{\theta}) \\ \psi_3(T, Z, \boldsymbol{\theta}) \\ \psi_4(T, Z, \boldsymbol{\theta}) \end{pmatrix} := \begin{pmatrix} (1-T)(Z - \theta_0) \\ T(Z - \theta_1) \\ (1-T)\left(Z\epsilon(0) - ZZ^\top\left(\theta_2 - \gamma_{(0)}\right)\right) \\ T\left(Z\epsilon(1) - ZZ^\top\left(\theta_3 - \gamma(1)\right)\right) \\ T - \theta_4 \end{pmatrix}$$

where $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2, \theta_3, \theta_4)$. We still have that $\hat{\boldsymbol{\theta}}_n = (\bar{Z}_{(0)}, \bar{Z}_{(1)}, \hat{\gamma}_{(0)}, \hat{\gamma}_{(1)}, \hat{e})$ is an M-estimator of type $\psi$ (see Stefanski and Boos, 2002) since

$$\sum_{i=1}^n \psi(T_i, Z_i, \hat{\boldsymbol{\theta}}_n) = 0.$$

We now demonstrate that $\mathbb{E}\left[\psi(T, Y, \boldsymbol{\theta}_\infty)\right] = 0$. We directly have that $\mathbb{E}\left[\psi_4(T, Y, \boldsymbol{\theta}_\infty)\right] = 0$. For the other terms we have:

$$\begin{aligned}
\mathbb{E}\left[\psi_1(T, Z, \boldsymbol{\theta}_\infty)\right] &= \mathbb{E}\left[T\left(Z - \mathbb{E}[Z|T=1]\right)\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[T\left(Z - \mathbb{E}[Z|T=1]\right)|T\right]\right] \\
&= \mathbb{E}\left[T\left(\mathbb{E}\left[Z|T\right] - \mathbb{E}[Z|T=1]\right)\right] \\
&= \mathbb{E}\left[T\left(\mathbb{E}\left[Z|T\right] - \mathbb{E}[Z|T=1]\right)\right] \\
&= \mathbb{P}\left[T=1\right]\left(\mathbb{E}\left[Z|T=1\right] - \mathbb{E}[Z|T=1]\right) \\
&= 0
\end{aligned}$$

We also have that:

$$\begin{aligned}\mathbb{E}\left[\psi_3(T,Z,\boldsymbol{\theta}_\infty)\right] &= \mathbb{E}\left[TZ\epsilon_{(1)}\right]\\ &= \mathbb{E}\left[Z\mathbb{E}\left[T\epsilon_{(1)}|Z\right]\right]\\ &= \mathbb{E}\left[Z\mathbb{E}\left[\epsilon_{(1)}|Z,T=1\right]\right]\\ &= 0.\end{aligned}$$

Similarly, we can show:

$$\mathbb{E}\left[\psi_0(T,Z,\boldsymbol{\theta}_\infty)\right]=0 \quad \text{and} \quad \mathbb{E}\left[\psi_2(T,Z,\boldsymbol{\theta}_\infty)\right]=0.$$

At this point, we note that since $\psi(T,Z,\boldsymbol{\theta})$ is a linear function of $\boldsymbol{\theta}$, $\theta_\infty$ is the only value of $\boldsymbol{\theta}$ such that $\mathbb{E}\left[\psi(T,Z,\boldsymbol{\theta})\right]=0$ We proceed by defining:

$$A\left(\theta_\infty\right)=\mathbb{E}\left[\frac{\partial\psi}{\partial\theta}\Big|_{\theta=\theta_\infty}\right] \quad \text{and} \quad B(\theta_\infty)=\mathbb{E}\left[\psi(T,Z,\theta_\infty)\psi(T,Z,\theta_\infty)^T\right].$$

Next, we check the conditions of Theorem 7.2 in Stefanski and Boos (2002). First, we compute $A\left(\theta_\infty\right)$ and $B\left(\theta_\infty\right)$. Since:

$$\frac{\partial\psi}{\partial\theta}(T,Z,\theta)=\begin{pmatrix} -(1-T) & 0 & 0 & 0 & 0\\ 0 & -T & 0 & 0 & 0\\ 0 & 0 & -(1-T)ZZ^\top & 0 & 0\\ 0 & 0 & 0 & -TZZ^\top & 0\\ 0 & 0 & 0 & 0 & -1\end{pmatrix},$$

we obtain:

$$A\left(\boldsymbol{\theta_\infty}\right)=\begin{pmatrix} -(1-e) & 0 & 0 & 0 & 0\\ 0 & -e & 0 & 0 & 0\\ 0 & 0 & -(1-e)Q_0 & 0 & 0\\ 0 & 0 & 0 & -eQ_1 & 0\\ 0 & 0 & 0 & 0 & -1\end{pmatrix}, \quad \text{where} \quad Q_t=\mathbb{E}\left[ZZ^\top|T=t\right].$$

which leads to:

$$A^{-1}\left(\boldsymbol{\theta_\infty}\right)=\begin{pmatrix} -\frac{1}{1-e} & 0 & 0 & 0 & 0\\ 0 & -\frac{1}{e} & 0 & 0 & 0\\ 0 & 0 & -\frac{Q_0^{-1}}{1-e} & 0 & 0\\ 0 & 0 & 0 & -\frac{Q_1^{-1}}{e} & 0\\ 0 & 0 & 0 & 0 & -1\end{pmatrix}.$$

Regarding $B(\boldsymbol{\theta_\infty})$, since we have $T(1-T)=0$, elementary calculations show that:

$$\begin{matrix} B(\boldsymbol{\theta_\infty})_{1,2}=B(\boldsymbol{\theta_\infty})_{2,1}=0 \\ B(\boldsymbol{\theta_\infty})_{3,4}=B(\boldsymbol{\theta_\infty})_{4,3}=0\end{matrix} \quad \text{and} \quad \begin{matrix} B(\boldsymbol{\theta_\infty})_{1,4}=B(\boldsymbol{\theta_\infty})_{4,1}=0 \\ B(\boldsymbol{\theta_\infty})_{2,3}=B(\boldsymbol{\theta_\infty})_{3,2}=0.\end{matrix}$$

Besides

$$\begin{aligned}B(\boldsymbol{\theta_\infty})_{2,2} &= E\left[T^2(Z-E\left[Z|T=1\right])(Z-E\left[Z|T=1\right])^\top\right]\\ &= E\left[T(Z-E\left[Z|T=1\right])(Z-E\left[Z|T=1\right])^\top\right]\\ &= E\left[TE\left[(Z-E\left[Z|T=1\right])(Z-E\left[Z|T=1\right])^\top|T\right]\right]\\ &= \mathbb{P}\left[T=1\right]E\left[(Z-E\left[Z|T=1\right])(Z-E\left[Z|T=1\right])^\top|T=1\right]\\ &= e\,\mathrm{Var}\left[Z|T=1\right],\end{aligned}$$

and similarly,

$$B(\boldsymbol{\theta_\infty})_{1,1}=(1-e)\,\mathrm{Var}\left[Z|T=0\right].$$

We can also note that:

$$
\begin{aligned}
B(\boldsymbol{\theta_\infty})_{4,4} = E\left[T^2 Z Z^\top \epsilon_{(1)}^2\right] &= E\left[T Z Z^\top \epsilon_{(1)}^2\right] \\
&= E\left[T E\left[Z Z^\top \epsilon_{(1)}^2 | T\right]\right] \\
&= \mathbb{P}\left[T=1\right] E\left[Z Z^\top \epsilon_{(1)}^2 | T=1\right] \\
&= e E\left[Z Z^\top E\left[\epsilon_{(1)}^2 | T=1, Z\right] | T=1\right] \\
&= e\sigma^2 E\left[Z Z^\top | T=1\right] := e\sigma^2 Q_1,
\end{aligned}
$$

and similarly,

$$
B(\boldsymbol{\theta_\infty})_{3,3} = (1-e)\sigma^2 Q_0.
$$

Finally,

$$
\begin{aligned}
B(\boldsymbol{\theta_\infty})_{2,4} = B(\boldsymbol{\theta_\infty})_{4,2} &= E\left[T^2 (Z - E\left[Z|T=1\right]) Z^\top \epsilon_{(1)}\right] \\
&= E\left[T (Z - E\left[Z|T=1\right]) Z^\top \epsilon_{(1)}\right] \\
&= \mathbb{P}\left[T=1\right] E\left[(Z - E\left[Z|T=1\right]) Z^\top \epsilon_{(1)} | T=1\right] \\
&= e E\left[(Z - E\left[Z|T=1\right]) Z^\top E\left[\epsilon_{(1)} | T=1, Z\right] | T=1\right] \\
&= 0,
\end{aligned}
$$

and similarly,

$$
B(\boldsymbol{\theta_\infty})_{1,3} = B(\boldsymbol{\theta_\infty})_{3,1} = 0.
$$

We also have that:

$$
\begin{aligned}
B(\boldsymbol{\theta_\infty})_{2,5} = B(\boldsymbol{\theta_\infty})_{5,2} &= E\left[T(Z - E\left[Z|T=1\right])(T-e)\right] \\
&= E\left[T^2 Z - T^2 E\left[Z|T=1\right] - eTZ + eTE\left[Z|T=1\right]\right] \\
&= E\left[T^2 Z - T^2 E\left[Z|T=1\right] - eTZ + eTE\left[Z|T=1\right]\right] \\
&= eE\left[Z|T=1\right] - eE\left[Z|T=1\right] - e^2 E\left[Z|T=1\right] + e^2 E\left[Z|T=1\right] \\
&= 0
\end{aligned}
$$

and similarly,

$$
B(\boldsymbol{\theta_\infty})_{1,5} = B(\boldsymbol{\theta_\infty})_{5,1} = 0.
$$

We also have that :

$$
\begin{aligned}
B(\boldsymbol{\theta_\infty})_{4,5} = B(\boldsymbol{\theta_\infty})_{5,4} &= E\left[(T-e)TZ\epsilon_{(0)}\right] \\
&= E\left[(TZ\epsilon_{(0)}\right] - eE\left[(TZ\epsilon_{(0)}\right] \\
&= (1-e)E\left[TZ\epsilon_{(0)}\right] \\
&= (1-e)\mathbb{E}\left[Z\mathbb{E}\left[T\epsilon_{(0)}|Z\right]\right] \\
&= (1-e)\mathbb{E}\left[Z\mathbb{E}\left[\epsilon_{(0)}|Z, T=1\right]\right] \\
&= 0
\end{aligned}
$$

and similarly,

$$
B(\boldsymbol{\theta_\infty})_{3,5} = B(\boldsymbol{\theta_\infty})_{5,3} = 0.
$$

Gathering all calculations, and since $B(\boldsymbol{\theta_\infty})_{5,5} = e(1-e)$, we have

$$
B(\boldsymbol{\theta_\infty}) = \begin{pmatrix}
(1-e)\operatorname{Var}\left[Z|T=0\right] & 0 & 0 & 0 & 0 \\
0 & e\operatorname{Var}\left[Z|T=1\right] & 0 & 0 & 0 \\
0 & 0 & (1-e)\sigma^2 Q_0 & 0 & 0 \\
0 & 0 & 0 & e\sigma^2 Q_1 & 0 \\
0 & 0 & 0 & 0 & e(1-e)
\end{pmatrix},
$$

Based on the previous calculations, we have:

- $\psi(z, \boldsymbol{\theta})$ and its first two partial derivatives with respect to $\boldsymbol{\theta}$ exist for all $z$ and for all $\boldsymbol{\theta}$ in the neighborhood of $\boldsymbol{\theta_\infty}$.
- For each $\boldsymbol{\theta}$ in the neighborhood of $\boldsymbol{\theta_\infty}$, we have for all $i, j, k \in \{0, 2\}$:

$$\left| \frac{\partial^2}{\partial \theta_i \partial \theta_j} \psi_k(z, \boldsymbol{\theta}) \right| \leq 1$$

  and 1 is integrable.
- $A(\theta_\infty)$ exists and is nonsingular.
- $B(\theta_\infty)$ exists and is finite.

Since we have:

$$\sum_{i=1}^n \psi(T_i, Z_i, \hat{\boldsymbol{\theta}}_n) = 0 \quad \text{and} \quad \hat{\boldsymbol{\theta}}_n \xrightarrow{p} \theta_\infty.$$

Then, the conditions of Theorem 7.2 in Stefanski and Boos (2002) are satisfied, we have:

$$\sqrt{n} \left( \hat{\boldsymbol{\theta}}_n - \theta_\infty \right) \xrightarrow{d} \mathcal{N}\left(0, A(\theta_\infty)^{-1} B(\theta_\infty) (A(\theta_\infty)^{-1})^\top\right),$$

where:

$$A(\theta_\infty)^{-1} B(\theta_\infty) (A(\theta_\infty)^{-1})^\top = \begin{pmatrix} \frac{\text{Var}[Z|T=0]}{(1-e)} & 0 & 0 & 0 & 0 \\ 0 & \frac{\text{Var}[Z|T=1]}{e} & 0 & 0 & 0 \\ 0 & 0 & \frac{\sigma^2 Q_0^{-1}}{1-e} & 0 & 0 \\ 0 & 0 & 0 & \frac{\sigma^2 Q_1^{-1}}{e} & 0 \\ 0 & 0 & 0 & 0 & e(1-e) \end{pmatrix},$$

□

**Proposition 9** (asymptotical normality of $\hat{\tau}_{\text{RR,OLS}}$)**.** *Assume we have linear model then we have:*

$$\sqrt{n}(\hat{\tau}_{RR,OLS} - \tau_{RR}) \xrightarrow{d} \mathcal{N}\left(0, V_{RR\text{-}OLS}\right)$$

*with*

$$\frac{V_{RR\text{-}OLS}}{\tau_{RR}^2} = \left\| \frac{\beta_{(1)}}{\mathbb{E}\left[Y^{(1)}\right]} - \frac{\beta_{(0)}}{\mathbb{E}\left[Y^{(0)}\right]} \right\|_\Sigma^2 + \sigma^2 \left( \frac{1 + (1-e)^2 \|\nu_1 - \nu_0\|^2_{\Sigma_1^{-1}}}{e \mathbb{E}\left[Y^{(1)}\right]^2} + \frac{1 + e^2 \|\nu_1 - \nu_0\|^2_{\Sigma_0^{-1}}}{(1-e) \mathbb{E}\left[Y^{(0)}\right]^2} \right).$$

*Proof.* Let $\hat{\beta}_{(1)}$ and $\hat{c}_{(1)}$ be the parameters obtained via fitting an ordinary least square method on the treated individuals only, that is

$$(\hat{\beta}_{(1)}, \hat{c}_{(1)}) \in \arg \min_{c_{(1)}, \beta_{(1)}} \sum_{i=1}^n (Y_i^{(1)} - c_{(1)} - \beta_{(1)} X_i)^2 \mathbb{1}_{T_i=1}. \tag{50}$$

Similarly, let $\hat{\beta}_{(0)}$ and $\hat{c}_{(0)}$ be the parameters obtained via fitting an ordinary least square method on the control individuals only, that is

$$(\hat{\beta}_{(0)}, \hat{c}_{(0)}) \in \arg \min_{c_{(0)}, \beta_{(0)}} \sum_{i=1}^n (Y_i^{(0)} - c_{(0)} - \beta_{(0)} X_i)^2 \mathbb{1}_{T_i=0}. \tag{51}$$

An estimator of the RR using the G-formula approach is thus given by

$$\hat{\tau}_{\text{RR,OLS}} = \frac{\sum_{i=1}^n \left( \hat{c}_{(1)} + X_i^\top \hat{\beta}_{(1)} \right)}{\sum_{i=1}^n \left( \hat{c}_{(0)} + X_i^\top \hat{\beta}_{(0)} \right)} \tag{52}$$

$$= \frac{\hat{c}_{(1)} + \bar{X}^\top \hat{\beta}_{(1)}}{\hat{c}_{(0)} + \bar{X}^\top \hat{\beta}_{(0)}}. \tag{53}$$

Besides, note that assuming a linear model implies that

$$\hat{\tau}_{\mathrm{RR,OLS}} = \frac{c_{(1)} + \mathbb{E}[X]^\top \beta_{(1)}}{c_{(0)} + \mathbb{E}[X]^\top \beta_{(0)}}. \tag{54}$$

Let, for all $i$, $Z_i = (1, X_i)$ and $\gamma_{(j)} = (c_{(j)}, \beta_{(j)})$ for all $j \in \{0,1\}$. Expanding the following difference, we have:

$$\begin{aligned}
\sqrt{n}(\tau_{\mathrm{RR,OLS}} - \tau_{\mathrm{RR}}) &= \sqrt{n}\left(\frac{\hat{c}_{(1)} + \bar{X}^\top \hat{\beta}_{(1)}}{\hat{c}_{(0)} + \bar{X}^\top \hat{\beta}_{(0)}} - \frac{c_{(1)} + \mathbb{E}[X]^\top \beta_{(1)}}{c_{(0)} + \mathbb{E}[X]^\top \beta_{(0)}}\right) && (55)\\
&= \sqrt{n}\left(\hat{c}_{(1)} + \bar{X}^\top \hat{\beta}_{(1)}\right)\left(\frac{1}{\hat{c}_{(0)} + \bar{X}^\top \hat{\beta}_{(0)}} - \frac{1}{c_{(0)} + \mathbb{E}[X]^\top \beta_{(0)}}\right) && (56)\\
&\quad + \frac{\sqrt{n}}{c_{(0)} + \mathbb{E}[X]^\top \beta_{(0)}}\left(\hat{c}_{(1)} + \bar{X}^\top \hat{\beta}_{(1)} - c_{(1)} - \mathbb{E}[X]^\top \beta_{(1)}\right) && (57)\\
&= \sqrt{n}\left(\bar{Z}^\top \hat{\gamma}_{(1)}\right)\left(\frac{1}{\bar{Z}^\top \hat{\gamma}_{(0)}} - \frac{1}{\mathbb{E}[Z]^\top \gamma_{(0)}}\right) && (58)\\
&\quad + \frac{\sqrt{n}}{\mathbb{E}[Z]^\top \gamma_{(0)}}\left(\bar{Z}^\top \hat{\gamma}_{(1)} - \mathbb{E}[Z]^\top \gamma_{(1)}\right) && (59)\\
&= \sqrt{n}\frac{\bar{Z}^\top \hat{\gamma}_{(1)}}{\bar{Z}^\top \hat{\gamma}_{(0)}\mathbb{E}[Z]^\top \gamma_{(0)}}\left(\mathbb{E}[Z]^\top \gamma_{(0)} - \bar{Z}^\top \hat{\gamma}_{(0)}\right) && (60)\\
&\quad + \frac{\sqrt{n}}{\mathbb{E}[Z]^\top \gamma_{(0)}}\left(\bar{Z}^\top \hat{\gamma}_{(1)} - \mathbb{E}[Z]^\top \gamma_{(1)}\right). && (61)
\end{aligned}$$

Besides, we have

$$\begin{aligned}
\bar{Z} - \mathbb{E}[Z] &= \hat{e}\bar{Z}_{(1)} + (1-\hat{e})\bar{Z}_{(0)} - e\mathbb{E}[Z|T=1] - (1-e)\mathbb{E}[Z|T=0]\\
&= (1-e)\left(\bar{Z}_{(0)} - \mathbb{E}[Z|T=0]\right) + e\left(\bar{Z}_{(1)} - \mathbb{E}[Z|T=1]\right) + \left(\bar{Z}_{(1)} - \bar{Z}_{(0)}\right)(\hat{e} - e)\\
&= \zeta(\theta_n - \theta_\infty),
\end{aligned}$$

where $\zeta = \left[(1-e)I_{d+1},\ eI_{d+1},\ 0_{d+1},\ 0_{d+1},\ (\bar{Z}_{(1)} - \bar{Z}_{(0)})\right] \in \mathbb{R}^{(d+1)\times 4(d+1)+1}$ and

$$\boldsymbol{\theta}_n = \begin{pmatrix} \bar{Z}_{(0)} \\ \bar{Z}_{(1)} \\ \hat{\gamma}_{(0)} \\ \hat{\gamma}_{(1)} \\ \hat{e} \end{pmatrix}, \quad \boldsymbol{\theta}_\infty = \begin{pmatrix} E[Z|T=0] \\ E[Z|T=1] \\ \gamma_{(0)} \\ \gamma_{(1)} \\ e \end{pmatrix}.$$

Note that for all $t \in \{0,1\}$,

$$\begin{aligned}
\bar{Z}^\top \hat{\gamma}_{(t)} - \mathbb{E}[Z]^\top \gamma_{(t)} &= \hat{\gamma}_{(t)}^\top\left(\bar{Z} - \mathbb{E}[Z]\right) + \mathbb{E}[Z]^\top\left(\hat{\gamma}_{(t)} - \gamma_{(t)}\right)\\
&= \hat{\gamma}_{(t)}^\top \zeta(\theta_n - \theta_\infty) + \mathbb{E}[Z]^\top\left(\hat{\gamma}_{(t)} - \gamma_{(t)}\right)\\
&= \hat{\alpha}_{(t)}^\top(\theta_n - \theta_\infty),
\end{aligned}$$

with

$$\hat{\alpha}_{(t)} = \begin{pmatrix} (1-e)\hat{\gamma}_{(t)} \\ e\hat{\gamma}_{(t)} \\ \mathbb{1}_{t=0}\mathbb{E}[Z] \\ \mathbb{1}_{t=1}\mathbb{E}[Z] \\ (\bar{Z}_{(1)} - \bar{Z}_{(0)})^\top \hat{\gamma}_{(t)}. \end{pmatrix}$$

. Therefore

$$\begin{aligned}\sqrt{n}(\tau_{\mathrm{RR,OLS}}-\tau_{\mathrm{RR}}) &= \sqrt{n}\frac{\bar{Z}^\top\hat{\gamma}_{(1)}}{\bar{Z}^\top\hat{\gamma}_{(0)}\mathbb{E}[Z]^\top\gamma_{(0)}}\left(\mathbb{E}[Z]^\top\gamma_{(0)}-\bar{Z}^\top\hat{\gamma}_{(0)}\right)\\ &+\frac{\sqrt{n}}{\mathbb{E}[Z]^\top\gamma_{(0)}}\left(\bar{Z}^\top\hat{\gamma}_{(1)}-\mathbb{E}[Z]^\top\gamma_{(1)}\right)\\ &=\frac{\sqrt{n}}{\mathbb{E}[Z]^\top\gamma_{(0)}}\hat{\alpha}_{(1)}^\top(\theta_n-\theta_\infty)\\ &-\sqrt{n}\frac{\bar{Z}^\top\hat{\gamma}_{(1)}}{\bar{Z}^\top\hat{\gamma}_{(0)}\mathbb{E}[Z]^\top\gamma_{(0)}}\hat{\alpha}_{(0)}^\top(\theta_n-\theta_\infty)\end{aligned}$$

Therefore, we get that

$$\sqrt{n}(\tau_{\mathrm{RR,OLS}}-\tau_{\mathrm{RR}})=\sqrt{n}\left(\frac{1}{\mathbb{E}[Z]^\top\gamma_{(0)}}\hat{\alpha}_{(1)}-\frac{\bar{Z}^\top\hat{\gamma}_{(1)}}{\bar{Z}^\top\hat{\gamma}_{(0)}\mathbb{E}[Z]^\top\gamma_{(0)}}\hat{\alpha}_{(0)}\right)^\top(\theta_n-\theta_\infty).$$

According to the Law of Large Numbers,

$$\frac{1}{\mathbb{E}[Z]^\top\gamma_{(0)}}\hat{\alpha}_{(1)}-\frac{\bar{Z}^\top\hat{\gamma}_{(1)}}{\bar{Z}^\top\hat{\gamma}_{(0)}\mathbb{E}[Z]^\top\gamma_{(0)}}\hat{\alpha}_{(0)}\xrightarrow{p}\frac{\mathbb{E}[Z]^\top\gamma_{(1)}}{\mathbb{E}[Z]^\top\gamma_{(0)}}\left(\frac{\alpha_{(1)}}{\mathbb{E}[Z]^\top\gamma_{(1)}}-\frac{\alpha_{(0)}}{\mathbb{E}[Z]^\top\gamma_{(0)}}\right):=\alpha_\infty,$$

with, for all $t\in\{0,1\}$,

$$\alpha_{(t)}=\begin{pmatrix}(1-e)\gamma_{(t)}\\ e\gamma_{(t)}\\ \mathbb{1}_{t=0}\mathbb{E}[Z]\\ \mathbb{1}_{t=1}\mathbb{E}[Z]\\ (\mathbb{E}[Z|T=1]-\mathbb{E}[Z|T=0])^\top\gamma_{(t)}\end{pmatrix}$$

and

$$\alpha_\infty=\frac{\mathbb{E}[Z]^\top\gamma_{(1)}}{\mathbb{E}[Z]^\top\gamma_{(0)}}\begin{pmatrix}\frac{(1-e)\gamma_{(1)}}{\mathbb{E}[Z]^\top\gamma_{(1)}}-\frac{(1-e)\gamma_{(0)}}{\mathbb{E}[Z]^\top\gamma_{(0)}}\\ \frac{e\gamma_{(1)}}{\mathbb{E}[Z]^\top\gamma_{(1)}}-\frac{e\gamma_{(0)}}{\mathbb{E}[Z]^\top\gamma_{(0)}}\\ -\frac{\mathbb{E}[Z]}{\mathbb{E}[Z]^\top\gamma_{(0)}}\\ \frac{\mathbb{E}[Z]}{\mathbb{E}[Z]^\top\gamma_{(1)}}\\ \frac{\gamma_{(0)}^\top(\mathbb{E}[Z|T=1]-\mathbb{E}[Z|T=0])}{\mathbb{E}[Z]^\top\gamma_{(0)}}-\frac{\gamma_{(1)}^\top(\mathbb{E}[Z|T=1]-\mathbb{E}[Z|T=0])}{\mathbb{E}[Z]^\top\gamma_{(1)}}\end{pmatrix}. \tag{62}$$

According to Proposition 8, letting $Q_t=\mathbb{E}\left[ZZ^\top|T=t\right]$ for all $t\in\{0,1\}$, we have

$$\sqrt{n}(\theta_n-\theta_\infty)\xrightarrow{d}\mathcal{N}(0,\Sigma)\quad\text{where}\quad\Sigma=\begin{pmatrix}\frac{\mathrm{Var}[Z|T=0]}{(1-e)}&0&0&0&0\\ 0&\frac{\mathrm{Var}[Z|T=1]}{e}&0&0&0\\ 0&0&\frac{\sigma^2Q_0^{-1}}{1-e}&0&0\\ 0&0&0&\frac{\sigma^2Q_1^{-1}}{e}&0\\ 0&0&0&0&e(1-e)\end{pmatrix}.$$

By Slutsky's theorem,

$$\sqrt{n}\left(\frac{1}{\mathbb{E}[Z]^\top\gamma_{(0)}}\hat{\alpha}_{(1)}-\frac{\bar{Z}^\top\hat{\gamma}_{(1)}}{\bar{Z}^\top\hat{\gamma}_{(0)}\mathbb{E}[Z]^\top\gamma_{(0)}}\hat{\alpha}_{(0)}\right)^\top(\theta_n-\theta_\infty)\xrightarrow{d}\mathcal{N}\left(0,\alpha_\infty^\top\Sigma\alpha_\infty\right). \tag{63}$$

We now compute the covariance matrix

$$\frac{\alpha_\infty^\top\Sigma\alpha_\infty}{\left(\frac{\mathbb{E}[Z]^\top\gamma_{(1)}}{\mathbb{E}[Z]^\top\gamma_{(0)}}\right)^2}=(1-e)\left\|\frac{\gamma_{(1)}}{\mathbb{E}[Z]^\top\gamma_{(1)}}-\frac{\gamma_{(0)}}{\mathbb{E}[Z]^\top\gamma_{(0)}}\right\|^2_{\mathrm{Var}[Z|T=0]}+e\left\|\frac{\gamma_{(1)}}{\mathbb{E}[Z]^\top\gamma_{(1)}}-\frac{\gamma_{(0)}}{\mathbb{E}[Z]^\top\gamma_{(0)}}\right\|^2_{\mathrm{Var}[Z|T=1]} \tag{64}$$

$$+\frac{\sigma^2}{1-e}\left\|\frac{\mathbb{E}[Z]}{\mathbb{E}[Z]^\top\gamma_{(0)}}\right\|^2_{Q_0^{-1}}+\frac{\sigma^2}{e}\left\|\frac{\mathbb{E}[Z]}{\mathbb{E}[Z]^\top\gamma_{(1)}}\right\|^2_{Q_1^{-1}}+e(1-e)\left\|\frac{\gamma_{(1)}}{\mathbb{E}[Z]^\top\gamma_{(1)}}-\frac{\gamma_{(0)}}{\mathbb{E}[Z]^\top\gamma_{(0)}}\right\|^2_{\Delta\Delta^\top}, \tag{65}$$

where $\Delta = \mathbb{E}[Z \mid T = 1] - \mathbb{E}[Z \mid T = 0]$. This variance can be rewritten as follows. Summing the first two terms and the last term in (65) leads to

$$\left\| \frac{\gamma_{(1)}}{\mathbb{E}[Z]^\top \gamma_{(1)}} - \frac{\gamma_{(0)}}{\mathbb{E}[Z]^\top \gamma_{(0)}} \right\|_J^2 ,$$

where $J = (1-e)\operatorname{Var}(Z \mid T = 0) + e \operatorname{Var}(Z \mid T = 1) + e(1-e)\Delta\Delta^\top$. Let us prove that $J = \operatorname{Var}(Z)$. Letting $Z_i$ the components of $Z$ for all $1 \le i \le d+1$, by the law of total covariance, we have

$$\operatorname{Cov}[Z_i, Z_j] = \mathbb{E}[\operatorname{Cov}[Z_i, Z_j|T]] + \operatorname{Cov}[\mathbb{E}[Z_i|T], \mathbb{E}[Z_j|T]], \tag{66}$$

with, since $T \in \{0, 1\}$,

$$\mathbb{E}[\operatorname{Cov}[Z_i, Z_j|T]] = e \operatorname{Cov}[Z_i, Z_j|T = 1] + (1-e)\operatorname{Cov}[Z_i, Z_j|T = 0]. \tag{67}$$

Besides, since $\mathbb{E}[Z] = (1-e)\mathbb{E}[Z \mid T = 0] + e\mathbb{E}[Z \mid T = 1]$, we can compute the deviations from the unconditional mean:

$$\begin{aligned}
\mathbb{E}[Z \mid T = 0] - \mathbb{E}[Z] &= \mathbb{E}[Z \mid T = 0] - ((1-e)\mathbb{E}[Z \mid T = 0] + e\mathbb{E}[Z \mid T = 1]) \\
&= (1 - (1-e))\mathbb{E}[Z \mid T = 0] - e\mathbb{E}[Z \mid T = 1] \\
&= -e\left(\mathbb{E}[Z \mid T = 1] - \mathbb{E}[Z \mid T = 0]\right) \\
&= -e\Delta.
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}[Z \mid T = 1] - \mathbb{E}[Z] &= \mathbb{E}[Z \mid T = 1] - ((1-e)\mathbb{E}[Z \mid T = 0] + e\mathbb{E}[Z \mid T = 1]) \\
&= (1-e)\left(\mathbb{E}[Z \mid T = 1] - \mathbb{E}[Z \mid T = 0]\right) \\
&= (1-e)\Delta.
\end{aligned}$$

Now, we can compute the second term in (66)

$$\begin{aligned}
\operatorname{Cov}[\mathbb{E}[Z_i|T], \mathbb{E}[Z_j|T]] &= \mathbb{E}[(\mathbb{E}[Z_i|T] - \mathbb{E}[Z_i])(\mathbb{E}[Z_j|T] - \mathbb{E}[Z_j])] && (68)\\
&= e(\mathbb{E}[Z_i|T = 1] - \mathbb{E}[Z_i])(\mathbb{E}[Z_j|T = 1] - \mathbb{E}[Z_j]) && (69)\\
&\quad + (1-e)(\mathbb{E}[Z_i|T = 0] - \mathbb{E}[Z_i])(\mathbb{E}[Z_j|T = 0] - \mathbb{E}[Z_j]) && (70)\\
&= e(1-e)^2\Delta_i\Delta_j + e^2(1-e)\Delta_i\Delta_j && (71)\\
&= e(1-e)\Delta_i\Delta_j. && (72)
\end{aligned}$$

Consequently, according to (66),

$$\operatorname{Cov}[Z_i, Z_j] = e \operatorname{Cov}[Z_i, Z_j|T = 1] + (1-e)\operatorname{Cov}[Z_i, Z_j|T = 0] + e(1-e)\Delta_i\Delta_j, \tag{73}$$

which leads to

$$\operatorname{Var}[Z] = (1-e)\operatorname{Var}(Z \mid T = 0) + e \operatorname{Var}(Z \mid T = 1) + e(1-e)\Delta\Delta^\top. \tag{74}$$

Similarly the last two remaining terms we have for $t \in \{0, 1\}$:

$$\begin{aligned}
\left\| \frac{\mathbb{E}[Z]}{\mathbb{E}[Z]^\top \gamma_{(t)}} \right\|_{Q_t^{-1}}^2 &= \frac{1}{(\mathbb{E}[Z]^\top \gamma_{(t)})^2} \left\|\mathbb{E}[Z]\right\|_{Q_t^{-1}}^2 \\
&= \frac{1}{(\mathbb{E}[Z]^\top \gamma_{(t)})^2} \mathbb{E}[Z]^\top Q_t^{-1} \mathbb{E}[Z]
\end{aligned}$$

Note that we have $\mathbb{E}[Z] = e\mathbb{E}[Z|T = 1] + (1-e)\mathbb{E}[Z|T = 0]$ and that for $t \in \{0, 1\}$,

$$\mathbb{E}[Z|T = t]^\top Q_t^{-1} \mathbb{E}[Z|T = t] = \begin{pmatrix} 1 \\ \nu_t \end{pmatrix}^\top \begin{pmatrix} 1 + \nu_t^\top \Sigma_t^{-1} \nu_t & -\nu_t^\top \Sigma_t^{-1} \\ -\Sigma_t^{-1} \nu_t & \Sigma_t^{-1} \end{pmatrix} \begin{pmatrix} 1 \\ \nu_t \end{pmatrix} = 1,$$

and

$$\begin{aligned}\mathbb{E}[Z|T=1-t]^\top Q_t^{-1}\mathbb{E}[Z|T=1-t] &= \begin{pmatrix}1\\ \nu_{1-t}\end{pmatrix}^\top \begin{pmatrix}1+\nu_t^\top\Sigma_t^{-1}\nu_t & -\nu_t^\top\Sigma_t^{-1}\\ -\Sigma_t^{-1}\nu_t & \Sigma_t^{-1}\end{pmatrix}\begin{pmatrix}1\\ \nu_{1-t}\end{pmatrix}\\ &= 1+\|\nu_{1-t}-\nu_t\|^2_{\Sigma_t^{-1}},\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}[Z|T=t]^\top Q_t^{-1}\mathbb{E}[Z|T=1-t] &= \begin{pmatrix}1\\ \nu_t\end{pmatrix}^\top \begin{pmatrix}1+\nu_t^\top\Sigma_t^{-1}\nu_t & -\mu_t^\top\Sigma_t^{-1}\\ -\Sigma_t^{-1}\nu_t & \Sigma_t^{-1}\end{pmatrix}\begin{pmatrix}1\\ \nu_{1-t}\end{pmatrix}\\ &= 1.\end{aligned}$$

Therefore, we have

$$\begin{aligned}\mathbb{E}[Z]^\top Q_0^{-1}\mathbb{E}[Z] &= e^2\|\mathbb{E}[Z|T=1]\|_{Q_0^{-1}} + (1-e)^2\|\mathbb{E}[Z|T=0]\|_{Q_0^{-1}} + 2e(1-e)\langle\mathbb{E}[Z|T=0],\mathbb{E}[Z|T=1]\rangle_{Q_0^{-1}}\\ &= e^2\|\mathbb{E}[Z|T=1]\|_{Q_0^{-1}} + (1-e)^2 + 2e(1-e)\langle\mathbb{E}[Z|T=0],\mathbb{E}[Z|T=1]\rangle_{Q_0^{-1}}\\ &= (1-e)^2 + e^2\left(1+\|\mu_1-\mu_0\|^2_{\Sigma_0^{-1}}\right) + 2e(1-e)\\ &= 1+e^2\|\nu_1-\nu_0\|^2_{\Sigma_0^{-1}},\end{aligned}$$

and similarly $\mathbb{E}[Z]^\top Q_1^{-1}\mathbb{E}[Z] = 1+(1-e)^2\|\nu_1-\nu_0\|^2_{\Sigma_1^{-1}}$. Finally, noting that for all $t\in\{0,1\}$

$$\mathbb{E}[Z]^\top\gamma_{(t)} = \mathbb{E}\left[Y^{(t)}\right] \quad \text{and} \quad \mathrm{Var}\left[Z\right] = \begin{pmatrix}0 & \cdots & 0\\ \vdots & \mathrm{Var}\left[X\right] & \\ 0 & & \end{pmatrix},$$

we have, letting $\Sigma = \mathrm{Var}\left[X\right]$

$$V_{\mathrm{RR,G,OLS}} = \tau_{\mathrm{RR}}^2\left(\left\|\frac{\beta_{(1)}}{\mathbb{E}\left[Y^{(1)}\right]} - \frac{\beta_{(0)}}{\mathbb{E}\left[Y^{(0)}\right]}\right\|^2_\Sigma + \sigma^2\left(\frac{1+(1-e)^2\|\nu_1-\nu_0\|^2_{\Sigma_1^{-1}}}{e\mathbb{E}\left[Y^{(1)}\right]^2} + \frac{1+e^2\|\nu_1-\nu_0\|^2_{\Sigma_0^{-1}}}{(1-e)\mathbb{E}\left[Y^{(0)}\right]^2}\right)\right).$$

□

**Lemma 2 (Comparison of the asymptotic variances of $\hat{\tau}_{\mathbf{RR,N}}$ and $\hat{\tau}_{\mathbf{RR,G}}$ under a linear model).** *Grant Assumption 1, Assumption 2 and Assumption 4. Recalling that $V_{RR,G,OLS}$ (resp. $V_{RR,G,OLS}$) is the asymptotic variance of the G-formula when oracle surface responses are used (resp. when they are estimated via OLS), we have*

$$V_{RR,N} = \tau_{RR}^2\left(\frac{\left\|\beta_{(1)}\right\|^2_\Sigma+\sigma^2}{e\mathbb{E}\left[Y^{(1)}\right]^2} + \frac{\left\|\beta_{(0)}\right\|^2_\Sigma+\sigma^2}{(1-e)\mathbb{E}\left[Y^{(0)}\right]^2}\right), \tag{75}$$

$$V_{RR,G,OLS} = \tau_{RR}^2\left(\left\|\frac{\beta_{(1)}}{\mathbb{E}\left[Y^{(1)}\right]} - \frac{\beta_{(0)}}{\mathbb{E}\left[Y^{(0)}\right]}\right\|^2_\Sigma + \sigma^2\left(\frac{1}{e\mathbb{E}\left[Y^{(1)}\right]^2} + \frac{1}{(1-e)\mathbb{E}\left[Y^{(0)}\right]^2}\right)\right) \tag{76}$$

$$= V_{RR,G} + \tau_{RR}^2\sigma^2\left(\frac{1}{e\mathbb{E}\left[Y^{(1)}\right]^2} + \frac{1}{(1-e)\mathbb{E}\left[Y^{(0)}\right]^2}\right), \tag{77}$$

*and*

$$V_{RR,N} - V_{RR,G,OLS} = \tau_{RR}^2\left(e(1-e)\left\|\frac{\beta_{(1)}}{e\mathbb{E}\left[Y^{(1)}\right]} - \frac{\beta_{(0)}}{(1-e)\mathbb{E}\left[Y^{(0)}\right]}\right\|^2_\Sigma\right) \geq 0. \tag{78}$$

*Proof of Lemma 2.*

**First equality** The variance of $Y^{(a)}$ satisfies

$$\begin{aligned}
\operatorname{Var}[Y^{(a)}] &= \operatorname{Var}[c_{(t)} + X^\top \beta_{(t)} + \varepsilon_{(t)}] \\
&= \operatorname{Var}[X^\top \beta_{(t)} + \varepsilon_{(t)}] && c_{(t)} \text{ is a constant} \\
&= \operatorname{Var}[X^\top \beta_{(t)}] + \operatorname{Var}[\varepsilon_{(t)}] + 2\operatorname{Cov}(X^\top \beta_{(t)}, \varepsilon_{(t)}) && \text{Bienaymé's identity} \\
&= ||\beta_{(t)}||_\Sigma + \sigma^2, && \text{(by linear model)}
\end{aligned}$$

since

$$\begin{aligned}
\operatorname{Cov}(X^\top \beta_{(t)}, \varepsilon_{(t)}) &= \mathbb{E}[X^\top \beta_{(t)} \varepsilon_{(t)}] - \mathbb{E}[X^\top \beta_{(t)}]\mathbb{E}[\varepsilon_{(t)}] \\
&= \mathbb{E}[X^\top \beta_{(t)} \mathbb{E}[\varepsilon_{(t)}|X]] - \mathbb{E}[X^\top \beta_{(t)}]\mathbb{E}[\mathbb{E}[\varepsilon_{(t)}|X]] && \text{(by total expectation)} \\
&= 0, && \mathbb{E}[\varepsilon_{(t)}|X] = 0,
\end{aligned}$$

and, using Eve's law, $\operatorname{Var}[\varepsilon_{(t)}] = \mathbb{E}[\operatorname{Var}[\varepsilon_{(t)}|X]] + \operatorname{Var}[\mathbb{E}[\varepsilon_{(t)}|X]] = \sigma^2$. Thus, $V_{\mathrm{RR,N}}$ satisfies

$$V_{\mathrm{RR,N}} = \tau_{\mathrm{RR}}^2 \left( \frac{\operatorname{Var}(Y^{(1)})}{e\mathbb{E}[Y^{(1)}]^2} + \frac{\operatorname{Var}(Y^{(0)})}{(1-e)\mathbb{E}[Y^{(0)}]^2} \right) \tag{79}$$

$$= \tau_{\mathrm{RR}}^2 \left( \frac{||\beta_{(1)}||_\Sigma^2 + \sigma^2}{e\mathbb{E}[Y^{(1)}]^2} + \frac{||\beta_{(0)}||_\Sigma^2 + \sigma^2}{(1-e)\mathbb{E}[Y^{(0)}]^2} \right). \tag{80}$$

**Second and third equality** According to Proposition 4 ,

$$\frac{V_{\mathrm{RR,G,OLS}}}{\tau_{\mathrm{RR}}^2} = \left\| \frac{\beta_{(1)}}{\mathbb{E}\left[Y^{(1)}\right]} - \frac{\beta_{(0)}}{\mathbb{E}\left[Y^{(0)}\right]} \right\|_\Sigma^2 + \sigma^2 \left( \frac{1 + (1-e)^2 \|\nu_1 - \nu_0\|^2_{\Sigma_1^{-1}}}{e\mathbb{E}\left[Y^{(1)}\right]^2} + \frac{1 + e^2 \|\nu_1 - \nu_0\|^2_{\Sigma_0^{-1}}}{(1-e)\mathbb{E}\left[Y^{(0)}\right]^2} \right)$$

Since we are in a RCT setting, we have that $\nu_1 = \nu_0$ and $\Sigma_1 = \Sigma_0 = \Sigma$. Therefore

$$\frac{V_{\mathrm{RR,G,OLS}}}{\tau_{\mathrm{RR}}^2} = \left\| \frac{\beta_{(1)}}{\mathbb{E}\left[Y^{(1)}\right]} - \frac{\beta_{(0)}}{\mathbb{E}\left[Y^{(0)}\right]} \right\|_\Sigma^2 + \sigma^2 \left( \frac{1}{e\mathbb{E}\left[Y^{(1)}\right]^2} + \frac{1}{(1-e)\mathbb{E}\left[Y^{(0)}\right]^2} \right)$$

The first term corresponds to the Oracle variance of the G-formula. Indeed, for all $t \in \{0, 1\}$,

$$\begin{aligned}
\operatorname{Var}[\mu_{(t)}(X)] &= \operatorname{Var}[\mathbb{E}[Y^{(t)}|X]] \\
&= \operatorname{Var}[\mathbb{E}[c_{(t)}|X] + \mathbb{E}[X^\top \beta_{(t)}|X] + \mathbb{E}[\varepsilon_{(t)}|X]] \\
&= \operatorname{Var}[c_{(t)} + \mathbb{E}[X^\top \beta_{(t)}|X]] \\
&= \operatorname{Var}[\mathbb{E}[X^\top \beta_{(t)}|X]] \\
&= \operatorname{Var}[X^\top \beta_{(t)}] \\
&= \|\beta_{(t)}\|_\Sigma^2.
\end{aligned}$$

Besides, the covariance between $\mu_1(X)$ and $\mu_0(X)$ satisfies

$$\begin{aligned}
\operatorname{Cov}(\mu_{(1)}(X), \mu_{(0)}(X)) =& \mathbb{E}[\mu_{(1)}(X)\mu_{(0)}(X)] - \mathbb{E}[Y^{(0)}]\mathbb{E}[Y^{(1)}] \\
=& \mathbb{E}[(c_{(1)} + X^\top \beta_{(1)})(c_{(0)} + X^\top \beta_{(0)})] - \mathbb{E}[Y^{(0)}]\mathbb{E}[Y^{(1)}] \\
=& \mathbb{E}[c_{(1)}c_{(0)}] + \mathbb{E}[c_{(1)}X^\top \beta_{(0)}] + \mathbb{E}[c_{(0)}X^\top \beta_{(1)}] \\
& + \mathbb{E}[X^\top \beta_{(0)} X^\top \beta_{(1)}] - \mathbb{E}[Y^{(0)}]\mathbb{E}[Y^{(1)}] \\
=& \mathbb{E}[X^\top \beta_{(0)} X^\top \beta_{(1)}] \\
=& \mathbb{E}\left[ \sum_j X_j \beta_{(0),j} \sum_k X_k \beta_{(1),k} \right] \\
=& \sum_j \sum_k \beta_{(0),j} \beta_{(1),k} \mathbb{E}[X_k X_j] \\
=& \langle \beta_{(0)}, \beta_{(1)} \rangle_\Sigma.
\end{aligned}$$

Therefore,

$$\begin{aligned}
V_{\mathrm{RR,G}} = \tau_{\mathrm{RR}}^2 \operatorname{Var}\left(\frac{\mu_1(X)}{\mathbb{E}\left[Y^{(1)}\right]} - \frac{\mu_0(X)}{\mathbb{E}\left[Y^{(0)}\right]}\right) &= \tau_{\mathrm{RR}}^2 \left(\frac{\operatorname{Var}(\mu_1(X))}{\mathbb{E}\left[Y^{(1)}\right]^2} + \frac{\operatorname{Var}(\mu_0(X))}{\mathbb{E}\left[Y^{(0)}\right]^2} - 2\frac{\operatorname{Cov}(\mu_0(X),\mu_1(X))}{\mathbb{E}\left[Y^{(0)}\right]\mathbb{E}\left[Y^{(1)}\right]}\right) \\
&= \tau_{\mathrm{RR}}^2 \left(\left\|\frac{\beta_{(1)}}{\mathbb{E}\left[Y^{(1)}\right]}\right\|_\Sigma^2 + \left\|\frac{\beta_{(0)}}{\mathbb{E}\left[Y^{(0)}\right]}\right\|_\Sigma^2 - 2\frac{\langle \beta_{(0)}, \beta_{(1)}\rangle_\Sigma}{\mathbb{E}\left[Y^{(0)}\right]\mathbb{E}\left[Y^{(1)}\right]}\right) \\
&= \tau_{\mathrm{RR}}^2 \left\|\frac{\beta_{(1)}}{\mathbb{E}\left[Y^{(1)}\right]} - \frac{\beta_{(0)}}{\mathbb{E}\left[Y^{(0)}\right]}\right\|_\Sigma^2.
\end{aligned}$$

**Last inequality** A simple computation leads to

$$\begin{aligned}
\frac{V_{\mathrm{RR,N}} - V_{\mathrm{RR,G}}}{\tau_{\mathrm{RR}}^2} &= \frac{\left\|\beta_{(1)}\right\|_\Sigma^2 + \sigma^2}{e\mathbb{E}\left[Y^{(1)}\right]^2} + \frac{\left\|\beta_{(0)}\right\|_\Sigma^2 + \sigma^2}{(1-e)\mathbb{E}\left[Y^{(0)}\right]^2} \\
&\quad - \left(\left\|\frac{\beta_{(1)}}{\mathbb{E}\left[Y^{(1)}\right]} - \frac{\beta_{(0)}}{\mathbb{E}\left[Y^{(0)}\right]}\right\|_\Sigma^2 + \sigma^2\left(\frac{1}{e\mathbb{E}\left[Y^{(1)}\right]^2} + \frac{1}{(1-e)\mathbb{E}\left[Y^{(0)}\right]^2}\right)\right) \\
&= \left(\frac{1-e}{e}\right)\frac{\left\|\beta_{(1)}\right\|_\Sigma^2}{\mathbb{E}\left[Y^{(1)}\right]^2} + \left(\frac{e}{1-e}\right)\frac{\left\|\beta_{(0)}\right\|_\Sigma^2}{\mathbb{E}\left[Y^{(0)}\right]^2} + \frac{2\langle \beta_{(1)}, \beta_{(0)}\rangle_\Sigma}{\mathbb{E}\left[Y^{(1)}\right]\mathbb{E}\left[Y^{(0)}\right]} \\
&= e(1-e)\left\|\frac{\beta_{(1)}}{e\mathbb{E}\left[Y^{(1)}\right]} - \frac{\beta_{(0)}}{(1-e)\mathbb{E}\left[Y^{(0)}\right]}\right\|_\Sigma^2.
\end{aligned}$$

□

### 6.3.4 Risk Ratio one-step estimator

*Proof of Definition 4.* We will use Kennedy (2022, 2015) notation in this proof. If you are not familiar on how to compute an influence function, note that it is very similar to compute the derivative of a function. We define our estimand quantity

$$\psi = \frac{\mathbb{E}\left[\mathbb{E}\left[Y|T=1,X\right]\right]}{\mathbb{E}\left[\mathbb{E}\left[Y|T=0,X\right]\right]} = \frac{\psi_1}{\psi_0}.$$

We can now compute the influence function $\varphi$ of $\psi$.

$$\begin{aligned}
\varphi = \mathbb{IF}\left(\psi\right) = \mathbb{IF}\left(\frac{\psi_1}{\psi_0}\right) &= \frac{\mathbb{IF}\left(\psi_1\right)\psi_0 - \mathbb{IF}\left(\psi_0\right)\psi_1}{\psi_0^2} \\
&= \frac{\mathbb{IF}\left(\psi_1\right)}{\psi_0} - \psi\frac{\mathbb{IF}\left(\psi_0\right)}{\psi_0}.
\end{aligned}$$

According to Example 2 in Kennedy (2022), we have

$$\begin{aligned}
\mathbb{IF}\left(\psi_1\right) &= \mu_1(X) + T\frac{Y-\mu_1(X)}{e(X)} - \psi_1 \\
\text{and}\quad \mathbb{IF}\left(\psi_0\right) &= \mu_0(X) + (1-T)\frac{Y-\mu_0(X)}{1-e(X)} - \psi_0.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\varphi &= \frac{\mathbb{IF}(\psi_1)}{\psi_0} - \psi\frac{\mathbb{IF}(\psi_0)}{\psi_0} \\
&= \frac{\mu_1(X) + T\frac{Y-\mu_1(X)}{e(X)} - \psi_1}{\psi_0} - \psi\frac{\mu_0(X) + (1-T)\frac{Y-\mu_0(X)}{1-e(X)} - \psi_0}{\psi_0} \\
&= \frac{\mu_1(X) + T\frac{Y-\mu_1(X)}{e(X)}}{\psi_0} - \psi - \psi\left(\frac{\mu_0(X) + (1-T)\frac{Y-\mu_0(X)}{1-e(X)}}{\psi_0} - 1\right) \\
&= \frac{\mu_1(X) + T\frac{Y-\mu_1(X)}{e(X)}}{\psi_0} - \psi\frac{\mu_0(X) + (1-T)\frac{Y-\mu_0(X)}{1-e(X)}}{\psi_0}.
\end{aligned}$$

As referenced in Kennedy (2022) regarding the semiparametric von Mises expansion, consider the functional $\psi : P \to \mathbb{R}$, where $P$ represents the true data distribution and $\hat{P}$ its estimation. The expansion is formulated as:

$$\psi(\hat{P}) - \psi(P) = \int \varphi(z;\hat{P})d(\hat{P}-P)(z) + R_2(\hat{P},P), \tag{81}$$

for all distributions $\hat{P}$ and $P$. The influence function $\varphi(z;P)$, associated with $\psi$, is a function with zero mean and finite variance as defined by Tsiatis (2006)

$$\int \varphi(z;P)dP(z) = 0 \quad \text{and} \quad \int \varphi(z;P)^2 dP(z) < \infty, \tag{82}$$

and $R_2(\hat{P},P)$ denotes a second-order remainder term. According to the expansion in (81), most plug-in estimators $\psi(\hat{P})$ are biased to the first order, evidenced by:

$$\psi(P) = \psi(\hat{P}) + \int \varphi(z;\hat{P})dP(z) + R_2(\hat{P},P),$$

since $\int \varphi(z;\hat{P})d\hat{P}(z) = 0$. Therefore, a first-order approximation of $\psi(P)$ is given by $\psi(\hat{P}) + \int \varphi(z;\hat{P})dP(z)$ which can be estimated via

$$\begin{aligned}
\hat{\tau}_{\text{RR-OS}} &= \hat{\psi} + \frac{1}{n}\sum_{i=1}^{n}\varphi(Z_i) \\
&= \hat{\psi} + \frac{1}{n}\sum_{i=1}^{n}\frac{\mu_1(X_i) + T_i\frac{Y_i-\mu_1(X_i)}{e(X_i)}}{\hat{\psi}_0} - \hat{\psi}\frac{\mu_0(X_i) + (1-T_i)\frac{Y_i-\mu_0(X_i)}{1-e(X_i)}}{\hat{\psi}_0} \\
&= \hat{\psi}\left(1 - \frac{\frac{1}{n}\sum_{i=1}^{n}\mu_0(X_i) + (1-T_i)\frac{Y_i-\mu_0(X_i)}{1-e(X_i)}}{\hat{\psi}_0}\right) + \frac{\frac{1}{n}\sum_{i=1}^{n}\mu_1(X_i) + T_i\frac{Y_i-\mu_1(X_i)}{e(X_i)}}{\hat{\psi}_0} \\
&= \frac{\sum_{i=1}^{n}\hat{\mu}_1(X_i)}{\sum_{i=1}^{n}\hat{\mu}_0(X_i)}\left(1 - \frac{\sum_{i=1}^{n}\hat{\mu}_0(X_i) + \frac{(1-T_i)(Y_i-\hat{\mu}_0(X_i))}{1-\hat{e}(X_i)}}{\sum_{i=1}^{n}\hat{\mu}_0(X_i)}\right) + \frac{\sum_{i=1}^{n}\hat{\mu}_1(X_i) + \frac{T_i(Y_i-\hat{\mu}_1(X_i))}{\hat{e}(X_i)}}{\sum_{i=1}^{n}\hat{\mu}_0(X_i)}.
\end{aligned}$$

□

*Proof of Proposition 5.*
**Asymptotic bias and variance of the cross-fitted One-step estimator** Recall that

$$\psi(P) = \frac{\mathbb{E}_P\left[\mathbb{E}_P[Y \mid X, T=1]\right]}{\mathbb{E}_P\left[\mathbb{E}_P[Y \mid X, T=0]\right]} = \frac{\psi_1}{\psi_0} \tag{83}$$

$$\psi(\hat{P}) = \frac{\sum_{i=1}^{n}\hat{\mu}_1(X_i)}{\sum_{i=1}^{n}\hat{\mu}_0(X_i)} = \frac{\hat{\psi}_1}{\hat{\psi}_0} \tag{84}$$

$$\varphi(Z;\hat{P}) = \frac{\hat{\mu}_1(X_i) + T_i\frac{Y_i-\hat{\mu}_1(X_i)}{\hat{e}(X_i)}}{\hat{\psi}_0} - \hat{\psi}\frac{\hat{\mu}_0(X_i) + (1-T_i)\frac{Y_i-\hat{\mu}_0(X_i)}{1-\hat{e}(X_i)}}{\hat{\psi}_0} \tag{85}$$

where $P$ represents the true underlying data distribution and $\hat{P}$ the distribution where oracle quantities have been replaced by plug-in estimates. We express $\psi(P)$ as follows:

$$\psi(P) = \psi(\hat{P}) + \int \varphi(z;\hat{P})dP(z) + R_2(\hat{P}, P),$$

where $R_2$ encapsulates higher order remainder terms.

To elucidate, we rearrange to find $\psi(\hat{P}) - \psi(P)$:

$$\begin{aligned}\psi(\hat{P}) - \psi(P) &= R_2(P,\hat{P}) - \int \varphi(z;\hat{P})dP(z) \\ &= \frac{1}{n}\sum_{i=1}^{n}\varphi(Z_i;P) - \frac{1}{n}\sum_{i=1}^{n}\varphi(Z_i;\hat{P}) \\ &\quad + \frac{1}{n}\sum_{i=1}^{n}\Big(\varphi(Z_i;\hat{P}) - \varphi(Z_i;P)\Big) - \int \Big(\varphi(z;\hat{P}) - \varphi(z;P)\Big)\,dP(z) \\ &\quad + R_2(P,\hat{P}).\end{aligned}$$

Recalling that $\hat{\tau}_{\text{RR-OS}} = \psi(\hat{P}) + \frac{1}{n}\sum_{i=1}^{n}\varphi(Z_i;\hat{P})$ and $\tau_{\text{RR}} = \psi(P)$, we have

$$\begin{aligned}\hat{\tau}_{\text{RR-OS}} - \tau_{\text{RR}} &= \psi(\hat{P}) + \frac{1}{n}\sum_{i=1}^{n}\varphi(Z_i;\hat{P}) - \psi(P) && (86)\\ &= \frac{1}{n}\sum_{i=1}^{n}\varphi(Z_i;P) && (87)\\ &\quad + \frac{1}{n}\sum_{i=1}^{n}\Big(\varphi(Z_i;\hat{P}) - \varphi(Z_i;P)\Big) - \int \Big(\varphi(z;\hat{P}) - \varphi(z;P)\Big)\,dP(z) && (88)\\ &\quad + R_2(P,\hat{P}). && (89)\end{aligned}$$

The first term is a sample average of centered i.i.d. terms since, by definition (82), $\int \varphi(z;P)dP(z) = 0$. According to the central limit theorem, it converges to a normally distributed random variable with variance $\text{Var}(\varphi(Z))/n$.

Following the work of Vaart (1998), we consider the second term in (89), that is

$$\frac{1}{n}\sum_{i=1}^{n}\Big(\varphi(Z_i;\hat{P}) - \varphi(Z_i;P)\Big) - \int \Big(\varphi(z;\hat{P}) - \varphi(z;P)\Big)\,dP(z).$$

Since our estimator is built on a cross-fitting strategy with $K$ folds $\mathcal{I}_1, \dots \mathcal{I}_K$, containing respectively $n_1, \dots, n_K$ observations, the above quantity may be written as

$$\frac{1}{n}\sum_{k=1}^{K}\sum_{i\in\mathcal{I}_k}\Big(\varphi(Z_i;\hat{P}^{-k}) - \varphi(Z_i;P)\Big) - \int \Big(\varphi(z;\hat{P}) - \varphi(z;P)\Big)\,dP(z),$$

where $\hat{P}^{-k}$ corresponds to a data distribution where oracle quantity are replaced by plug-in estimates built on all observations except those in $\mathcal{I}_k$. We denote this set of observations as $\mathcal{I}_{-k}$. We let $\hat{\varphi}^{-k}(Z) = \varphi(Z;\hat{P}^{-k})$ and

$$U_k = \Big(\mathbb{P}_n^{(k)} - P\Big)\left(\hat{\varphi}^{-k}(Z) - \varphi(Z)\right), \tag{90}$$

where $\mathbb{P}_n^{(k)}$ is the empirical measure over $\mathcal{I}_k$. The quantity of interest can thus be written as

$$\frac{1}{n}\sum_{k=1}^{K}\sum_{i\in\mathcal{I}_k}\left(\varphi^{-k}(Z_i) - \varphi(Z_i;P)\right) - \int \Big(\varphi(z;\hat{P}) - \varphi(z;P)\Big)\,dP(z) = \frac{1}{n}\sum_{k=1}^{K} n_k U_k. \tag{91}$$

The expectation and variance of $U_k$ satisfy

$$\begin{aligned}\mathbb{E}\left[U_k \mid \mathcal{I}_{-k}\right] &= \mathbb{E}\left[\left(\mathbb{P}_n^{(k)} - P\right)\left(\hat{\varphi}^{-k} - \varphi\right) \mid \mathcal{I}_{-k}\right] &(92)\\ &= \mathbb{E}\left[\mathbb{P}_n^{(k)}(\hat{\varphi}^{-k} - \varphi) \mid \mathcal{I}_{-k}\right] - \mathbb{E}\left[P(\hat{\varphi}^{-k} - \varphi) \mid \mathcal{I}_{-k}\right] &(93)\\ &= \mathbb{E}\left[\hat{\varphi}^{-k}(Z) - \varphi(Z)\right] - \mathbb{E}\left[\hat{\varphi}^{-k}(Z) - \varphi(Z)\right] &(94)\\ &= 0, &(95)\end{aligned}$$

and

$$\begin{aligned}\mathrm{Var}\left[U_k \mid \mathcal{I}_{-k}\right] &= \mathrm{Var}\left[\left(\mathbb{P}_n^{(k)} - P\right)\left(\hat{\varphi}^{-k} - \varphi\right) \mid \mathcal{I}_{-k}\right] &(96)\\ &= \mathrm{Var}\left[\mathbb{P}_n^{(k)}(\hat{\varphi}^{-k} - \varphi) - P(\hat{\varphi}^{-k} - \varphi) \mid \mathcal{I}_{-k}\right] &(97)\\ &= \mathrm{Var}\left[\frac{1}{n_k}\sum_{i=1}^{n_k}\left(\hat{\varphi}^{-k}(Z_i) - \varphi(Z_i)\right) \mid \mathcal{I}_{-k}\right] &(98)\\ &= \frac{1}{n_k}\mathrm{Var}\left[\hat{\varphi}^{-k}(Z) - \varphi(Z) \mid \mathcal{I}_{-k}\right] &(99)\\ &\leq \frac{1}{n_k}\mathbb{E}\left[(\hat{\varphi}^{-k}(Z) - \varphi(Z))^2 \mid \mathcal{I}_{-k}\right]. &(100)\end{aligned}$$

Let $a > 0$. Applying Chebyshev's inequality leads to

$$\begin{aligned}&\mathbb{P}\left(\frac{|U_k - \mathbb{E}[U_k \mid \mathcal{I}_{-k}]|}{\sqrt{\mathrm{Var}[U_k \mid \mathcal{I}_{-k}]}} \geq a \mid \mathcal{I}_{-k}\right) \leq \frac{1}{a^2} &(101)\\ \Longleftrightarrow &\mathbb{P}\left(\frac{|U_k|}{\sqrt{\mathrm{Var}[U_k \mid \mathcal{I}_{-k}]}} \geq a \mid \mathcal{I}_{-k}\right) \leq \frac{1}{a^2}. &(102)\end{aligned}$$

Thus,

$$\mathbb{P}\left(\frac{|U_k|\sqrt{n_k}}{\sqrt{\mathbb{E}\left[(\hat{\varphi}^{-k}(Z) - \varphi(Z))^2 \mid \mathcal{I}_{-k}\right]}} \geq a \mid \mathcal{I}_{-k}\right) \leq \mathbb{P}\left(\frac{|U_k|}{\sqrt{\mathrm{Var}[U_k \mid \mathcal{I}_{-k}]}} \geq a \mid \mathcal{I}_{-k}\right) \leq \frac{1}{a^2}, \tag{103}$$

which leads to

$$\mathbb{P}\left(|U_k|\sqrt{n_k} \geq a \mid \mathcal{I}_{-k}\right) \leq \frac{\mathbb{E}\left[(\hat{\varphi}^{-k}(Z) - \varphi(Z))^2 \mid \mathcal{I}_{-k}\right]}{a^2}. \tag{104}$$

Finally, taking the expectation on both sides leads to

$$\mathbb{P}\left(|U_k|\sqrt{n_k} \geq a\right) \leq \frac{\mathbb{E}\left[(\hat{\varphi}^{-k}(Z) - \varphi(Z))^2\right]}{a^2}. \tag{105}$$

According to (91), the quantity of interest takes the form

$$\frac{1}{n}\sum_{k=1}^{K} n_k U_k = \sum_{k=1}^{K}\frac{n_k}{n}U_k. \tag{106}$$

Hence,

$$\mathbb{P}\left(\sqrt{n}\frac{n_k}{n}|U_k| \leq a\frac{\sqrt{n_k}}{\sqrt{n}}\right) \geq 1 - \frac{\mathbb{E}\left[(\hat{\varphi}^{-k}(Z) - \varphi(Z))^2\right]}{a^2}. \tag{107}$$

Therefore,

$$\begin{aligned}&\mathbb{P}\left(\sqrt{n}\sum_{k=1}^{K}\frac{n_k}{n}|U_k| \leq a\sum_{k=1}^{K}\frac{\sqrt{n_k}}{\sqrt{n}}\right) \geq 1 - \sum_{k=1}^{K}\frac{\mathbb{E}\left[(\hat{\varphi}^{-k}(Z) - \varphi(Z))^2\right]}{a^2} &(108)\\ \Rightarrow\quad &\mathbb{P}\left(\sqrt{n}\sum_{k=1}^{K}\frac{n_k}{n}|U_k| \leq aK\right) \geq 1 - \sum_{k=1}^{K}\frac{\mathbb{E}\left[(\hat{\varphi}^{-k}(Z) - \varphi(Z))^2\right]}{a^2}, &(109)\end{aligned}$$

which proves that $\sum_{k=1}^{K} \frac{n_k}{n} U_k = o_P(1/\sqrt{n})$ as $K$ is fixed and $\varphi^{-k}$ is $L^2$ consistent.

Regarding the last term, note that

$$
\begin{aligned}
R_2(P, \hat{P}) &= \psi(\hat{P}) - \psi(P) + \int \varphi(z; \hat{P}) dP(z) && (110)\\
&= \psi(\hat{P}) - \psi(P) + \mathbb{E}[\varphi(Z; \hat{P})] && (111)\\
&= \psi(\hat{P}) - \psi(P) + \mathbb{E}\left[\frac{\hat{\mu}_1(X) + T\frac{Y-\hat{\mu}_1(X)}{\hat{e}(X)}}{\hat{\psi}_0} - \hat{\psi}\frac{\hat{\mu}_0(X) + (1-T)\frac{Y-\hat{\mu}_0(X)}{1-\hat{e}(X)}}{\hat{\psi}_0}\right] && (112)\\
&= \psi(\hat{P}) - \psi(P) + \frac{\mathbb{E}\left[\hat{\mu}_1(X) + T\frac{Y-\hat{\mu}_1(X)}{\hat{e}(X)}\right]}{\hat{\psi}_0} - \hat{\psi}\frac{\mathbb{E}\left[\hat{\mu}_0(X) + (1-T)\frac{Y-\hat{\mu}_0(X)}{1-\hat{e}(X)}\right]}{\hat{\psi}_0} && (113)\\
&= \psi(\hat{P}) - \psi(P) + \frac{\mathbb{E}\left[\hat{\mu}_1(X) - \mu_1(X) + T\frac{Y-\hat{\mu}_1(X)}{\hat{e}(X)}\right]}{\hat{\psi}_0} + \frac{\psi_1}{\hat{\psi}_0} && (114)\\
&\quad - \hat{\psi}\frac{\mathbb{E}\left[\hat{\mu}_0(X) - \mu_0(X) + (1-T)\frac{Y-\hat{\mu}_0(X)}{1-\hat{e}(X)}\right]}{\hat{\psi}_0} - \hat{\psi}\frac{\psi_0}{\hat{\psi}_0}. && (115)
\end{aligned}
$$

Note that

$$
\begin{aligned}
\mathbb{E}\left[\hat{\mu}_1(X) - \mu_1(X) + T\frac{Y-\hat{\mu}_1(X)}{\hat{e}(X)}\right] &= \mathbb{E}\left[\frac{1}{\hat{e}(X)}(\mu_1(X) - \hat{\mu}_1(X))(\hat{e}(X) - e(X))\right] && (116)\\
\text{Positivity} \qquad &\leq \frac{1}{\eta}\mathbb{E}\left[(\mu_1(X) - \hat{\mu}_1(X))(\hat{e}(X) - e(X))\right] && (117)\\
\text{Cauchy-Schwarz} \qquad &\leq \frac{1}{\eta}\mathbb{E}\left[(\hat{e}(X) - e(X))^2\right]^{1/2}\mathbb{E}\left[(\hat{\mu}_1(X) - \mu_1(X))^2\right]^{1/2} && (118)\\
&= o_p\left(\frac{1}{\sqrt{n}}\right). && (119)
\end{aligned}
$$

Similarly,

$$
\mathbb{E}\left[\hat{\mu}_1(X) - \mu_1(X) + T\frac{Y-\hat{\mu}_1(X)}{\hat{e}(X)}\right] = o_p\left(\frac{1}{\sqrt{n}}\right). \qquad (120)
$$

For the last term in (115), since $\psi = \psi_1/\psi_0$ and $\hat{\psi} = \hat{\psi}_1/\hat{\psi}_0$,

$$
\begin{aligned}
\hat{\psi} - \psi + \frac{\psi_1}{\hat{\psi}_0} - \hat{\psi}\frac{\psi_0}{\hat{\psi}_0} &= \psi_1\left(\frac{1}{\hat{\psi}_0} - \frac{1}{\psi_0}\right) + \hat{\psi}\left(1 - \frac{\psi_0}{\hat{\psi}_0}\right)\\
&= \psi_1\frac{\psi_0 - \hat{\psi}_0}{\psi_0\hat{\psi}_0} + \hat{\psi}\left(\frac{\hat{\psi}_0 - \psi_0}{\hat{\psi}_0}\right)\\
&= \left(\frac{\hat{\psi}_0 - \psi_0}{\hat{\psi}_0}\right)\left(\hat{\psi} - \psi\right)\\
&= \left(\frac{\hat{\psi}_0 - \psi_0}{\hat{\psi}_0}\right)\left(\left(\frac{1}{\hat{\psi}_0} - \frac{1}{\psi_0}\right)\hat{\psi}_1 + \frac{1}{\psi_0}\left(\hat{\psi}_1 - \psi_1\right)\right)\\
&= \frac{1}{\psi_0\hat{\psi}_0}\left((\hat{\psi}_0 - \psi_0)(\hat{\psi}_1 - \psi_1) - \hat{\psi}(\hat{\psi}_0 - \psi_0)(\hat{\psi}_0 - \psi_0)\right).
\end{aligned}
$$

By assumption, we have

$$
\begin{aligned}
(\hat{\psi}_0 - \psi_0)(\hat{\psi}_1 - \psi_1) &= \mathbb{E}\left[\hat{\mu}_0(X) - \mu_0(X)\right]\mathbb{E}\left[\hat{\mu}_1(X) - \mu_1(X)\right] && (121)\\
&\leq (\mathbb{E}\left[(\hat{\mu}_0(X) - \mu_0(X))^2\right])^{1/2}(\mathbb{E}\left[(\hat{\mu}_1(X) - \mu_1(X))^2\right])^{1/2} && (122)\\
&= o_p\left(\frac{1}{\sqrt{n}}\right) && (123)
\end{aligned}
$$

and

$$
\begin{aligned}
(\hat{\psi}_0 - \psi_0)(\hat{\psi}_0 - \psi_0) &= \left(\mathbb{E}\left[\hat{\mu}_0(X)\right] - \mathbb{E}\left[\mu_0(X)\right]\right)^2 \\
&= o_p\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}
$$

By assumption, $\mathbb{E}\left[(\hat{\mu}_0(X) - \mu_0(X))^2\right]$ tends to zero. Thus, $\hat{\psi}_0 = \mathbb{E}[\hat{\mu}_0]$ tends to $\psi_0 = \mathbb{E}[\mu_0(X)]$. Thus,

$$
\hat{\psi} - \psi + \frac{\psi_1}{\hat{\psi}_0} - \hat{\psi}\frac{\psi_0}{\hat{\psi}_0} = o_p\left(\frac{1}{\sqrt{n}}\right),
$$

which implies that $R_2(P, \hat{P}) = o_p\left(n^{-1/2}\right)$. Finally,

$$
\sqrt{n}\left(\hat{\tau}_{\text{RR-OS}} - \tau_{\text{RR}}\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varphi(Z_i; P) + o_p\left(\frac{1}{\sqrt{n}}\right)
$$

and thus

$$
\sqrt{n}\left(\hat{\tau}_{\text{RR-OS}} - \tau_{\text{RR}}\right) \xrightarrow{d} \mathcal{N}\left(0, \operatorname{Var}(\varphi)\right),
$$

where

$$
\begin{aligned}
\operatorname{Var}(\varphi) &= \operatorname{Var}\left(\frac{\mu_1(X) + T\frac{Y-\mu_1(X)}{e(X)}}{\psi_0} - \psi\frac{\mu_0(X) + (1-T)\frac{Y-\mu_0(X)}{1-e(X)}}{\psi_0}\right) && (124) \\
&= \psi^2 \operatorname{Var}\left(\frac{\mu_1(X) + T\frac{Y-\mu_1(X)}{e(X)}}{\psi_1}\frac{\mu_0(X) + (1-T)\frac{Y-\mu_0(X)}{1-e(X)}}{\psi_0}\right) && (125) \\
&= \tau_{\text{RR}}^2 \operatorname{Var}\left(\frac{g_1(Z)}{\mathbb{E}\left[Y^{(1)}\right]} - \frac{g_0(Z)}{\mathbb{E}\left[Y^{(0)}\right]}\right). && (126)
\end{aligned}
$$

Using Bienaymé's identity, we get

$$
\begin{aligned}
\operatorname{Var}\left(\frac{g_1(Z)}{\mathbb{E}\left[Y^{(1)}\right]} - \frac{g_0(Z)}{\mathbb{E}\left[Y^{(0)}\right]}\right) &= \operatorname{Var}\left(\frac{\mu_1(X)}{\mathbb{E}\left[Y^{(1)}\right]} - \frac{\mu_0(X)}{\mathbb{E}\left[Y^{(0)}\right]}\right) + \operatorname{Var}\left(\frac{T(Y-\mu_1(X))}{\mathbb{E}\left[Y^{(1)}\right]e(X)} - \frac{(1-T)(Y-\mu_0(X))}{\mathbb{E}\left[Y^{(0)}\right](1-e(X))}\right) && (127) \\
&\quad + 2\operatorname{Cov}\left(\frac{\mu_1(X)}{\mathbb{E}\left[Y^{(1)}\right]} - \frac{\mu_0(X)}{\mathbb{E}\left[Y^{(0)}\right]}; \frac{T(Y-\mu_1(X))}{\mathbb{E}\left[Y^{(1)}\right]e(X)} - \frac{(1-T)(Y-\mu_0(X))}{\mathbb{E}\left[Y^{(0)}\right](1-e(X))}\right). && (128)
\end{aligned}
$$

The second term can be rewritten as

$$
\begin{aligned}
&\operatorname{Var}\left(\frac{T(Y-\mu_1(X))}{\mathbb{E}\left[Y^{(1)}\right]e(X)} - \frac{(1-T)(Y-\mu_0(X))}{\mathbb{E}\left[Y^{(0)}\right](1-e(X))}\right) && (129) \\
&= \operatorname{Var}\left(\frac{T(Y-\mu_1(X))}{\mathbb{E}\left[Y^{(1)}\right]e(X)}\right) + \operatorname{Var}\left(\frac{(1-T)(Y-\mu_0(X))}{\mathbb{E}\left[Y^{(0)}\right](1-e(X))}\right) - 2\operatorname{Cov}\left(\frac{T(Y-\mu_1(X))}{\mathbb{E}\left[Y^{(1)}\right]e(X)}, \frac{(1-T)(Y-\mu_0(X))}{\mathbb{E}\left[Y^{(0)}\right](1-e(X))}\right), && (130)
\end{aligned}
$$

with

$$
\operatorname{Var}\left(\frac{T(Y-\mu_1(X))}{\mathbb{E}\left[Y^{(1)}\right]e(X)}\right) = \mathbb{E}\left[\left(\frac{T(Y-\mu_1(X))}{\mathbb{E}\left[Y^{(1)}\right]e(X)}\right)^2\right] - \mathbb{E}\left[\frac{T(Y-\mu_1(X))}{\mathbb{E}\left[Y^{(1)}\right]e(X)}\right]^2. \qquad (131)
$$

For the first term in (131),

$$
\begin{aligned}
&\mathbb{E}\left[\left(T\frac{Y-\mu_1(X)}{e(X)\mathbb{E}\left[Y^{(1)}\right]}\right)^2\right] \\
&= \mathbb{E}\left[\left(T\frac{Y^{(1)}-\mu_1(X)}{e(X)\mathbb{E}\left[Y^{(1)}\right]}\right)^2\right] && \text{Consistency} \\
&= \mathbb{E}\left[\mathbb{E}\left[\left(T\frac{Y^{(1)}-\mu_1(X)}{e(X)\mathbb{E}\left[Y^{(1)}\right]}\right)^2 \mid X\right]\right] && \text{Total expectation} \\
&= \mathbb{E}\left[\mathbb{E}\left[T\left(\frac{Y^{(1)}-\mu_1(X)}{e(X)\mathbb{E}\left[Y^{(1)}\right]}\right)^2 \mid X\right]\right] && \text{T is binary} \\
&= \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}_{\{T=1\}}\left(\frac{Y^{(1)}-\mu_1(X)}{e(X)\mathbb{E}\left[Y^{(1)}\right]}\right)^2 \mid X\right]\right] && \text{T written as an indicator} \\
&= \mathbb{E}\left[\frac{1}{e(X)^2\mathbb{E}\left[Y^{(1)}\right]^2}\mathbb{E}\left[\mathbf{1}_{\{T=1\}}\left(Y^{(1)}-\mu_1(X)\right)^2 \mid X\right]\right] && e(X) \text{ is a function of } X \\
&= \mathbb{E}\left[\frac{\mathrm{Var}\left(Y^{(1)}|X\right)}{e(X)^2\mathbb{E}\left[Y^{(1)}\right]^2}\mathbb{E}\left[\mathbf{1}_{\{T=1\}} \mid X\right]\right] && \text{Uncounf. \& } \mu_1(.) \text{ is func. of } X \\
&= \mathbb{E}\left[\frac{\mathrm{Var}\left(Y^{(1)}|X\right)}{e(X)^2\mathbb{E}\left[Y^{(1)}\right]^2}e(X)\right] && \text{Definition of } e(X) \\
&= \mathbb{E}\left[\frac{\mathrm{Var}\left(Y^{(1)}|X\right)}{e(X)\mathbb{E}\left[Y^{(1)}\right]^2}\right].
\end{aligned}
$$

For the second term in (131),

$$
\begin{aligned}
&\mathbb{E}\left[\frac{T(Y-\mu_1(X))}{\mathbb{E}\left[Y^{(1)}\right]e(X)}\right] \\
&= \mathbb{E}\left[\frac{T(Y^{(1)}-\mu_1(X))}{\mathbb{E}\left[Y^{(1)}\right]e(X)}\right] && \text{Consistency} \\
&= \mathbb{E}\left[\mathbb{E}\left[T\frac{Y^{(1)}-\mu_1(X)}{e(X)\mathbb{E}\left[Y^{(1)}\right]} \mid X\right]\right] && \text{Total expectation} \\
&= \mathbb{E}\left[\frac{1}{e(X)\mathbb{E}\left[Y^{(1)}\right]}\mathbb{E}\left[T(Y^{(1)}-\mu_1(X)) \mid X\right]\right] && e(X) \text{ is a function of } X \\
&= \mathbb{E}\left[\frac{e(X)}{e(X)\mathbb{E}\left[Y^{(1)}\right]}(\mu_1(X)-\mu_1(X))\right] && \text{Uncounf. \& } \mu_1(.) \text{ is func. of } X \\
&= 0.
\end{aligned}
$$

Therefore

$$
\mathrm{Var}\left(\frac{T(Y-\mu_1(X))}{\mathbb{E}\left[Y^{(1)}\right]e(X)}\right) = \mathbb{E}\left[\frac{\mathrm{Var}\left(Y^{(1)}|X\right)}{e(X)\mathbb{E}\left[Y^{(1)}\right]^2}\right],
$$

and similarly

$$
\mathrm{Var}\left(\frac{(1-T)(Y-\mu_0(X))}{\mathbb{E}\left[Y^{(0)}\right](1-e(X))}\right) = \mathbb{E}\left[\frac{\mathrm{Var}\left(Y^{(0)}|X\right)}{(1-e(X))\mathbb{E}\left[Y^{(0)}\right]^2}\right].
$$

Besides,

$$\begin{aligned}\operatorname{Cov}\left(\frac{T(Y-\mu_1(X))}{\mathbb{E}\left[Y^{(1)}\right]e(X)},\frac{(1-T)(Y-\mu_0(X))}{\mathbb{E}\left[Y^{(0)}\right](1-e(X))}\right) &= \mathbb{E}\left[\frac{T(Y-\mu_1(X))}{\mathbb{E}\left[Y^{(1)}\right]e(X)}\frac{(1-T)(Y-\mu_0(X))}{\mathbb{E}\left[Y^{(0)}\right](1-e(X))}\right]\\ &\quad - \mathbb{E}\left[\frac{T(Y-\mu_1(X))}{\mathbb{E}\left[Y^{(1)}\right]e(X)}\right]\mathbb{E}\left[\frac{(1-T)(Y-\mu_0(X))}{\mathbb{E}\left[Y^{(0)}\right](1-e(X))}\right]\\ &= 0.\end{aligned}$$

Gathering all these results into (130), we obtain

$$\operatorname{Var}\left(\frac{T(Y-\mu_1(X))}{\mathbb{E}\left[Y^{(1)}\right]e(X)}-\frac{(1-T)(Y-\mu_0(X))}{\mathbb{E}\left[Y^{(0)}\right](1-e(X))}\right) = \mathbb{E}\left[\frac{\operatorname{Var}\left(Y^{(1)}|X\right)}{e(X)\mathbb{E}\left[Y^{(1)}\right]^2}\right]+\mathbb{E}\left[\frac{\operatorname{Var}\left(Y^{(0)}|X\right)}{(1-e(X))\mathbb{E}\left[Y^{(0)}\right]^2}\right].$$

In order to rewrite the last term in (128), note that

$$\begin{aligned}\operatorname{Cov}\left(\mu_1(X),\frac{T(Y-\mu_1(X))}{e(X)}\right) &= \mathbb{E}\left[\mu_1(X)\frac{T(Y-\mu_1(X))}{e(X)}\right]-\mathbb{E}\left[\mu_1(X)\right]\mathbb{E}\left[\frac{T(Y-\mu_1(X))}{e(X)}\right]\\ &= \mathbb{E}\left[\mu_1(X)\frac{T(Y-\mu_1(X))}{e(X)}\right]\\ &= \mathbb{E}\left[\frac{\mu_1(X)}{e(X)}\mathbb{E}\left[T(Y^{(1)}-\mu_1(X))|X\right]\right]\\ &= \mathbb{E}\left[\frac{\mu_1(X)}{e(X)}\mathbb{E}\left[\frac{T(Y^{(1)}-\mu_1(X))}{e(X)}|X\right]\right]\\ &= \mathbb{E}\left[\frac{\mu_1(X)e(X)}{e(X)}(\mu_1(X))-\mu_1(X))\right]\\ &= 0.\end{aligned}$$

Similar calculations leads to

$$\operatorname{Cov}\left(\frac{\mu_1(X)}{\mathbb{E}\left[Y^{(1)}\right]}-\frac{\mu_0(X)}{\mathbb{E}\left[Y^{(0)}\right]};\frac{T(Y-\mu_1(X))}{\mathbb{E}\left[Y^{(1)}\right]e(X)}-\frac{(1-T)(Y-\mu_0(X))}{\mathbb{E}\left[Y^{(0)}\right](1-e(X))}\right)=0.$$

Finally

$$V_{\mathrm{RR,OS}} = \tau_{\mathrm{RR}}^2\left(\operatorname{Var}\left(\frac{\mu_1(X)}{\mathbb{E}\left[Y^{(1)}\right]}-\frac{\mu_0(X)}{\mathbb{E}\left[Y^{(0)}\right]}\right)+\mathbb{E}\left[\frac{\operatorname{Var}\left(Y^{(1)}|X\right)}{e(X)\mathbb{E}\left[Y^{(1)}\right]^2}\right]+\mathbb{E}\left[\frac{\operatorname{Var}\left(Y^{(0)}|X\right)}{(1-e(X))\mathbb{E}\left[Y^{(0)}\right]^2}\right]\right).$$

An estimator $\hat{V}_{RR,OS}$ can be derived as follows:

$$\hat{V}_{RR,OS} = \frac{\hat{\tau}^2_{\mathrm{RR,OS,n}}}{n}\sum_{i=1}^{n}\left(\Delta_i-\frac{1}{n}\sum_{i=1}^{n}\Delta_i\right)^2 \tag{132}$$

where

$$\Delta_i = \frac{\hat{\Gamma}_i(1)}{\hat{S}(1)}-\frac{\hat{\Gamma}_i(0)}{\hat{S}(0)}$$

with the intermediate for $t\in\{0,1\}$ quantities defined as:

$$\hat{\Gamma}_i(t) = \hat{\mu}_t(X_i)+\mathbb{1}_{T_i=t}\frac{Y_i-\hat{\mu}_t(X_i)}{\hat{e}_t(X_i)} \quad \text{and} \quad \hat{S}(t)=\frac{1}{n}\sum_{j=1}^{n}\hat{\Gamma}_j(t)$$

□

#### 6.3.5 Risk Ratio Augmented Inverse Propensity Weighting

*Proof of Proposition 6.* We use the derivations established in the proof of Proposition 5. Indeed, we showed in Section 6.3.4 that the influence function $\varphi$ of $\psi = \frac{\mathbb{E}[\mathbb{E}[Y|T=1,X]]}{\mathbb{E}[\mathbb{E}[Y|T=0,X]]}$ can be written:

$$\varphi(Z;P) = \frac{\mu_1(X) + T\frac{Y-\mu_1(X)}{e(X)}}{\psi_0} - \psi\frac{\mu_0(X) + (1-T)\frac{Y-\mu_0(X)}{1-e(X)}}{\psi_0}.$$

Using Equation (81), and knowing that $\int \varphi(z;\hat{P})d\hat{P}(z) = 0$, we have:

$$\begin{aligned}
\psi(\hat{P}) - \psi(P) &= \int \varphi(z;\hat{P})d(\hat{P}-P)(z) + R_2(\hat{P},P) \\
&= R_2(P,\hat{P}) - \int \varphi(z;\hat{P})dP(z) \\
&= \frac{1}{n}\sum_{i=1}^{n}\varphi(Z_i;P) - \frac{1}{n}\sum_{i=1}^{n}\varphi(z_i;\hat{P}) \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\left(\varphi(Z_i;\hat{P}) - \varphi(Z_i;P)\right) - \int \left(\varphi(z;\hat{P}) - \varphi(z;P)\right)dP(z) \\
&\quad + R_2(P,\hat{P}).
\end{aligned}$$

As outlined in A.Schuler (2024), in the estimating equation approach, we assume that the efficient influence function for any given distribution depends solely on the target parameter $\psi$ and a set of nuisance parameters $\eta$. Therefore, instead of denoting the efficient influence function as $\varphi(z;P)$, we can express it as $\varphi(Z;\psi,\eta)$. If the influence function can be represented in this form, we proceed by first estimating $\hat{\eta} = (\hat{e},\hat{\mu}_1,\hat{\mu}_0)$ with crossfitting. For any fixed value $\hat{\eta}$, we find a value $\hat{\psi}$ such that $P_n\varphi_{\hat{\psi},\hat{\eta}} = 0$, that is

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\hat{\mu}_1(X_i) + T_i\frac{Y_i-\hat{\mu}_1(X_i)}{\hat{e}(X_i)}}{\hat{\psi}_0} - \hat{\psi}\frac{\hat{\mu}_0(X_i) + (1-T_i)\frac{Y_i-\hat{\mu}_0(X_i)}{1-\hat{e}(X_i)}}{\hat{\psi}_0} = 0,$$

which implies

$$\hat{\psi} = \frac{\sum_{i=1}^{n}\hat{\mu}_1(X_i) + T_i\frac{Y_i-\hat{\mu}_1(X_i)}{\hat{e}(X_i)}}{\sum_{i=1}^{n}\hat{\mu}_0(X_i) + (1-T_i)\frac{Y_i-\hat{\mu}_0(X_i)}{1-\hat{e}(X_i)}}.$$

Using this $\hat{\psi}$ we have that

$$\begin{aligned}
\hat{\psi} - \psi &= \frac{1}{n}\sum_{i=1}^{n}\varphi(Z_i;P) - \frac{1}{n}\sum_{i=1}^{n}\varphi(Z_i;\hat{P}) + R_2(P,\hat{P}) \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\left(\varphi(Z_i;\hat{P}) - \varphi(Z_i;P)\right) - \int \left(\varphi(z;\hat{P}) - \varphi(z;P)\right)dP(z) \\
&= \frac{1}{n}\sum_{i=1}^{n}\varphi(Z_i;P) + R_2(P,\hat{P}) \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\left(\varphi(Z_i;\hat{P}) - \varphi(Z_i;P)\right) - \int \left(\varphi(z;\hat{P}) - \varphi(z;P)\right)dP(z).
\end{aligned}$$

As detailed in Section 6.3.4, we have $R_2(P,\hat{P}) = o_p\left(n^{-1/2}\right)$ and

$$\frac{1}{n}\sum_{i=1}^{n}\left(\varphi(Z_i;\hat{P}) - \varphi(Z_i;P)\right) - \int \left(\varphi(z;\hat{P}) - \varphi(z;P)\right)dP(z) = o_P(1/\sqrt{n}).$$

Therefore,

$$\sqrt{n}\left(\hat{\tau}_{\text{RR-AIPW}} - \tau_{\text{RR}}\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varphi(z_i;P) + o_p\left(\frac{1}{\sqrt{n}}\right),$$

which leads to

$$\sqrt{n}\left(\hat{\tau}_{\text{RR-AIPW}} - \tau_{\text{RR}}\right) \xrightarrow{d} \mathcal{N}\left(0, \text{Var}(\varphi)\right),$$

where $\text{Var}(\varphi) = V_{\text{RR,OS}}$

□

Hereafter, we propose another proof of Proposition 6 which does not use the influence function theory

*Proof of Proposition 6.*

**Asymptotic bias and variance of the crossfitted Ratio AIPW estimator** In this alternative proof, we further assume that $\text{Var}[Y|X] \leq \sigma^2$ for some $\sigma > 0$. Recall that we want to analyze $\sqrt{n}\left(\hat{\tau}_{\text{RR,AIPW}} - \tau_{\text{RR}}\right)$. Letting

$$\tau^{\star}_{\text{RR,AIPW}} = \frac{\sum_{i=1}^{n}\mu_1(X_i) + \frac{T_i(Y_i - \mu_1(X_i))}{e(X_i)}}{\sum_{i=1}^{n}\mu_0(X_i) + \frac{(1-T_i)(Y_i-\mu_0(X_i))}{1-e(X_i)}} := \frac{\tau^{\star}_{\text{RR,AIPW, 1}}}{\tau^{\star}_{\text{RR,AIPW, 0}}} \tag{133}$$

be the oracle version of $\hat{\tau}_{\text{RR,AIPW}}$ where the propensity score and both response surfaces are assumed to be known, we can rewrite

$$\hat{\tau}_{\text{RR,AIPW}} - \tau_{\text{RR}} = \hat{\tau}_{\text{RR,AIPW}} - \tau^{\star}_{\text{RR,AIPW}} + \tau^{\star}_{\text{RR,AIPW}} - \tau_{\text{RR}}. \tag{134}$$

Regarding the first term in (134), we have

$$\begin{aligned}
&\left|\hat{\tau}_{\text{RR,AIPW}} - \tau^{\star}_{\text{RR,AIPW}}\right| && (135)\\
&= \left|\frac{\hat{\tau}_{\text{RR,AIPW, 1}}}{\hat{\tau}_{\text{RR,AIPW, 0}}} - \frac{\tau^{\star}_{\text{RR,AIPW, 1}}}{\tau^{\star}_{\text{RR,AIPW, 0}}}\right| && (136)\\
&= \left|\left(\left(\hat{\tau}_{\text{RR,AIPW, 0}}\right)^{-1} - \left(\tau^{\star}_{\text{RR,AIPW, 0}}\right)^{-1}\right)\hat{\tau}_{\text{RR,AIPW, 1}}\right. && (137)\\
&\quad \left. + \left(\tau^{\star}_{\text{RR,AIPW, 0}}\right)^{-1}\left(\hat{\tau}_{\text{RR,AIPW, 1}} - \tau^{\star}_{\text{RR,AIPW, 1}}\right)\right| && (138)\\
&\leq \left|\left(\left(\hat{\tau}_{\text{RR,AIPW, 0}}\right)^{-1} - \left(\tau^{\star}_{\text{RR,AIPW, 0}}\right)^{-1}\right)\hat{\tau}_{\text{RR,AIPW, 1}}\right| && (139)\\
&\quad + \left|\left(\tau^{\star}_{\text{RR,AIPW, 0}}\right)^{-1}\left(\hat{\tau}_{\text{RR,AIPW, 1}} - \tau^{\star}_{\text{RR,AIPW, 1}}\right)\right|. && (140)
\end{aligned}$$

We now show that

$$\left|\hat{\tau}_{\text{RR,AIPW, 1}} - \tau^{\star}_{\text{RR,AIPW, 1}}\right| = o_p\left(\frac{1}{\sqrt{n}}\right) \quad \text{and} \quad \left|\hat{\tau}_{\text{RR,AIPW, 0}} - \tau^{\star}_{\text{RR,AIPW, 0}}\right| = o_p\left(\frac{1}{\sqrt{n}}\right).$$

The following decomposition holds

$$\begin{aligned}
\sqrt{n}\left|\hat{\tau}_{\text{RR,AIPW, 1}} - \tau^{\star}_{\text{RR,AIPW, 1}}\right| &= \frac{1}{\sqrt{n}}\sum_{i\in\mathcal{I}_k}\left(\hat{\mu}_1^{\mathcal{I}_{-k}}(X_i) + T_i\frac{Y_i - \hat{\mu}_1^{\mathcal{I}_{-k}}(X_i)}{\hat{e}(X_i)} - \mu_1(X_i) - T_i\frac{Y_i - \mu_1(X_i)}{e(X_i)}\right)\\
\text{Further denoted } A_n^k \quad &= \frac{1}{\sqrt{n}}\sum_{i\in\mathcal{I}_k}\left(\left(\hat{\mu}_1^{\mathcal{I}_{-k}}(X_i) - \mu_1(X_i)\right)\left(1 - \frac{T_i}{e(X_i)}\right)\right)\\
\text{Further denoted } B_n^k \quad &+ \frac{1}{\sqrt{n}}\sum_{i\in\mathcal{I}_k}T_i\left(\left(Y_i - \mu_1(X_i)\right)\left(\frac{1}{\hat{e}(X_i)} - \frac{1}{e(X_i)}\right)\right)\\
\text{Further denoted } C_n^k \quad &- \frac{1}{\sqrt{n}}\sum_{i\in\mathcal{I}_k}T_i\left(\left(\hat{\mu}_1^{\mathcal{I}_{-k}}(X_i) - \mu_1(X_i)\right)\left(\frac{1}{\hat{e}(X_i)} - \frac{1}{e(X_i)}\right)\right).
\end{aligned}$$

In the following, we prove that the first two terms tend to zero in $L^2$.

**Regarding** $A_n^k$ One can show that the expectation of $A_n^k/\sqrt{n}$ is null:

$$\begin{aligned}
\mathbb{E}\left[\frac{A_n^k}{\sqrt{n}} \mid \mathcal{I}_{-k}\right] &= \frac{1}{n}\sum_{i\in\mathcal{I}_k}\mathbb{E}\left[\left(\hat{\mu}_1^{\mathcal{I}_{-k}}(X_i)-\mu_1(X_i)\right)\left(1-\frac{T_i}{e(X_i)}\right)\mid \mathcal{I}_{-k}\right] \\
&= \frac{|\mathcal{I}_k|}{n}\mathbb{E}\left[\left(\hat{\mu}_1^{\mathcal{I}_{-k}}(X)-\mu_1(X)\right)\left(1-\frac{T}{e(X)}\right)\mid \mathcal{I}_{-k}\right] && \text{i.i.d.} \\
&= \frac{|\mathcal{I}_k|}{n}\mathbb{E}\left[\mathbb{E}\left[\left(\hat{\mu}_1^{\mathcal{I}_{-k}}(X)-\mu_1(X)\right)\left(1-\frac{T}{e(X)}\right)\mid X,\mathcal{I}_{-k}\right]\mid \mathcal{I}_{-k}\right] \\
&= \frac{|\mathcal{I}_k|}{n}\mathbb{E}\left[\left(\hat{\mu}_1^{\mathcal{I}_{-k}}(X)-\mu_1(X)\right)\mathbb{E}\left[\left(1-\frac{T}{e(X)}\right)\mid X,\mathcal{I}_{-k}\right]\mid \mathcal{I}_{-k}\right] \\
&= \frac{|\mathcal{I}_k|}{n}\mathbb{E}\left[\left(\hat{\mu}_1^{\mathcal{I}_{-k}}(X)-\mu_1(X)\right)\left(1-\frac{e(X)}{e(X)}\right)\mid \mathcal{I}_{-k}\right] \\
&= 0.
\end{aligned}$$

We will make use of this results in several calculations. Now,

$$\begin{aligned}
\mathbb{E}\left[\left(\frac{A_n^k}{\sqrt{n}}\right)^2 \mid \mathcal{I}_{-k}\right] &= \operatorname{Var}\left[\frac{1}{n}\sum_{i\in\mathcal{I}_k}\left(\left(\hat{\mu}_1^{\mathcal{I}_{-k}}(X_i)-\mu_1(X_i)\right)\left(1-\frac{T_i}{e(X_i)}\right)\right)\mid \mathcal{I}_{-k}\right] \\
&= \frac{1}{n^2}\operatorname{Var}\left[\sum_{i\in\mathcal{I}_k}\left(\left(\hat{\mu}_1^{\mathcal{I}_{-k}}(X_i)-\mu_1(X_i)\right)\left(1-\frac{T_i}{e(X_i)}\right)\right)\mid \mathcal{I}_{-k}\right] \\
&= \frac{1}{n^2}\sum_{i\in\mathcal{I}_k}\operatorname{Var}\left[\left(\hat{\mu}_1^{\mathcal{I}_{-k}}(X_i)-\mu_1(X_i)\right)\left(1-\frac{T_i}{e(X_i)}\right)\mid \mathcal{I}_{-k}\right] && \text{iid} \\
&= \frac{|\mathcal{I}_k|}{n^2}\mathbb{E}\left[\left(\left(\hat{\mu}_1^{\mathcal{I}_{-k}}(X)-\mu_1(X)\right)\left(1-\frac{T}{e(X)}\right)\right)^2\mid \mathcal{I}_{-k}\right] \\
&= \frac{|\mathcal{I}_k|}{n^2}\mathbb{E}\left[\mathbb{E}\left[\left(\left(\hat{\mu}_1^{\mathcal{I}_{-k}}(X)-\mu_1(X)\right)\left(1-\frac{T}{e(X)}\right)\right)^2|X,\mathcal{I}_{-k}\right]\mid \mathcal{I}_{-k}\right] \\
&= \frac{|\mathcal{I}_k|}{n^2}\mathbb{E}\left[\left(\hat{\mu}_1^{\mathcal{I}_{-k}}(X)-\mu_1(X)\right)^2\mathbb{E}\left[\left(1-\frac{T}{e(X)}\right)^2|X,\mathcal{I}_{-k}\right]\mid \mathcal{I}_{-k}\right] \\
&= \frac{|\mathcal{I}_k|}{n^2}\mathbb{E}\left[\left(\hat{\mu}_1^{\mathcal{I}_{-k}}(X)-\mu_1(X)\right)^2\frac{1}{e(X)^2}\mathbb{E}\left[\left(e(X)-T\right)^2|X,\mathcal{I}_{-k}\right]\mid \mathcal{I}_{-k}\right] \\
&= \frac{|\mathcal{I}_k|}{n^2}\mathbb{E}\left[\left(\hat{\mu}_1^{\mathcal{I}_{-k}}(X)-\mu_1(X)\right)^2\frac{e(X)\left(1-e(X)\right)}{e(X)^2}\mid \mathcal{I}_{-k}\right] \\
&= \frac{|\mathcal{I}_k|}{n^2}\mathbb{E}\left[\left(\hat{\mu}_1^{\mathcal{I}_{-k}}(X)-\mu_1(X)\right)^2\left(\frac{1}{e(X)}-1\right)\mid \mathcal{I}_{-k}\right] \\
&\leq \frac{|\mathcal{I}_k|}{\eta n^2}\mathbb{E}\left[\left(\hat{\mu}_1^{\mathcal{I}_{-k}}(X)-\mu_1(X)\right)^2\mid \mathcal{I}_{-k}\right] && \text{Overlap.}
\end{aligned}$$

Taking the expectation, we obtain

$$\mathbb{E}\left[\left(\frac{A_n^k}{\sqrt{n}}\right)^2\right] \leq \frac{|\mathcal{I}_k|}{\eta n^2}\mathbb{E}\left[\left(\hat{\mu}_1^{\mathcal{I}_{-k}}(X)-\mu_1(X)\right)^2\right], \tag{141}$$

that is

$$\mathbb{E}\left[\left(A_n^k\right)^2\right] \leq \frac{1}{\eta}\mathbb{E}\left[\left(\hat{\mu}_1^{\mathcal{I}_{-k}}(X)-\mu_1(X)\right)^2\right]. \tag{142}$$

Thus $A_n^k$ converges to zero in $L^2$ and thus in probability.

**Regarding** $B_n^k$ The second term $B_n^k$ can also be controlled using similar arguments. By assumption,

$$\frac{\eta}{2} \leq \hat{e}(X) \leq 1 - \frac{\eta}{2}.$$

Thus,

$$\frac{1}{\hat{e}(X)} - \frac{1}{e(X)} = \frac{e(X) - \hat{e}(X)}{\hat{e}(X)e(X)} \leq 2\left(\frac{e(X) - \hat{e}(X)}{\eta^2}\right).$$

Derivations are very close to the ones for the first term, noting that,

$$\mathbb{E}\left[\mathbb{E}\left[\frac{1}{n}\sum_{i\in\mathcal{I}_k} T_i\left(\left(Y_i - \mu_1\left(X_i\right)\right)\left(\frac{1}{\hat{e}^{\mathcal{I}_{-k}}\left(X_i\right)} - \frac{1}{e\left(X_i\right)}\right)\right) \mid X_i, \mathcal{I}_{-k}\right] \mid \mathcal{I}_{-k}\right] = 0,$$

so that,

$$\begin{aligned}
\mathbb{E}\left[\left(\frac{B_n^k}{\sqrt{n}}\right)^2 \mid \mathcal{I}_{-k}\right] &= \operatorname{Var}\left[\frac{1}{n}\sum_{i\in\mathcal{I}_k} T_i\left(Y_i - \mu_1\left(X_i\right)\right)\left(\frac{1}{\hat{e}^{\mathcal{I}_{-k}}\left(X_i\right)} - \frac{1}{e\left(X_i\right)}\right) \mid \mathcal{I}_{-k}\right] \\
&= \frac{1}{n^2}\sum_{i\in\mathcal{I}_k} \operatorname{Var}\left[T_i\left(Y_i - \mu_1\left(X_i\right)\right)\left(\frac{1}{\hat{e}^{\mathcal{I}_{-k}}\left(X_i\right)} - \frac{1}{e\left(X_i\right)}\right) \mid \mathcal{I}_{-k}\right] && \text{iid} \\
&= \frac{|\mathcal{I}_k|}{n^2}\mathbb{E}\left[T\left(Y - \mu_1\left(X\right)\right)^2\left(\frac{1}{\hat{e}^{\mathcal{I}_{-k}}\left(X\right)} - \frac{1}{e\left(X\right)}\right)^2 \mid \mathcal{I}_{-k}\right] \\
&\leq \frac{4|\mathcal{I}_k|}{\eta^4 n^2}\mathbb{E}\left[T\left(Y - \mu_1\left(X\right)\right)^2\left(\hat{e}^{\mathcal{I}_{-k}}\left(X\right) - e\left(X\right)\right)^2 \mid \mathcal{I}_{-k}\right] \\
&\leq \frac{4|\mathcal{I}_k|}{\eta^4 n^2}\mathbb{E}\left[\left(Y - \mu_1\left(X\right)\right)^2\left(\hat{e}^{\mathcal{I}_{-k}}\left(X\right) - e\left(X\right)\right)^2 \mid \mathcal{I}_{-k}\right] && \text{Sicne } T \leq 1 \\
&\leq \frac{4|\mathcal{I}_k|}{\eta^4 n^2}\mathbb{E}\left[\mathbb{E}\left[\left(Y - \mu_1\left(X\right)\right)^2\left(\hat{e}^{\mathcal{I}_{-k}}\left(X\right) - e\left(X\right)\right)^2 | X, \mathcal{I}_{-k}\right] \mid \mathcal{I}_{-k}\right] \\
&\leq \frac{4|\mathcal{I}_k|}{\eta^4 n^2}\mathbb{E}\left[\mathbb{E}\left[\left(Y - \mu_1\left(X\right)\right)^2 | X, \mathcal{I}_{-k}\right]\left(\hat{e}^{\mathcal{I}_{-k}}\left(X\right) - e\left(X\right)\right)^2 \mid \mathcal{I}_{-k}\right] \\
&\leq \frac{4|\mathcal{I}_k|}{\eta^4 n^2}\mathbb{E}\left[\operatorname{Var}\left[Y|X\right]\left(\hat{e}^{\mathcal{I}_{-k}}\left(X\right) - e\left(X\right)\right)^2 \mid \mathcal{I}_{-k}\right] \\
&\leq \frac{4\operatorname{Var}\left[Y|X\right]|\mathcal{I}_k|}{\eta^4 n^2}\mathbb{E}\left[\left(\hat{e}^{\mathcal{I}_{-k}}\left(X\right) - e\left(X\right)\right)^2 \mid \mathcal{I}_{-k}\right].
\end{aligned}$$

Taking the expectation on both sides, since $\operatorname{Var}\left[Y|X\right] \leq \sigma^2$, we get

$$\mathbb{E}\left[\left(\frac{B_n^k}{\sqrt{n}}\right)^2\right] \leq \frac{4\sigma^2|\mathcal{I}_k|}{\eta^4 n^2}\mathbb{E}\left[\left(\hat{e}^{\mathcal{I}_{-k}}\left(X\right) - e\left(X\right)\right)^2 \mid \mathcal{I}_{-k}\right], \tag{143}$$

which leads to

$$\mathbb{E}\left[\left(B_n^k\right)^2\right] \leq \frac{4\sigma^2}{\eta^4}\mathbb{E}\left[\left(\hat{e}^{\mathcal{I}_{-k}}\left(X\right) - e\left(X\right)\right)^2 \mid \mathcal{I}_{-k}\right]. \tag{144}$$

Since, by assumption, the right-hand side term converges to zero, $B_n^k$ converges to zero in $L^2$.

**Regarding** $C_n^k$ Regarding the last term, the approach is different and will involve another assumption on the product of residuals. More precisely,

$$
\begin{aligned}
\mathbb{E}[|C_n^k|] &= \sqrt{n}\frac{1}{n}\sum_{i\in\mathcal{I}_k}\mathbb{E}\left[\left|T_i\left(\hat{\mu}_1^{\mathcal{I}_{-k}}(X_i)-\mu_1(X_i)\right)\left(\frac{1}{\hat{e}^{\mathcal{I}_{-k}}(X_i)}-\frac{1}{e(X_i)}\right)\right|\right] \\
&= \frac{\sqrt{n}}{\eta^2}\frac{1}{n}\sum_{i\in\mathcal{I}_k}\mathbb{E}\left[\left|T_i\left(\hat{\mu}_1^{\mathcal{I}_{-k}}(X_i)-\mu_1(X_i)\right)\left(e(X_i)-\hat{e}^{\mathcal{I}_{-k}}(X_i)\right)\right|\right] \\
&= \frac{\sqrt{n}|\mathcal{I}_k|}{\eta^2}\frac{1}{n}\mathbb{E}\left[\left|T\left(\hat{\mu}_1^{\mathcal{I}_{-k}}(X)-\mu_1(X)\right)\left(e(X)-\hat{e}^{\mathcal{I}_{-k}}(X)\right)\right|\right] \\
&\leq \frac{\sqrt{n}}{\eta^2}\mathbb{E}\left[\left|\left(\hat{\mu}_1^{\mathcal{I}_{-k}}(X)-\mu_1(X)\right)\left(e(X)-\hat{e}^{\mathcal{I}_{-k}}(X)\right)\right|\right] \\
&\leq \frac{\sqrt{n}}{\eta^2}\sqrt{\mathbb{E}\left[\left(\hat{\mu}_1^{\mathcal{I}_{-k}}(X)-\mu_1(X)\right)^2\right]\mathbb{E}\left[\left(e(X)-\hat{e}^{\mathcal{I}_{-k}}(X)\right)^2\right]},
\end{aligned}
$$

which tends to zero by assumption. Each term $A_n^k$, $B_n^k$, and $C_n^k$ has been shown to be bounded by a term in $o_{\mathbb{P}}(1)$. Thus,

$$
\sqrt{n}\left|\hat{\tau}_{\text{RR,AIPW, 1}}-\tau^\star_{\text{RR,AIPW, 1}}\right| = \sum_{k=1}^K A_n^k + B_n^k + C_n^k \tag{145}
$$

tends to zero in probability. Similarly, one can show that

$$
\sqrt{n}\left|\hat{\tau}_{\text{RR,AIPW, 1}}-\tau^\star_{\text{RR,AIPW, 1}}\right| \xrightarrow{p} 0. \tag{146}
$$

According to (140), since for all $t\in\{0,1\}$, $|\hat{\tau}_{\text{RR,AIPW,t}}|$ tends to $\tau^\star_{\text{RR,AIPW,t}}$ which is lower and upper bounded, we have

$$
\begin{aligned}
\sqrt{n}\left|\hat{\tau}_{\text{RR,AIPW}}-\tau^\star_{\text{RR,AIPW}}\right| \leq{}& \sqrt{n}\left|\frac{\hat{\tau}_{\text{RR,AIPW, 1}}}{\hat{\tau}_{\text{RR,AIPW, 0}}\tau^\star_{\text{RR,AIPW, 0}}}\right|\left|\hat{\tau}_{\text{RR,AIPW, 0}}-\tau^\star_{\text{RR,AIPW, 0}}\right| \\
&+\sqrt{n}\left|\frac{1}{\tau^\star_{\text{RR,AIPW, 0}}}\right|\left|\hat{\tau}_{\text{RR,AIPW, 1}}-\tau^\star_{\text{RR,AIPW, 1}}\right|
\end{aligned}
$$

which tends to zero.

Regarding the second term in (134), we can use Theorem 1 with $g_1(Z)=\mu_1(X)+\frac{T(Y-\mu_1(X))}{e(X)}$ and $g_0(Z)=\mu_0(X)+\frac{(1-T)(Y-\mu_0(X))}{(1-e(X))}$ where $Z=(T,X,Y)$. Hence, we have that $g_1(Z)$ is square integrable:

$$
\mathbb{E}\left[g_1(Z)^2\right]\leq 2\mathbb{E}\left[(\mu_1(X)^2\right]+2\mathbb{E}\left[\left(\frac{T(Y-\mu_1(X))}{e(X)}\right)^2\right],
$$

where $\mathbb{E}\left[(\mu_1(X)^2\right]=\text{Var}(Y^{(1)})+\mathbb{E}\left[Y^{(1)}\right]^2$ is finite. Using Consistency, Unconfoundedness, and definition or $\mu_1(X)=\mathbb{E}[Y\mid X,T=1]$, simple calculations show that

$$
\begin{aligned}
\mathbb{E}\left[\left(T\frac{Y-\mu_1(X)}{e(X)}\right)^2\right] &= \mathbb{E}\left[\left(T\frac{Y^{(1)}-\mu_1(X)}{e(X)}\right)^2\right] && \text{Consistency} \\
&= \mathbb{E}\left[\mathbb{E}\left[\left(T\frac{Y^{(1)}-\mu_1(X)}{e(X)}\right)^2\mid X\right]\right] && \text{Total expectation} \\
&= \mathbb{E}\left[\frac{\left(Y^{(1)}-\mu_1(X)\right)^2}{e(X)}\right] \\
&\leq \frac{\text{Var}(\mu_1(X))}{\eta}.
\end{aligned}
$$

Similarly, we can show that $g_0(Z)$ is square integrable. Since $\mathbb{E}\left[g_0(Z)\right] = \mathbb{E}\left[Y^{(0)}\right]$ and $\mathbb{E}\left[g_1(Z)\right] = \mathbb{E}\left[Y^{(1)}\right]$, we can apply Theorem 1 and conclude that

$$\sqrt{n}(\tau^{\star}_{\mathrm{RR,AIPW}} - \tau_{\mathrm{RR}}) \rightarrow \mathcal{N}(0, V_{\mathrm{RR,OS}}). \tag{147}$$

Finally,

$$\sqrt{n}(\hat{\tau}_{\mathrm{AIPW}} - \tau_{\mathrm{RR}}) = \underbrace{\sqrt{n}(\hat{\tau}_{\mathrm{RR,AIPW}} - \tau^{\star}_{\mathrm{RR,AIPW}})}_{\xrightarrow{p} 0} + \underbrace{\sqrt{n}(\tau^{\star}_{\mathrm{RR,AIPW}} - \tau_{\mathrm{RR}})}_{\xrightarrow{d} \mathcal{N}\left(0, V_{\mathrm{RR,OS}}\right)},$$

where

$$V_{\mathrm{RR,OS}} = \tau_{\mathrm{RR}}^2 \left( \mathrm{Var}\left( \frac{\mu_1(X)}{\mathbb{E}\left[Y^{(1)}\right]} - \frac{\mu_0(X)}{\mathbb{E}\left[Y^{(0)}\right]} \right) + \mathbb{E}\left[ \frac{\mathrm{Var}\left(Y^{(1)}|X\right)}{e(X)\mathbb{E}\left[Y^{(1)}\right]^2} \right] + \mathbb{E}\left[ \frac{\mathrm{Var}\left(Y^{(0)}|X\right)}{(1-e(X))\mathbb{E}\left[Y^{(0)}\right]^2} \right] \right)$$

□

# 7 Simulation

For the simulations we have implemented all estimators in Python using Scikit-Learn for our regression and classification models. All our experiments were run on a 8GB M1 Mac. The propensity scores is estimated based on the provided training data and covariate names. Depending on the chosen method, it either uses logistic regression with a high regularization parameter (parametric) or a random forest classifier with parameters determined by the training data size (non-parametric). The response surface is estimated based on the training data, covariate names, the method (parametric or non-parametric), and whether the response is binary or continuous. For parametric methods, it uses a stochastic gradient descent classifier for binary responses and a linear regression model for continuous responses. For non-parametric methods, it employs a random forest classifier for binary responses and a random forest regressor for continuous responses. Both methods fit the model using the training data to estimate the respective scores and surfaces, enabling flexible handling of various datasets and assumptions for causal inference analysis.

## 7.1 Randomized Controlled Trials

In this part we will simulate Randomized Controlled Trials (RCT) and test the following Ratio estimators: Ratio Neyman, Ratio Horvitz Thomson and the Ratio G-formula. Since we are in a Randomized Controlled Trials, the propensity score $e(.)$ is constant.

### 7.1.1 Linear RCT

The first DGP has linear outcome models (linear treatment effect and the baseline). The data is generated using:

$$\begin{array}{llll} m(X) = (c_1 - c_0) + (\beta_1 - \beta_0)^\top X & c_0 & = 6, & c_1 = 12 \\ b(X) = c_0 + \beta_0^\top X & \beta_1 & = (2, -5, 2, 8, -2, 8) & \\ e(X) = 0.5 & \beta_0 & = (3, -7, 1, 4, -2, 2) & \end{array}$$

Given that $X$ has a zero mean, it follows that $\tau_{\mathrm{RR}} = c_1/c_0 = 2$. This scenario aligns with the linear setting outlined in Assumption 4. Referring to Figure 4, as proved in the previous sections all estimators converge to the true Risk Ratio as $n$ increases. Additionally, within this linear framework as per Lemma 2, the variance of the Neyman estimator exceeds the one of the G-formula. In such a linear environment, the parametric G-formula performs better than its non-parametric counterpart. Additionally, the Ratio Neyman estimator demonstrates lower variance compared to the Horvitz-Thomson estimator as indicated in Equation (5).

### 7.1.2 Non-Linear RCT

This DGP is also a Randomized Controlled Trials however, the outcomes are not linear this time:

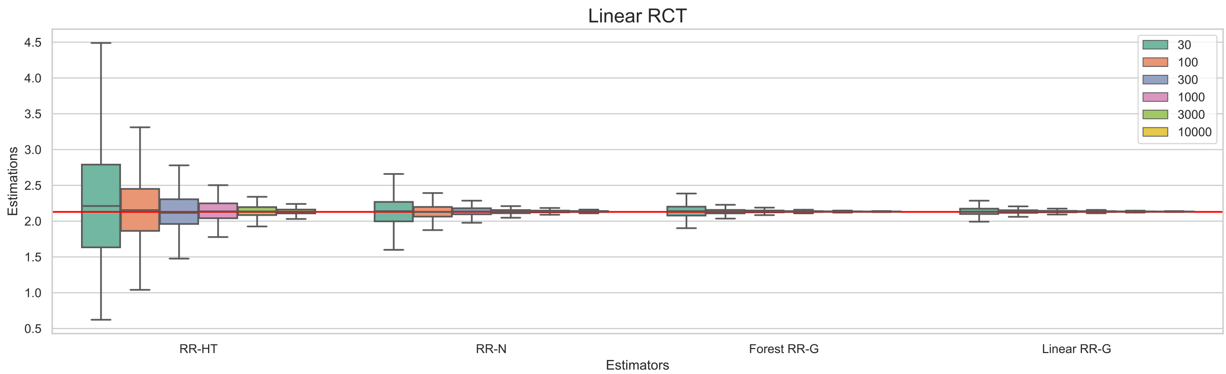


Figure 4: Comparison of RCT estimators in a Linear RCT

$$m(X) = \sin(X_1) \cdot X_2^2 + \frac{X_3}{X_4 + 1} - \log(X_5 + 1) + X_6^3 + 1$$

$$b(X) = 4 * \max(X_1 + X_2 + X_3, 0) - \min(X_4 + X_6, 0) \quad \text{and} \quad e(X) = 0.5$$

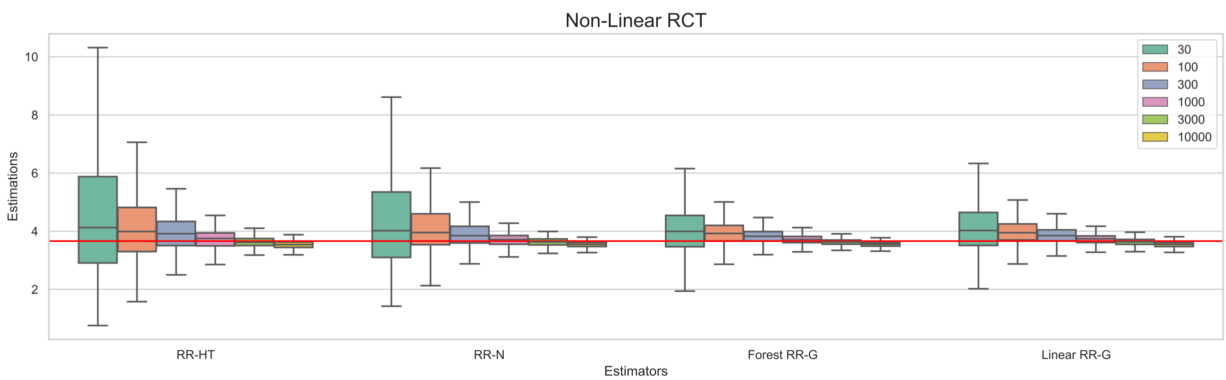


Figure 5: Comparison of RCT estimators in a Non-Linear RCT

The presence of trigonometric, exponential, logarithmic, and polynomial terms makes this setting non-linear. It's important to note that since we are in a Randomized Controlled Trial (RCT), the propensity function remains constant. As the sample size $(n)$ increases, all proposed estimators converge. A bias can be seen in 5 but decreases to 0 as $(n)$ increases as predicted in previous sections. Linear regression struggles with small $n$ values, failing to capture the intricate relationships between features and non-linearities. On the other hand, Random Forest, a non-parametric method, excels in capturing these complexities by segmenting the feature space and predicting based on response averages within those segments. However, predicting the complex function can be challenging, the Neyman estimator might outperform the G-formula, particularly when both parametric and non-parametric responses may lack consistency. Although we do not fall in assumptions of Equation (5) the Ratio Neyman estimator demonstrates lower variance compared to the Horvitz-Thomson estimator.

### 7.2 Observational Studies

#### 7.2.1 Non-Linear and Logistic DGP

Using the Non-Linear and Logistic DGP, Figure 6 represents the coverage and length of the 95% CI obtained for the AIPW, G-formula, and Neyman estimators. Only the linear RR-AIPW estimator has a satisfying coverage, which is expected given the non-linear logistic DGP. In this context, the propensity score is likely to be well-estimated, while estimating non-linear response surfaces using Linear or Forest methods, such as in Linear/Forest RR-G and

Forest RR-AIPW, proves challenging with a sample size of only $n = 1000$. As anticipated, the RR-N estimator shows almost no coverage, consistent with its application in observational studies.

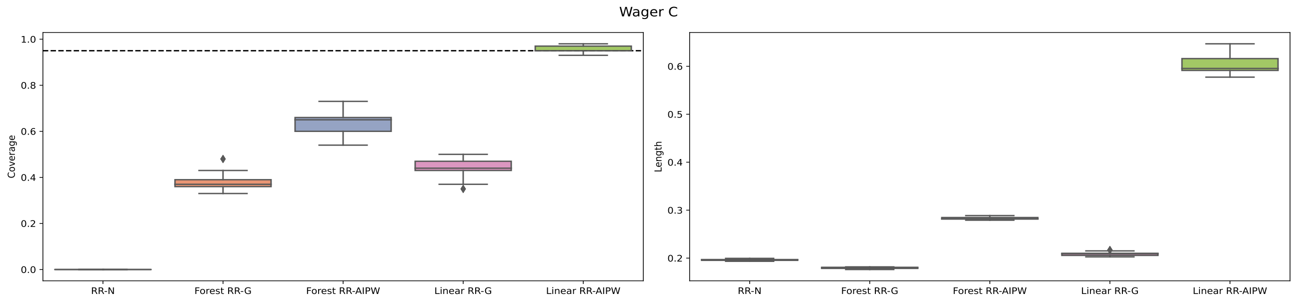


Figure 6: Average coverage (left) and average length (right) of asymptotic confidence interval derived from Section 2 and Section 3 for different estimators with $n = 1000$ and 300 repetitions for a Non-Linear and Logistic DGP.

### 7.2.2 Non-Linear and Non-Logistic DGP

We use the same simulations as in Nie and Wager (2020) using nonlinear models for every quantity, as detailed below, with $X \sim \text{Unif}(0, 1)^6$

$$\begin{aligned} m(X) &= \sin\left(\pi X_1 X_2\right) + 2\left(X_3 - 0.5\right)^2 + X_4 + 0.5X_5 - \left(X_1 + X_2\right)/4 \\ b(X) &= \left(X_1 + X_2\right)/2 \\ e(X) &= \max\{0.1, \min(\sin\left(\pi X_1\right), 0.9)\}. \end{aligned}$$

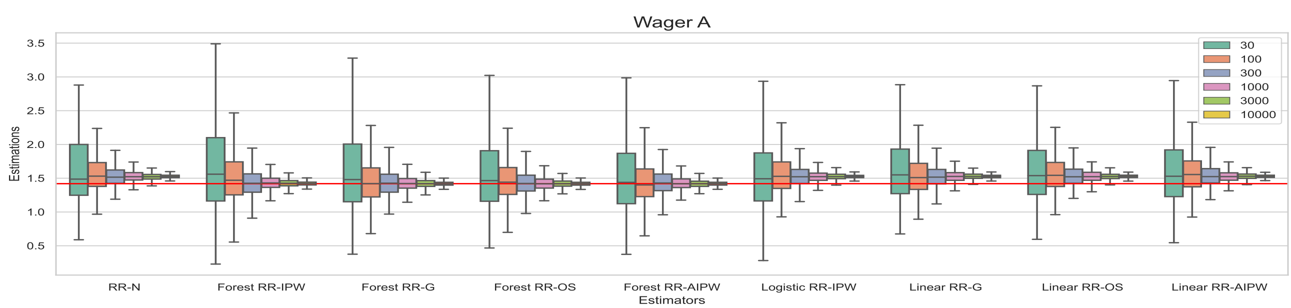


Figure 7: Estimations of the Risk Ratio with weighting, outcome based and augmented estimators as a function of the sample size for the non-Linear-non-Logistic DGP. Parametric (Regression) and non parametric (Forest) estimations of nuisance are displayed.

Results are presented in Figure 7. At first glance, all methods seem to have similar performances. However, estimators based on parametric estimators (last four) fail to converge to the correct quantity. They present an intrinsic bias, which does not vanish as the sample size increases. This was expected as linear methods are unable to model the complex non-linear generative process of this simulation. On the other hand, methods that employ random forests estimators achieve good performances: they are consistent and unbias even for small sample sizes. Note that RIPW has a larger variance than the other methods, with a small bias for very small sample sizes. Therefore, the G-formula and the two doubly-robust estimators that use random forests are competitive in this setting. Here again, both double robust estimators give similar performances. No estimator achieves 95% coverage, which is expected given the non-linear, non-logistic DGP. Linear estimators, such as Linear RR-G and RR-AIPW, struggle to accurately estimate the nuisance functions in this context. Additionally, the limited number of observations prevents the Forest estimators from converging effectively.

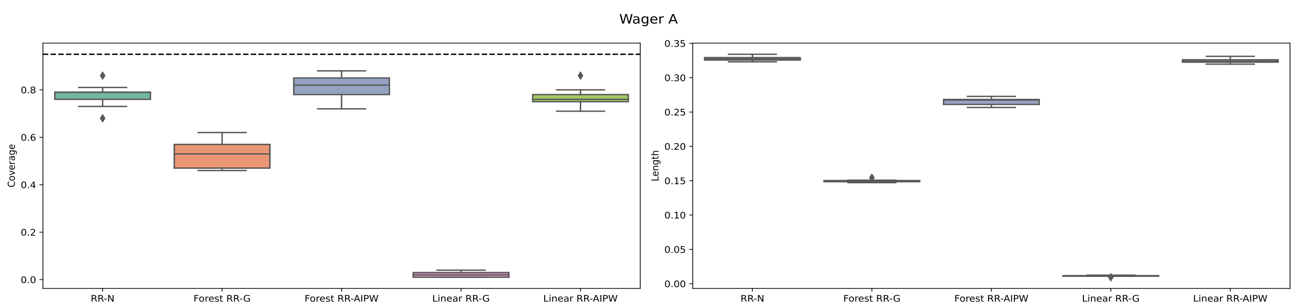


Figure 8: Average coverage (left) and average length (right) of asymptotic confidence interval derived from Section 2 and Section 3 for different estimators with $n = 1000$ and 300 repetitions for a Non-Linear and Non-Logistic DGP.