

Empirical likelihood inference for non-randomized pretest-posttest studies with missing data

Shixiao Zhang

*Department of Statistics and Actuarial Science, University of Waterloo,
200 University Avenue West, Waterloo, Ontario N2L 3G1, Canada
e-mail: shixiao.zhang@uwaterloo.ca*

Peisong Han

*Department of Biostatistics, School of Public Health,
University of Michigan, 1415 Washington Heights, Ann Arbor, Michigan 48109-2029,
U.S.A.
e-mail: peisong@umich.edu*

and

Changbao Wu

*Department of Statistics and Actuarial Science, University of Waterloo,
200 University Avenue West, Waterloo, Ontario N2L 3G1, Canada
e-mail: cbwu@uwaterloo.ca*

Abstract: Pretest-posttest studies are commonly used for assessing the effect of a treatment or an intervention. We propose an empirical likelihood based approach to both testing and estimation of the treatment effect in non-randomized pretest-posttest studies where the posttest outcomes are subject to missingness. The proposed empirical likelihood ratio test and the estimation procedure are multiply robust in the sense that multiple working models are allowed for the propensity score of treatment assignment, the missingness probability and the outcome regression, and the validity of the test and the estimation requires only a certain combination of those multiple working models to be correctly specified. An empirical likelihood ratio confidence interval can be constructed for the treatment effect and has better coverage probabilities than confidence intervals based on the Wald statistic. Simulations are conducted to demonstrate the finite-sample performances of the proposed methods.

Keywords and phrases: Auxiliary information, biased Sampling, empirical likelihood, missing at random, multiple robustness, observational study, treatment effect.

Received August 2018.

Contents

1	Introduction	2013
2	Notation and setup	2014
3	Empirical likelihood ratio test for the treatment effect	2016

3.1	Known propensity scores	2016
3.2	Unknown propensity scores	2020
3.3	A bootstrap calibration	2023
4	Estimation of the treatment effect	2023
5	Simulation study	2025
6	Additional remarks	2031
7	Appendix	2031
7.1	Expressions for the Scaling Constants in the Theorems	2031
7.1.1	Expressions for Theorem 3.1	2031
7.1.2	Expressions for Theorem 3.2	2032
7.1.3	Expressions for Theorem 3.3	2033
7.1.4	Expressions for Theorem 3.4	2033
7.2	Proofs of the Theorems	2034
7.2.1	Proof of Theorem 3.1	2034
7.2.2	Proof of Theorem 3.2	2035
7.2.3	Proof of Theorem 4.1	2037
7.2.4	Proof of Theorem 4.2	2040
	References	2041

1. Introduction

Pretest-posttest studies are commonly used to evaluate the effect of a treatment or an intervention. In randomized pretest-posttest studies, subjects are randomly assigned to one of the treatment group or the control group. The outcome of interest is first measured at the baseline prior to the treatment (pretest) along with certain auxiliary variables and then again at the end of the study after the treatment (posttest). The main parameter of interest is the treatment effect defined as the difference in change of mean responses between the treatment and the control groups. Inferences on the treatment effect for randomized pretest-posttest studies can be done by classic statistical methods such as the two-sample t -test, the paired t -test, analysis of covariance, or the generalized estimating equations. See, for instance, Brogan and Kutner [2], Laird [16], Stanek [25], Follmann [9], Singer and Andrade [24], Yang and Tsiatis [28], Bonate [1], among others. By adopting the framework of potential outcomes and treating the unobservable counterfactual outcomes as missing data, Leon et al. [17] constructed semiparametric estimators of the treatment effect based on the missing data theory of Robins et al. [21]. Chen et al. [5, 6] considered an imputation-based approach under the same framework.

The posttest outcomes are often subject to missingness. Under the assumption of missing at random as defined by Rubin [23], Davidian et al. [8] studied semiparametric estimation of the treatment effect and constructed a class of augmented inverse probability weighted estimators (Robins et al. 21) by deriving the influence functions for all regular and asymptotically linear estimators under the setting they considered. The augmented inverse probability weighted estimators require working models for both the missingness probability and the

outcome regression. Huang et al. [15] proposed an empirical likelihood method for randomized pretest-posttest studies with missing data.

The randomization step in pretest-posttest research designs is often not feasible in practice. In many social studies on the effectiveness of an intervention, it is mandatory that participants are allowed to choose to take the intervention or not. This is also the case for many medical studies involving two alternative treatments. For such self-selection-based treatment assignments, inference procedures developed for randomized designs are no longer suitable.

This paper presents an empirical likelihood approach to non-randomized pretest-posttest studies when the posttest outcomes are also subject to missingness. We develop a unified framework for both testing and estimation of the treatment effect and the inferential procedures are multiply robust in the sense that multiple working models are allowed for the propensity score of treatment assignment, the missingness probability and the outcome regression, and the validity of testing and estimation requires only a certain combination of those multiple working models to be correctly specified. Multiply robust inferences were first introduced by Han and Wang [14] for missing data problems. The methods have attracted considerable research interests due to added layers of protection against possible misspecifications of working models. See, for instance, Chan and Yam [3], Han [11, 10, 12, 13], Chen and Haziza [7], among others. Multiply robust methods usually lead to smaller biases under complete misspecification of all working models (Han 11, 12). Our general framework follows the two-sample empirical likelihood formulation with estimating equations (Qin and Lawless 19, Owen 18, Qin and Zhang 20, Huang et al. 15, Wu and Yan 27). However, the use of multiple working models for the propensity score of treatment assignment, the missingness probability and the outcome regression makes our proposed approach much more challenging in terms of theoretical development.

In addition to multiple robustness, the proposed empirical likelihood ratio test leads to data-adaptive confidence intervals for the treatment effect that give better finite-sample coverage probabilities compared to those based on the Wald statistic, and the proposed estimator of treatment effect always falls into the corresponding parameter space by construction.

2. Notation and setup

The setup for this paper is a generalization of that used by Davidian et al. [8] and Huang et al. [15] from randomized trials to non-randomized trials. Let T denote the treatment indicator with $T = 1$ if the subject chooses treatment and $T = 0$ if the subject selects control. Let Y_1 and Y_0 denote, respectively, the posttest potential outcomes that would have been observed had a subject chosen treatment ($T = 1$) and control ($T = 0$). Let $Y = TY_1 + (1 - T)Y_0$ be the actual observed posttest outcome. Let \mathbf{Z} denote a vector of variables, including the pretest outcome as well as auxiliary variables, collected at the baseline before the treatment or intervention. After the treatment assignment

but prior to the end of the study, some additional variables \mathbf{X}_t for $T = t$, $t = 0, 1$, are also measured during the follow-up, including possible intermediate outcome measures. To accommodate possible missingness on Y , let R_t denote the indicator variable of observing Y for subjects in group $T = t$, $t = 0, 1$, with $R_t = 1$ if Y is observed and $R_t = 0$ otherwise. For a random sample of size n at the baseline, let $n_1 = \sum_{i=1}^n T_i$ and $n_0 = n - n_1$ be the numbers of subjects in the treatment group and in the control group, respectively. In addition, let $n_{11} = \sum_{i:T_i=1} R_{1i}$ and $n_{01} = \sum_{i:T_i=0} R_{0i}$ be the numbers of subjects in the treatment group and in the control group with Y observed, respectively. The data structure and the notation for all associated variables are shown in Table 1, which is similar to Table 1 of Huang et al. [15], after suitable re-ordering of subjects within each group. The parameter of interest is the treatment effect defined as $\delta = E(Y_1) - E(Y_0)$. We make the following standard assumptions on the treatment assignment and the missing data mechanism.

TABLE 1
Data Structure for Pretest-Posttest Studies With Missing Responses

Subject i	Baseline	Group T	Intermediate Variables		Missingness Indicators		Posttest Outcomes	
	\mathbf{Z}		\mathbf{X}_1	\mathbf{X}_0	R_1	R_0	Y_1	Y_0
1	\mathbf{z}_1	1	\mathbf{x}_{11}	??	1	??	y_{11}	??
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
n_{11}	$\mathbf{z}_{n_{11}}$	1	$\mathbf{x}_{1n_{11}}$??	1	??	$y_{1n_{11}}$??
$n_{11} + 1$	$\mathbf{z}_{n_{11}+1}$	1	$\mathbf{x}_{1(n_{11}+1)}$??	0	??	?	??
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
n_1	\mathbf{z}_{n_1}	1	\mathbf{x}_{1n_1}	??	0	??	?	??
$n_1 + 1$	\mathbf{z}_{n_1+1}	0	??	$\mathbf{x}_{0(n_1+1)}$??	1	??	$y_{0(n_1+1)}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
$n_1 + n_{01}$	$\mathbf{z}_{n_1+n_{01}}$	0	??	$\mathbf{x}_{0(n_1+n_{01})}$??	1	??	$y_{0(n_1+n_{01})}$
$n_1 + n_{01} + 1$	$\mathbf{z}_{n_1+n_{01}+1}$	0	??	$\mathbf{x}_{0(n_1+n_{01}+1)}$??	0	??	?
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
n	\mathbf{z}_n	0	??	\mathbf{x}_{0n}	??	0	??	?

?: counterfactual values not observed in the actual world. ?: missing values.

Assumption 1 (No unmeasured confounders). The treatment assignment is independent of the potential outcomes and the intermediate measurements given all baseline variables, i.e., $T \perp (Y_1, Y_0, \mathbf{X}_1, \mathbf{X}_0, R_1, R_0) \mid \mathbf{Z}$.

Assumption 2 (Missing at random). Missingness of the posttest outcome is independent of the outcome itself given all the baseline and intermediate measurements and the treatment assignment, i.e., $R_t \perp Y_t \mid (\mathbf{Z}, T = t, \mathbf{X}_t)$, $t = 0, 1$.

Assumption 3 (Positivity). The propensity score for treatment assignment and the missingness probability satisfy (i) $0 < \mathbb{P}(T = 1 \mid \mathbf{Z}) < 1$ and (ii) $0 < \mathbb{P}(R_t = 1 \mid \mathbf{Z}, T = t, \mathbf{X}_t) < 1$, $t = 0, 1$.

Let $\pi(\mathbf{Z}) = \mathbb{P}(T = 1 \mid \mathbf{Z})$ be the propensity score of treatment assignment and $\varpi_t(\mathbf{Z}, \mathbf{X}_t) = \mathbb{P}(R_t = 1 \mid \mathbf{Z}, T = t, \mathbf{X}_t)$ be the non-missingness probability for a subject in the group with $T = t$, $t = 0, 1$. The combination of Assumption 1 and Assumption 3(i) is referred to as strongly ignorable treatment assignment by Rosenbaum and Rubin [22]. Assumption 3(ii) implies that there is a positive probability of observing the complete data for each of the treatment and control groups.

3. Empirical likelihood ratio test for the treatment effect

3.1. Known propensity scores

We propose empirical likelihood ratio tests for $H_0 : \delta = \delta_*$ against $H_1 : \delta \neq \delta_*$, where δ_* is a pre-specified value. Testing the existence of a treatment effect then corresponds to the special case with $\delta_* = 0$. For ease of presentation and to facilitate asymptotic development, we first consider the less practical scenario in Section 3.1 where the propensity score $\pi(\mathbf{Z})$ is assumed to be known. The scenario with unknown propensity score is considered in Section 3.2.

(i) *Single working model for the missingness probability.* Suppose that the missingness probability $\varpi_t(\mathbf{Z}, \mathbf{X}_t)$ is correctly modeled by $\varpi_t(\mathbf{Z}, \mathbf{X}_t; \boldsymbol{\alpha}_t)$ such that $\varpi_t(\mathbf{Z}, \mathbf{X}_t; \boldsymbol{\alpha}_{t*}) = \varpi_t(\mathbf{Z}, \mathbf{X}_t)$ for some $\boldsymbol{\alpha}_{t*}$, $t = 0, 1$. Let $\hat{\boldsymbol{\alpha}}_t$ be the estimator of $\boldsymbol{\alpha}_{t*}$ derived by maximizing the likelihood function

$$\prod_{i:T_i=t} \{\varpi_{ti}(\boldsymbol{\alpha}_t)\}^{R_{ti}} \{1 - \varpi_{ti}(\boldsymbol{\alpha}_t)\}^{1-R_{ti}}, \quad (3.1)$$

where for notational simplicity we used $\varpi_{ti}(\boldsymbol{\alpha}_t)$ to denote $\varpi_t(\mathbf{Z}_i, \mathbf{X}_{ti}; \boldsymbol{\alpha}_t)$. It is easy to verify that the inverse probability weighted estimator

$$\frac{1}{n} \sum_{i:T_i=1, R_{1i}=1} \frac{Y_{1i}}{\pi(\mathbf{Z}_i)\varpi_{1i}(\hat{\boldsymbol{\alpha}}_1)} - \frac{1}{n} \sum_{i:T_i=0, R_{0i}=1} \frac{Y_{0i}}{\{1 - \pi(\mathbf{Z}_i)\}\varpi_{0i}(\hat{\boldsymbol{\alpha}}_0)}$$

is consistent for δ . This motivates the use of constraint (3.4) in the discussions below for our proposed empirical likelihood method.

Let $\{w_i : T_i = 1, R_{1i} = 1\}$ be a discrete probability measure over the set of subjects $\{i : T_i = 1, R_{1i} = 1\}$ with observed posttest outcomes; let $\{v_i : T_i = 0, R_{0i} = 1\}$ be a discrete probability measure over the set of subjects $\{i : T_i = 0, R_{0i} = 1\}$. The empirical likelihood function for the combined sample is defined as

$$L(\mathbf{w}, \mathbf{v}) = \prod_{i:T_i=1, R_{1i}=1} w_i \prod_{i:T_i=0, R_{0i}=1} v_i. \quad (3.2)$$

Maximizing $L(\mathbf{w}, \mathbf{v})$ with respect to w_i and v_i under the normalization constraints

$$w_i > 0, \quad \sum_{i:T_i=1, R_{1i}=1} w_i = 1, \quad v_i > 0, \quad \sum_{i:T_i=0, R_{0i}=1} v_i = 1 \quad (3.3)$$

leads to $\hat{w}_i = 1/n_{11}$ and $\hat{v}_i = 1/n_{01}$. The constraint induced by the parameter of interest, the treatment effect δ , is constructed as

$$\frac{n_{11}}{n} \sum_{i:T_i=1, R_{1i}=1} w_i \frac{Y_{1i}}{\pi(\mathbf{Z}_i) \varpi_{1i}(\hat{\boldsymbol{\alpha}}_1)} - \frac{n_{01}}{n} \sum_{i:T_i=0, R_{0i}=1} v_i \frac{Y_{0i}}{\{1 - \pi(\mathbf{Z}_i)\} \varpi_{0i}(\hat{\boldsymbol{\alpha}}_0)} = \delta_* \quad (3.4)$$

for the given δ_* .

The formulation described above follows the general framework of empirical likelihood with estimating equations (Qin and Lawless 19). It provides a powerful platform for incorporating additional constraints induced by working models for the treatment assignment, the missingness probability and the outcome regression to construct multiply robust test and estimation procedures.

Let \tilde{w}_i and \tilde{v}_i be the maximizer of $L(\mathbf{w}, \mathbf{v})$ under both the normalization constraint (3.3) and the parameter constraint (3.4). The empirical likelihood ratio statistic for testing $H_0 : \delta = \delta_*$ is computed as

$$W(\delta_*) \equiv -2 \left(\sum_{i:T_i=1, R_{1i}=1} \log \frac{\tilde{w}_i}{\hat{w}_i} + \sum_{i:T_i=0, R_{0i}=1} \log \frac{\tilde{v}_i}{\hat{v}_i} \right), \quad (3.5)$$

where the dependence of $W(\delta_*)$ on δ_* is through the dependence of \tilde{w}_i and \tilde{v}_i on δ_* . Let χ_1^2 denote the chi-square distribution with one degree of freedom. The following result gives the asymptotic distribution of $W(\delta_*)$ under H_0 . Proof of the theorem and the exact expression for the positive scaling factor σ_1 are provided in the Appendix.

Theorem 3.1. *Under Assumptions 1–3 with known $\pi(\mathbf{Z})$ and correctly specified $\varpi_t(\boldsymbol{\alpha}_t)$ for $\varpi_t(\mathbf{Z}, \mathbf{X}_t)$, $t = 0, 1$, the empirical likelihood ratio statistic $W(\delta_*)$ has an asymptotic distribution $\sigma_1 \chi_1^2$ under $H_0 : \delta = \delta_*$.*

A $(1 - \alpha)$ -level empirical likelihood ratio confidence interval for δ can be constructed as $\{\delta : W(\delta) \leq \hat{\sigma}_1 \chi_{1, (1-\alpha)}^2\}$, where $\hat{\sigma}_1$ is the estimated value of σ_1 and $\chi_{1, (1-\alpha)}^2$ is the 100(1- α)% percentile of the chi-square distribution with one degree of freedom. One advantage of the empirical likelihood ratio confidence interval compared to the Wald confidence interval is that it is data-driven and range respecting and has better coverage probability for δ with moderate sample sizes, which are shown in our simulation studies.

The major computational task for calculating $W(\delta_*)$ is to maximize (3.2) subject to (3.3) and (3.4) with the given δ_* . To derive the \tilde{w}_i and \tilde{v}_i , we follow Wu and Yan [27] and reformulate the constrained maximization problem as to maximize

$$\frac{1}{2} \sum_{i:T_i=1, R_{1i}=1} \log w_i + \frac{1}{2} \sum_{i:T_i=0, R_{0i}=1} \log v_i$$

subject to $w_i > 0$, $v_i > 0$ and

$$\frac{1}{2} \sum_{i:T_i=1, R_{1i}=1} w_i + \frac{1}{2} \sum_{i:T_i=0, R_{0i}=1} v_i = 1,$$

$$\frac{1}{2} \sum_{i:T_i=1, R_{1i}=1} w_i \tilde{\mathbf{g}}_{1i} + \frac{1}{2} \sum_{i:T_i=0, R_{0i}=1} v_i \tilde{\mathbf{g}}_{0i} = \mathbf{0},$$

where

$$\tilde{\mathbf{g}}_{1i} = \left(\frac{1}{\frac{n_{11}}{n} \frac{Y_{1i}}{\pi(\mathbf{Z}_i) \varpi_{1i}(\hat{\boldsymbol{\alpha}}_1)} - \frac{\delta_*}{2}} \right), \quad \tilde{\mathbf{g}}_{0i} = - \left(\frac{1}{\frac{n_{01}}{n} \frac{Y_{0i}}{\{1-\pi(\mathbf{Z}_i)\} \varpi_{0i}(\hat{\boldsymbol{\alpha}}_0)} + \frac{\delta_*}{2}} \right).$$

It can be shown through the standard Lagrange multiplier method that the maximizers are given by

$$\begin{aligned} \tilde{w}_i &= \frac{2}{n_{11} + n_{01}} \frac{1}{1 + \tilde{\boldsymbol{\rho}}^T \tilde{\mathbf{g}}_{1i}}, & \{i : T_i = 1, R_{1i} = 1\}, \\ \tilde{v}_i &= \frac{2}{n_{11} + n_{01}} \frac{1}{1 + \tilde{\boldsymbol{\rho}}^T \tilde{\mathbf{g}}_{0i}}, & \{i : T_i = 0, R_{0i} = 1\}, \end{aligned}$$

where the Lagrange multiplier $\tilde{\boldsymbol{\rho}}$ satisfies

$$\sum_{i:T_i=1, R_{1i}=1} (1 + \tilde{\boldsymbol{\rho}}^T \tilde{\mathbf{g}}_{1i})^{-1} \tilde{\mathbf{g}}_{1i} + \sum_{i:T_i=0, R_{0i}=1} (1 + \tilde{\boldsymbol{\rho}}^T \tilde{\mathbf{g}}_{0i})^{-1} \tilde{\mathbf{g}}_{0i} = \mathbf{0}.$$

The solution $\tilde{\boldsymbol{\rho}}$ can be solved by using the modified Newton-Raphson algorithm in Chen et al. [4] or Han [11].

(ii) *Multiple working models for the missingness probability.* We now consider the case where there are multiple working models for the missingness probability and propose an empirical likelihood ratio test for H_0 that is valid if one of the working models in each of the treatment and control groups is correctly specified. Let $\mathcal{P}_t = \{\varpi_t^{(j)}(\boldsymbol{\alpha}_t^{(j)}), j = 1, \dots, J_t\}$ be a set of working models for $\varpi_t(\mathbf{Z}, \mathbf{X}_t)$ and $\hat{\boldsymbol{\alpha}}_t^{(j)}$ be the estimator for $\boldsymbol{\alpha}_t^{(j)}$ by maximizing (3.1) with $\varpi_t(\boldsymbol{\alpha}_t)$ replaced by $\varpi_t^{(j)}(\boldsymbol{\alpha}_t^{(j)})$, $j = 1, \dots, J_t$, $t = 0, 1$. The two sets of working models are postulated independently with possibly different numbers of models J_1 and J_0 .

It can be verified by following similar arguments in Han and Wang [14] that for any $h_1(\mathbf{Z}, \mathbf{X}_1)$ and $h_0(\mathbf{Z}, \mathbf{X}_0)$, assuming all relevant expectations exist, the following equalities hold:

$$\begin{aligned} E(w(\mathbf{Z}, \mathbf{X}_1)[h_1(\mathbf{Z}, \mathbf{X}_1) - E\{h_1(\mathbf{Z}, \mathbf{X}_1)\}] | T = 1, R_1 = 1) &= 0, \\ E(v(\mathbf{Z}, \mathbf{X}_0)[h_0(\mathbf{Z}, \mathbf{X}_0) - E\{h_0(\mathbf{Z}, \mathbf{X}_0)\}] | T = 0, R_0 = 1) &= 0, \end{aligned} \quad (3.6)$$

where

$$\begin{aligned} w(\mathbf{Z}, \mathbf{X}_1) &= 1/\{\pi(\mathbf{Z})\varpi_1(\mathbf{Z}, \mathbf{X}_1)\} \quad \text{and} \\ v(\mathbf{Z}, \mathbf{X}_0) &= 1/[\{1 - \pi(\mathbf{Z})\}\varpi_0(\mathbf{Z}, \mathbf{X}_0)]. \end{aligned}$$

It follows that multiple working models in \mathcal{P}_1 and \mathcal{P}_0 can be simultaneously accommodated by taking $h_1(\mathbf{Z}, \mathbf{X}_1)$ to be $\pi(\mathbf{Z})\varpi_1^{(j)}(\boldsymbol{\alpha}_1^{(j)})$ and $h_0(\mathbf{Z}, \mathbf{X}_0)$ to be $\{1 - \pi(\mathbf{Z})\}\varpi_0^{(j)}(\boldsymbol{\alpha}_0^{(j)})$ to construct the following empirical version of (3.6) as constraints for the empirical likelihood inference,

$$\sum_{i:T_i=1, R_{1i}=1} w_i \left\{ \pi(\mathbf{Z}_i) \varpi_{1i}^{(j)}(\hat{\boldsymbol{\alpha}}_1^{(j)}) - \hat{\theta}_1^{(j)}(\hat{\boldsymbol{\alpha}}_1^{(j)}) \right\} = 0, \quad j = 1, \dots, J_1, \quad (3.7)$$

$$\sum_{i:T_i=0, R_{0i}=1} v_i \left[\{1 - \pi(\mathbf{Z}_i)\} \varpi_{0i}^{(j)}(\hat{\boldsymbol{\alpha}}_0^{(j)}) - \hat{\theta}_0^{(j)}(\hat{\boldsymbol{\alpha}}_0^{(j)}) \right] = 0, \quad j = 1, \dots, J_0, \quad (3.8)$$

where

$$\hat{\theta}_1^{(j)}(\boldsymbol{\alpha}_1^{(j)}) = n^{-1} \sum_{i:T_i=1} \varpi_{1i}^{(j)}(\boldsymbol{\alpha}_1^{(j)}) \quad \text{and} \quad \hat{\theta}_0^{(j)}(\boldsymbol{\alpha}_0^{(j)}) = n^{-1} \sum_{i:T_i=0} \varpi_{0i}^{(j)}(\boldsymbol{\alpha}_0^{(j)})$$

are, respectively, consistent estimators for $E\{\pi(\mathbf{Z})\varpi_1^{(j)}(\boldsymbol{\alpha}_1^{(j)})\}$ and $E[\{1 - \pi(\mathbf{Z})\}\varpi_0^{(j)}(\boldsymbol{\alpha}_0^{(j)})]$.

Under the current setting, the empirical likelihood ratio statistic for the treatment effect δ is computed as $W(\delta)$ specified in (3.5) with \hat{w}_i and \hat{v}_i maximizing $L(\mathbf{w}, \mathbf{v})$ subject to the normalization constraint (3.3) and the model constraints (3.7) and (3.8), and \tilde{w}_i and \tilde{v}_i maximizing $L(\mathbf{w}, \mathbf{v})$ subject to (3.3), (3.7), (3.8) plus an additional parameter constraint given by

$$\sum_{i:T_i=1, R_{1i}=1} w_i Y_{1i} - \sum_{i:T_i=0, R_{0i}=1} v_i Y_{0i} = \delta. \quad (3.9)$$

Note that the parameter constraint (3.9) has a different form compared to (3.4), due to the inclusion of model constraints (3.7) and (3.8). The asymptotic distribution of the empirical likelihood ratio statistic $W(\delta_*)$ under $H_0 : \delta = \delta_*$ is presented in the following theorem. The proof and the exact expression for the positive scaling factor σ_2 are provided in the Appendix.

Theorem 3.2. *Under Assumptions 1–3 with known $\pi(\mathbf{Z})$ and also assuming that both \mathcal{P}_1 and \mathcal{P}_0 contain a correctly specified model, the empirical likelihood ratio statistic $W(\delta_*)$ has an asymptotic distribution $\sigma_2 \chi_1^2$ under $H_0 : \delta = \delta_*$.*

The empirical likelihood ratio confidence interval for the treatment effect can be constructed similarly to before. Theorem 3.2 states that the proposed test is multiply robust in the sense that it is valid if one of the working models is correctly specified in each of \mathcal{P}_1 and \mathcal{P}_0 . It can be shown that $\hat{w}_i = 1/\{n_{11}(1 + \hat{\boldsymbol{\rho}}_1^T \hat{\mathbf{g}}_{1i})\}$ for $\{i : T_i = 1, R_{1i} = 1\}$ and $\hat{v}_i = 1/\{n_{01}(1 + \hat{\boldsymbol{\rho}}_0^T \hat{\mathbf{g}}_{0i})\}$ for $\{i : T_i = 0, R_{0i} = 1\}$, where $\hat{\boldsymbol{\rho}}_1$ and $\hat{\boldsymbol{\rho}}_0$ satisfy, respectively, the equations

$$\sum_{i:T_i=1, R_{1i}=1} \frac{\hat{\mathbf{g}}_{1i}}{1 + \hat{\boldsymbol{\rho}}_1^T \hat{\mathbf{g}}_{1i}} = \mathbf{0}, \quad \sum_{i:T_i=0, R_{0i}=1} \frac{\hat{\mathbf{g}}_{0i}}{1 + \hat{\boldsymbol{\rho}}_0^T \hat{\mathbf{g}}_{0i}} = \mathbf{0},$$

with

$$\begin{aligned} \hat{\mathbf{g}}_{1i} &= \left\{ \pi(\mathbf{Z}_i) \varpi_{1i}^{(1)}(\hat{\boldsymbol{\alpha}}_1^{(1)}) - \hat{\theta}_1^{(1)}(\hat{\boldsymbol{\alpha}}_1^{(1)}), \dots, \right. \\ &\quad \left. \pi(\mathbf{Z}_i) \varpi_{1i}^{(J_1)}(\hat{\boldsymbol{\alpha}}_1^{(J_1)}) - \hat{\theta}_1^{(J_1)}(\hat{\boldsymbol{\alpha}}_1^{(J_1)}) \right\}^T, \\ \hat{\mathbf{g}}_{0i} &= \left[\{1 - \pi(\mathbf{Z}_i)\} \varpi_{0i}^{(1)}(\hat{\boldsymbol{\alpha}}_0^{(1)}) - \hat{\theta}_0^{(1)}(\hat{\boldsymbol{\alpha}}_0^{(1)}), \dots, \right. \\ &\quad \left. \{1 - \pi(\mathbf{Z}_i)\} \varpi_{0i}^{(J_0)}(\hat{\boldsymbol{\alpha}}_0^{(J_0)}) - \hat{\theta}_0^{(J_0)}(\hat{\boldsymbol{\alpha}}_0^{(J_0)}) \right]^T. \end{aligned}$$

It follows from Han and Wang [14] that, when both \mathcal{P}_1 and \mathcal{P}_0 contain a correctly specified model, we have $\hat{w}_i = \{n\pi(\mathbf{Z}_i)\varpi_1(\mathbf{Z}_i, \mathbf{X}_{1i})\}^{-1}\{1 + O_p(n^{-1/2})\}$ for $\{i : T_i = 1, R_{1i} = 1\}$ and $\hat{v}_i = [n\{1 - \pi(\mathbf{Z}_i)\}\varpi_0(\mathbf{Z}_i, \mathbf{X}_{0i})]^{-1}\{1 + O_p(n^{-1/2})\}$ for $\{i : T_i = 0, R_{0i} = 1\}$. Therefore, the estimator computed as $\sum_{i:T_i=1, R_{1i}=1} \hat{w}_i \times Y_{1i} - \sum_{i:T_i=0, R_{0i}=1} \hat{v}_i Y_{0i}$ is consistent for δ , which justifies the use of the constraint (3.9) for defining $W(\delta)$. Following the same arguments as in Wu and Yan [27], we have

$$\begin{aligned}\tilde{w}_i &= \frac{2}{n_{11} + n_{01}} \frac{1}{1 + \tilde{\boldsymbol{\rho}}^T \tilde{\mathbf{g}}_{1i}}, & \{i : T_i = 1, R_{1i} = 1\}, \\ \tilde{v}_i &= \frac{2}{n_{11} + n_{01}} \frac{1}{1 + \tilde{\boldsymbol{\rho}}^T \tilde{\mathbf{g}}_{0i}}, & \{i : T_i = 0, R_{0i} = 1\},\end{aligned}$$

where $\tilde{\boldsymbol{\rho}}$ satisfies

$$\sum_{i:T_i=1, R_{1i}=1} \frac{\tilde{\mathbf{g}}_{1i}}{1 + \tilde{\boldsymbol{\rho}}^T \tilde{\mathbf{g}}_{1i}} + \sum_{i:T_i=0, R_{0i}=1} \frac{\tilde{\mathbf{g}}_{0i}}{1 + \tilde{\boldsymbol{\rho}}^T \tilde{\mathbf{g}}_{0i}} = \mathbf{0},$$

and

$$\tilde{\mathbf{g}}_{1i} = \begin{pmatrix} 1/2 \\ Y_{1i} - \delta_*/2 \\ \hat{\mathbf{g}}_{1i} \\ \mathbf{0}_{J_0 \times 1} \end{pmatrix}, \quad \tilde{\mathbf{g}}_{0i} = \begin{pmatrix} -1/2 \\ -Y_{0i} - \delta_*/2 \\ \mathbf{0}_{J_1 \times 1} \\ \hat{\mathbf{g}}_{0i} \end{pmatrix}.$$

The three Lagrange multipliers $\hat{\boldsymbol{\rho}}_1$, $\hat{\boldsymbol{\rho}}_0$ and $\tilde{\boldsymbol{\rho}}$ can all be calculated using a Newton-Raphson-type algorithm similar to the one described in Chen et al. [4] and Han [11].

For the special case $J_1 = J_0 = 1$, the scaling constant σ_2 in Theorem 3.2 does not reduce to σ_1 in Theorem 3.1 due to different formulations of constraints. In this special case, the formulation in Section 3.1 involves computing \tilde{w}_i and \tilde{v}_i , whereas the formulation in Section 3.2 involves computing \hat{w}_i , \hat{v}_i , \tilde{w}_i and \tilde{v}_i . The former formulation can only deal with one model for the missingness probability and the latter can deal with multiple models.

3.2. Unknown propensity scores

(i) *Single working model for the propensity score.* We now consider the more practical scenario where the propensity score of treatment assignment $\pi(\mathbf{Z})$ is unknown. Let $\pi(\mathbf{Z}; \boldsymbol{\gamma})$ denote a parametric model for $\pi(\mathbf{Z})$ with parameter $\boldsymbol{\gamma}$ and $\pi_i(\boldsymbol{\gamma}) = \pi(\mathbf{Z}_i; \boldsymbol{\gamma})$. Let $\hat{\boldsymbol{\gamma}}$ be the estimator of $\boldsymbol{\gamma}$ derived by maximizing the likelihood function

$$\prod_{i=1}^n \{\pi_i(\boldsymbol{\gamma})\}^{T_i} \{1 - \pi_i(\boldsymbol{\gamma})\}^{1-T_i}. \quad (3.10)$$

With the single correctly specified model $\pi(\mathbf{Z}; \boldsymbol{\gamma})$, the empirical likelihood ratio statistic $W(\delta)$ can be calculated in the same way as in Section 3.1 but with

$\pi(\mathbf{Z}_i)$ substituted by $\pi_i(\hat{\gamma})$. Depending on whether there is a single model or there are multiple working models for the missingness probability, we have the following two theorems parallel to Theorems 3.1 and 3.2. The exact expressions for σ_3 and σ_4 are given in the Appendix but detailed proofs are omitted due to similarities to the proofs of Theorems 3.1 and 3.2.

Theorem 3.3. *Under Assumptions 1–3 and also assuming that $\pi(\gamma)$ and $\varpi_t(\alpha_t)$ are correctly specified models for $\pi(\mathbf{Z})$ and $\varpi_t(\mathbf{Z}, \mathbf{X}_t)$, respectively, $t = 0, 1$, the empirical likelihood ratio statistic $W(\delta_*)$ has an asymptotic distribution $\sigma_3\chi_1^2$ under $H_0 : \delta = \delta_*$.*

Theorem 3.4. *Under Assumptions 1–3 and also assuming that $\pi(\gamma)$ is a correctly specified model for $\pi(\mathbf{Z})$ and both \mathcal{P}_1 and \mathcal{P}_0 contain a correctly specified model, the empirical likelihood ratio statistic $W(\delta_*)$ has an asymptotic distribution $\sigma_4\chi_1^2$ under $H_0 : \delta = \delta_*$.*

(ii) *Multiple working models for the propensity score.* When multiple working models are considered for the unknown propensity score $\pi(\mathbf{Z})$, construction of an empirical likelihood ratio test becomes significantly more challenging since using the estimated propensity scores from one particular working model will not lead to valid results. The general concept of multiple robustness assumes that one of the multiple working models is correctly specified but does not require the identification of the correct model.

We propose a two-step strategy to construct the empirical likelihood ratio test on the treatment effect. The first step accommodates the multiple working models for $\pi(\mathbf{Z})$ and produces weights that are a consistent estimate for $\pi(\mathbf{Z})$ when one of the working models is correctly specified. The second step incorporates multiple working models for the missingness probability similar to the procedures presented in Section 3.1 by using the weights obtained from the first step. Such a two-step procedure ensures multiple robustness of the proposed test.

Let $\mathcal{Q} = \{\pi^{(l)}(\gamma^{(l)}), l = 1, \dots, L\}$ denote a set of working models for $\pi(\mathbf{Z})$. An estimator $\hat{\gamma}^{(l)}$ for $\gamma^{(l)}$ can be derived by maximizing (3.10) but with $\pi(\gamma)$ replaced by $\pi^{(l)}(\gamma^{(l)})$. For the first step, we consider empirical likelihood weights p_i for subjects in the treatment group $\{i : T_i = 1\}$ and q_i for the subjects in the control group $\{i : T_i = 0\}$. The estimated weights \hat{p}_i and \hat{q}_i are obtained by maximizing the empirical likelihood function $L(\mathbf{p}, \mathbf{q}) = \prod_{i:T_i=1} p_i \prod_{i:T_i=0} q_i$ subject to the constraints

$$\begin{aligned}
 p_i > 0, \quad \sum_{i:T_i=1} p_i &= 1, \quad q_i > 0, \quad \sum_{i:T_i=0} q_i = 1, \\
 \sum_{i:T_i=1} p_i \pi_i^{(l)}(\hat{\gamma}^{(l)}) &= n^{-1} \sum_{i=1}^n \pi_i^{(l)}(\hat{\gamma}^{(l)}), \quad l = 1, \dots, L, \\
 \sum_{i:T_i=0} q_i \pi_i^{(l)}(\hat{\gamma}^{(l)}) &= n^{-1} \sum_{i=1}^n \pi_i^{(l)}(\hat{\gamma}^{(l)}), \quad l = 1, \dots, L.
 \end{aligned} \tag{3.11}$$

The last two sets of constraints in (3.11) are, respectively, the empirical versions of moment equalities

$$\begin{aligned} E(p(\mathbf{Z})[h(\mathbf{Z}) - E\{h(\mathbf{Z})\}] | T = 1) &= 0, \\ E(q(\mathbf{Z})[h(\mathbf{Z}) - E\{h(\mathbf{Z})\}] | T = 0) &= 0, \end{aligned}$$

with $p(\mathbf{Z}) = 1/\pi(\mathbf{Z})$, $q(\mathbf{Z}) = 1/\{1 - \pi(\mathbf{Z})\}$ and $h(\mathbf{Z})$ taken to be $\pi^{(l)}(\boldsymbol{\gamma}^{(l)})$, $l = 1, \dots, L$.

The most important consequence from the proposed first step is that if \mathcal{Q} contains a correctly specified model for $\pi(\mathbf{Z})$, then $(n\hat{p}_i)^{-1} = \pi(\mathbf{Z}_i)\{1 + O_p(n^{-1/2})\}$ for $\{i : T_i = 1\}$ and $(n\hat{q}_i)^{-1} = \{1 - \pi(\mathbf{Z}_i)\}\{1 + O_p(n^{-1/2})\}$ for $\{i : T_i = 0\}$, as shown in Han and Wang [14]. This leads to the second step to obtain the final weights \hat{w}_i and \hat{v}_i similar to the procedures described in Section 3.1 but replacing $\pi(\mathbf{Z}_i)$ for $\{i : T_i = 1\}$ and $1 - \pi(\mathbf{Z}_i)$ for $\{i : T_i = 0\}$ by $(n\hat{p}_i)^{-1}$ and $(n\hat{q}_i)^{-1}$, respectively. More specifically, we replace constraints (3.7) and (3.8) based on missingness probability working models by

$$\sum_{i:T_i=1, R_{1i}=1} w_i \left\{ (n\hat{p}_i)^{-1} \varpi_{1i}^{(j)}(\hat{\boldsymbol{\alpha}}_1^{(j)}) - \hat{\theta}_1^{(j)}(\hat{\boldsymbol{\alpha}}_1^{(j)}) \right\} = 0, \quad j = 1, \dots, J_1, \quad (3.12)$$

$$\sum_{i:T_i=0, R_{0i}=1} v_i \left\{ (n\hat{q}_i)^{-1} \varpi_{0i}^{(j)}(\hat{\boldsymbol{\alpha}}_0^{(j)}) - \hat{\theta}_0^{(j)}(\hat{\boldsymbol{\alpha}}_0^{(j)}) \right\} = 0, \quad j = 1, \dots, J_0. \quad (3.13)$$

The \tilde{w}_i and \tilde{v}_i required for computing the empirical likelihood ratio statistic $W(\delta)$ can be obtained by maximizing $L(\mathbf{w}, \mathbf{v})$ given by (3.2) under (3.3), (3.12), (3.13) and (3.9).

The calculation of \hat{w}_i , \hat{v}_i , \tilde{w}_i and \tilde{v}_i needed for $W(\delta)$ is the same as in Section 3.1, and this calculation requires \hat{p}_i for $\{i : T_i = 1\}$ and \hat{q}_i for $\{i : T_i = 0\}$. It is easy to show that $\hat{p}_i = 1/\{n_1(1 + \hat{\boldsymbol{\psi}}_1^T \hat{\mathbf{u}}_i)\}$ and $\hat{q}_i = 1/\{n_0(1 + \hat{\boldsymbol{\psi}}_0^T \hat{\mathbf{u}}_i)\}$, where $\hat{\boldsymbol{\psi}}_1$ and $\hat{\boldsymbol{\psi}}_0$ satisfy, respectively,

$$\sum_{i:T_i=1} \frac{\hat{\mathbf{u}}_i}{1 + \hat{\boldsymbol{\psi}}_1^T \hat{\mathbf{u}}_i} = \mathbf{0} \quad \text{and} \quad \sum_{i:T_i=0} \frac{\hat{\mathbf{u}}_i}{1 + \hat{\boldsymbol{\psi}}_0^T \hat{\mathbf{u}}_i} = \mathbf{0},$$

with $\hat{\mathbf{u}}_i^T = \{\pi_i^{(1)}(\hat{\boldsymbol{\gamma}}^{(1)}) - \hat{\zeta}^{(1)}(\hat{\boldsymbol{\gamma}}^{(1)}), \dots, \pi_i^{(L)}(\hat{\boldsymbol{\gamma}}^{(L)}) - \hat{\zeta}^{(L)}(\hat{\boldsymbol{\gamma}}^{(L)})\}$ and $\hat{\zeta}^{(l)}(\hat{\boldsymbol{\gamma}}^{(l)}) = n^{-1} \sum_{i=1}^n \pi_i^{(l)}(\hat{\boldsymbol{\gamma}}^{(l)})$, $l = 1, \dots, L$.

It is possible to derive the asymptotic distribution of $W(\delta_*)$ under $H_0 : \delta = \delta_*$ when each of \mathcal{Q} , \mathcal{P}_1 and \mathcal{P}_0 contains a correctly specified model for $\pi(\mathbf{Z})$, $\varpi_1(\mathbf{Z}, \mathbf{X}_1)$ and $\varpi_0(\mathbf{Z}, \mathbf{X}_0)$, respectively. However, such a derivation is extremely tedious and the resulting scaling factor similar to those appeared in Theorems 3.1–3.4 has a very complex form. We propose to use a bootstrap method, presented in Section 3.3 below, to bypass the derivation and estimation of the scaling factor. Simulation results show that the bootstrap method performs well for finite samples.

3.3. A bootstrap calibration

We present a bootstrap calibration for the proposed empirical likelihood ratio test to bypass the derivation and estimation of scaling factors. Let $\hat{\delta}$ denote our proposed estimator of the treatment effect δ given in (4.1) that will be studied in Section 4. The bootstrap procedure is as follows.

Step 1. Take a random sample of size n with replacement from the original data set.

Step 2. Compute the empirical likelihood ratio statistic $W(\delta)$ at $\delta = \hat{\delta}$ based on the data set from Step 1 to obtain $W(\hat{\delta})$.

Step 3. Repeat Steps 1 and 2 B times to obtain $\{W^{(1)}(\hat{\delta}), \dots, W^{(B)}(\hat{\delta})\}$.

Step 4. Let $b_{(1-\alpha)}$ be the $100(1-\alpha)\%$ sample quantile of $\{W^{(1)}(\hat{\delta}), \dots, W^{(B)}(\hat{\delta})\}$.

Step 5. The empirical likelihood ratio test rejects $H_0 : \delta = \delta_*$ if $W(\delta_*) \geq b_{(1-\alpha)}$ and the $(1 - \alpha)$ -level confidence interval can be constructed as $\{\delta : W(\delta) < b_{(1-\alpha)}\}$.

As one referee pointed out, an alternative is to replace Steps 4 and 5 above by the following two steps.

Step 4'. Compute $\hat{\sigma} = B^{-1} \sum_{b=1}^B W^{(b)}(\hat{\delta})$.

Step 5'. The empirical likelihood ratio test rejects $H_0 : \delta = \delta_*$ if $W(\delta_*) \geq \hat{\sigma} \chi_{1,(1-\alpha)}^2$ and the $(1 - \alpha)$ -level confidence interval can be constructed as $\{\delta : W(\delta) < \hat{\sigma} \chi_{1,(1-\alpha)}^2\}$.

The $\hat{\sigma}$ computed in Step 4' is an estimate of the scale parameter that appears in all theorems. Our simulation studies in Section 5 indicate that these two bootstrap procedures have very similar performance.

4. Estimation of the treatment effect

Our proposed maximum empirical likelihood estimator for the treatment effect δ is

$$\hat{\delta} = \sum_{i:T_i=1, R_{1i}=1} \hat{w}_i Y_{1i} - \sum_{i:T_i=0, R_{0i}=1} \hat{v}_i Y_{0i}, \quad (4.1)$$

where \hat{w}_i and \hat{v}_i are derived by maximizing $L(\mathbf{w}, \mathbf{v})$ given in (3.2) subject to the constraints (3.3), (3.7) and (3.8) if there is only one working model for the propensity score $\pi(\mathbf{Z})$ or the constraints (3.3), (3.12) and (3.13) if there are multiple working models for $\pi(\mathbf{Z})$. The estimator $\hat{\delta}$ is multiply robust since $\hat{w}_i = \{n\pi(\mathbf{Z}_i)\varpi_1(\mathbf{Z}_i, \mathbf{X}_{1i})\}^{-1}\{1 + O_p(n^{-1/2})\}$ and $\hat{v}_i = [n\{1 - \pi(\mathbf{Z}_i)\}\varpi_0(\mathbf{Z}_i, \mathbf{X}_{0i})]^{-1}\{1 + O_p(n^{-1/2})\}$ when one of the multiple working models for each of $\pi(\mathbf{Z})$, $\varpi_1(\mathbf{Z}, \mathbf{X}_1)$ and $\varpi_0(\mathbf{Z}, \mathbf{X}_0)$ is correctly specified.

In the following we will discuss how to further improve the robustness of $\hat{\delta}$ by accommodating working models for outcome regression. The outcome regression

models can also lead to efficiency gain in estimation. We first consider the case where $\pi(\mathbf{Z})$ is either known or modeled by a single working model $\pi(\boldsymbol{\gamma})$. Let $\mathcal{A}_t = \{a_t^{(k)}(\boldsymbol{\beta}_t^{(k)}) : k = 1, \dots, K_t\}$ be a set of working models for the outcome regression $E(Y_t | \mathbf{Z}, \mathbf{X}_t)$, $t = 0, 1$. We allow working models to be postulated separately for each of the treatment and control groups, and the numbers of models K_1 and K_0 can be different. By Assumptions 1 and 2, we have $E(Y_t | \mathbf{Z}, \mathbf{X}_t) = E(Y_t | \mathbf{Z}, \mathbf{X}_t, T = t, R_t = 1)$, which implies that $\boldsymbol{\beta}_t^{(k)}$ can be estimated by fitting the model $a_t^{(k)}(\boldsymbol{\beta}_t^{(k)})$ based on the complete cases within the group $T = t$, $t = 0, 1$. Let $\hat{\boldsymbol{\beta}}_t^{(k)}$ denote the estimator for $\boldsymbol{\beta}_t^{(k)}$. Our proposed estimator for δ is still $\hat{\delta}$ given in (4.1), but the \hat{w}_i and \hat{v}_i are now derived by maximizing (3.2) subject to (3.3), (3.7), (3.8) and the following additional constraints formed based on the outcome regression working models:

$$\begin{aligned} \sum_{i:T_i=1, R_{0i}=1} w_i \left\{ a_{1i}^{(k)}(\hat{\boldsymbol{\beta}}_1^{(k)}) - \hat{\eta}_1^{(k)}(\hat{\boldsymbol{\beta}}_1^{(k)}) \right\} &= 0, & k = 1, \dots, K_1, \\ \sum_{i:T_i=0, R_{0i}=1} v_i \left\{ a_{0i}^{(k)}(\hat{\boldsymbol{\beta}}_0^{(k)}) - \hat{\eta}_0^{(k)}(\hat{\boldsymbol{\beta}}_0^{(k)}) \right\} &= 0, & k = 1, \dots, K_0, \end{aligned} \quad (4.2)$$

where

$$\hat{\eta}_1^{(k)}(\boldsymbol{\beta}_1^{(k)}) = \frac{1}{n} \sum_{i:T_i=1} \frac{a_{1i}^{(k)}(\boldsymbol{\beta}_1^{(k)})}{\pi(\mathbf{Z}_i)} \quad \text{and} \quad \hat{\eta}_0^{(k)}(\boldsymbol{\beta}_0^{(k)}) = \frac{1}{n} \sum_{i:T_i=0} \frac{a_{0i}^{(k)}(\boldsymbol{\beta}_0^{(k)})}{1 - \pi(\mathbf{Z}_i)}$$

are consistent estimators of $E\{a_1^{(k)}(\boldsymbol{\beta}_1^{(k)})\}$ and $E\{a_0^{(k)}(\boldsymbol{\beta}_0^{(k)})\}$, respectively. The constraints in (4.2) are empirical versions of (3.6) by taking $h_1(\mathbf{Z}, \mathbf{X}_1)$ to be $a_1^{(k)}(\boldsymbol{\beta}_1^{(k)})$ and $h_0(\mathbf{Z}, \mathbf{X}_0)$ to be $a_0^{(k)}(\boldsymbol{\beta}_0^{(k)})$. When $\pi(\mathbf{Z})$ is unknown but is modeled by $\pi(\boldsymbol{\gamma})$, the $\pi(\mathbf{Z})$ in all the constraints is replaced by $\pi(\hat{\boldsymbol{\gamma}})$. Multiple robustness of $\hat{\delta}$ is summarized by the theorem below, the proof of which is given in the Appendix.

Theorem 4.1. *Suppose that $\pi(\mathbf{Z})$ is either known or correctly modeled by $\pi(\boldsymbol{\gamma})$. If (i) \mathcal{P}_1 contains a correctly specified model for $\varpi_1(\mathbf{Z}, \mathbf{X}_1)$ or \mathcal{A}_1 contains a correctly specified model for $E(Y_1 | \mathbf{Z}, \mathbf{X}_1)$; and (ii) \mathcal{P}_0 contains a correctly specified model for $\varpi_0(\mathbf{Z}, \mathbf{X}_0)$ or \mathcal{A}_0 contains a correctly specified model for $E(Y_0 | \mathbf{Z}, \mathbf{X}_0)$, then $\hat{\delta}$ is a consistent estimator of δ .*

When $\pi(\mathbf{Z})$ is unknown and there are multiple working models $\mathcal{Q} = \{\pi^{(l)}(\boldsymbol{\gamma}^{(l)}), l = 1, \dots, L\}$, the two-step strategy described in Section 3.2 can be used to construct the estimator for δ . Let \hat{p}_i and \hat{q}_i be derived through (3.11) based on models in \mathcal{Q} . Let \hat{w}_i and \hat{v}_i be derived by maximizing $L(\mathbf{w}, \mathbf{v})$ given in (3.2) subject to (3.3), (3.12), (3.13) and the additional constraints (4.2) but with the model-averages in (4.2) redefined as $\hat{\eta}_1^{(k)}(\hat{\boldsymbol{\beta}}_1^{(k)}) = \sum_{i:T_i=1} \hat{p}_i a_{1i}^{(k)}(\hat{\boldsymbol{\beta}}_1^{(k)})$ and $\hat{\eta}_0^{(k)}(\hat{\boldsymbol{\beta}}_0^{(k)}) = \sum_{i:T_i=0} \hat{q}_i a_{0i}^{(k)}(\hat{\boldsymbol{\beta}}_0^{(k)})$. Theorem 4.2 summarizes the property of $\hat{\delta}$ and the proof is given in the Appendix.

Theorem 4.2. *If (i) \mathcal{Q} contains a correctly specified model for $\pi(\mathbf{Z})$; (ii) \mathcal{P}_1 contains a correctly specified model for $\varpi_1(\mathbf{Z}, \mathbf{X}_1)$ or \mathcal{A}_1 contains a correctly specified model for $E(Y_1 | \mathbf{Z}, \mathbf{X}_1)$; and (iii) \mathcal{P}_0 contains a correctly*

specified model for $\varpi_0(\mathbf{Z}, \mathbf{X}_0)$ or \mathcal{A}_0 contains a correctly specified model for $E(Y_0 | \mathbf{Z}, \mathbf{X}_0)$, then $\hat{\delta}$ is a consistent estimator of δ .

One might think that modeling $E(Y_t | \mathbf{Z})$, in addition to $\pi(\mathbf{Z})$, $\varpi_t(\mathbf{Z}, \mathbf{X}_t)$ and $E(Y_t | \mathbf{Z}, \mathbf{X}_t)$, could further improve the robustness of estimation. This is not the case in general. As pointed out in Davidian et al. [8], because $E(Y_t | \mathbf{Z}) \neq E(Y_t | \mathbf{Z}, T = t, R_t = 1)$, fitting a model for $E(Y_t | \mathbf{Z})$ cannot be based on a complete-case analysis but rather requires modeling other quantities, which complicates achieving multiple robustness since the parameters in a model for $E(Y_t | \mathbf{Z})$ may not be consistently estimated even if this model is correctly specified. For this reason, we do not consider modeling $E(Y_t | \mathbf{Z})$.

Standard error of $\hat{\delta}$ is oftentimes needed to make inference on δ . A practically useful approach to computing the standard error is the bootstrap method. It can be implemented easily by taking repeated bootstrap samples similar to Step 1 of the method described in Section 3.3 to obtain bootstrap copies of $\hat{\delta}$, which then lead to the bootstrap standard error. The reliability of the bootstrap method for standard error calculation for multiply robust estimators in missing data literature has been well documented (Han 11, 12). We will examine its numerical performance under the current setting in Section 5.

In addition to the multiple robustness property, the proposed maximum empirical likelihood estimator $\hat{\delta}$ has other advantages compared to existing ones. The weights \hat{w}_i and \hat{v}_i are positive and sum to one, and thus both $\sum_{i:T_i=1, R_{1i}=1} \hat{w}_i Y_{1i}$ and $\sum_{i:T_i=0, R_{0i}=1} \hat{v}_i Y_{0i}$ are convex combinations of the observed responses. The estimator $\hat{\delta}$ therefore always falls into the parameter space for the treatment effect. Issues with the conventional inverse probability weighted and augmented inverse probability weighted estimators when some estimated values of $\pi(\mathbf{Z})$ and/or $\varpi_t(\mathbf{Z}, \mathbf{X}_t)$ are close to zero are also significantly mitigated under our proposed empirical likelihood approach. Han [11] contains more detailed discussions and simulation results on this particular aspect in missing data analysis.

Note that the empirical likelihood ratio test described in Section 3 can also accommodate outcome regression models in the same way as for estimation, and the test based on the bootstrap calibration as in Section 3.3 remains multiply robust.

5. Simulation study

We conducted simulation studies to evaluate the finite-sample performance of the proposed empirical likelihood ratio test and the maximum empirical likelihood estimator for the treatment effect. The simulation sample data were generated as follows. At the baseline, a pretest measurement was generated as $Z \sim \text{Uniform}(-2.5, 2.5)$, and then the propensity score of treatment assignment was set to be $\pi(Z) = \{1 + \exp(1 - 0.8Z^2)\}^{-1}$, which leads to approximately 60% and 40% of subjects in the treatment ($T = 1$) and the control ($T = 0$) groups, respectively. For the scenario of no treatment effect ($\delta = 0$), the intermediate covariate was generated as $X_t \sim N(1 + Z, 1)$, and the posttest

potential outcome was generated as $Y_t | X_t \sim N\{a_t(X_t), 2X_t^2 + 2\}$, where $a_t(X_t) = 1 + 4X_t^2$, $t = 0, 1$. For the scenario of non-zero δ , the intermediate covariate was generated as $X_t \sim N(1 + t + Z, 1)$, $t = 0, 1$, and the posttest potential outcome was generated as $Y_t | X_t \sim N\{a_t(X_t), 2X_t^2 + 2\}$, where $a_1(X_1) = 1 + 4X_1^2$, $a_0(X_0) = \beta_{00*} + 2X_0^2$ and $\beta_{00*} = 1, 5, 9$ and 13 , which leads to $\delta = 20.2, 16.2, 12.2$ and 8.2 , respectively. For all scenarios, the missingness probabilities are set to be $\varpi_1(Z, X_1) = \{1 + \exp(0.6 - 0.1Z - 0.7X_1)\}^{-1}$ and $\varpi_0(Z, X_0) = \{1 + \exp(-0.4 + 0.1Z - 0.6X_0)\}^{-1}$, resulting in a missingness rate of 29% for the control group in all scenarios, and 49% for treatment group when $\delta = 0$ and 37% for the treatment group when $\delta \neq 0$.

We considered two parametric working models listed below for each of $\pi(Z)$, $\varpi_t(Z, X_t)$ and $E(Y_t | Z, X_t)$. The first working model in each pair, $\pi^{(1)}(\gamma^{(1)})$, $\varpi_t^{(1)}(\alpha_t^{(1)})$ and $a_t^{(1)}(\beta_t^{(1)})$, is correctly specified:

$$\begin{aligned}\pi^{(1)}(\gamma^{(1)}) &= \{1 + \exp(\gamma_0^{(1)} + \gamma_1^{(1)}Z^2)\}^{-1}, \\ \pi^{(2)}(\gamma^{(2)}) &= 1 - \exp[-\exp\{\gamma_0^{(2)} + \gamma_1^{(2)}Z + \gamma_2^{(2)}\exp(Z)\}], \\ \varpi_t^{(1)}(\alpha_t^{(1)}) &= \{1 + \exp(\alpha_{t0}^{(1)} + \alpha_{t1}^{(1)}Z + \alpha_{t2}^{(1)}X_t)\}^{-1}, \\ \varpi_t^{(2)}(\alpha_t^{(2)}) &= 1 - \exp\{-\exp(\alpha_{t0}^{(2)} + \alpha_{t1}^{(2)}X_t^2)\}, \\ a_t^{(1)}(\beta_t^{(1)}) &= \beta_{t0}^{(1)} + \beta_{t1}^{(1)}X_t^2, \\ a_t^{(2)}(\beta_t^{(2)}) &= \beta_{t0}^{(2)} + \beta_{t1}^{(2)}X_t + \beta_{t2}^{(2)}\exp(X_t).\end{aligned}$$

Here we demonstrate the steps for computing the multiply robust estimator and the EL ratio test that use all of the above working models. Estimators and tests that use other combinations of these working models can be computed similarly.

Step 1. Calculate $\hat{\gamma}^{(l)}$, $l = 1, 2$, by maximizing

$$\prod_{i=1}^n \{\pi_i^{(l)}(\gamma^{(l)})\}^{T_i} \{1 - \pi_i^{(l)}(\gamma^{(l)})\}^{1-T_i}.$$

Step 2. Calculate $\hat{\alpha}_t^{(j)}$, $j = 1, 2$, by maximizing

$$\prod_{i:T_i=t} \{\varpi_{ti}^{(j)}(\alpha_t^{(j)})\}^{R_{ti}} \{1 - \varpi_{ti}^{(j)}(\alpha_t^{(j)})\}^{1-R_{ti}}.$$

Step 3. Calculate $\hat{\beta}_t^{(k)}$, $k = 1, 2$, by fitting the outcome regression model $a_t^{(k)}(\beta_t^{(k)})$ based on the complete cases $\{i : T_i = t, R_{ti} = 1\}$ in each intervention group.

Step 4. Calculate the weights

$$\begin{aligned}\hat{p}_i &= \{n_1(1 + \hat{\psi}_1^T \hat{\mathbf{u}}_i)\}^{-1}, \quad \{i : T_i = 1\}, \\ \hat{q}_i &= \{n_0(1 + \hat{\psi}_0^T \hat{\mathbf{u}}_i)\}^{-1}, \quad \{i : T_i = 0\},\end{aligned}$$

where $\hat{\boldsymbol{\psi}}_1$ and $\hat{\boldsymbol{\psi}}_0$ minimize, respectively,

$$F_1(\boldsymbol{\psi}_1) = -n^{-1} \sum_{i=1}^n T_i \log(1 + \boldsymbol{\psi}_1^T \hat{\boldsymbol{u}}_i),$$

$$F_0(\boldsymbol{\psi}_0) = -n^{-1} \sum_{i=1}^n (1 - T_i) \log(1 + \boldsymbol{\psi}_0^T \hat{\boldsymbol{u}}_i).$$

Here $\hat{\boldsymbol{u}}_i = \{\pi_i^{(1)}(\hat{\boldsymbol{\gamma}}^{(1)}) - \hat{\zeta}^{(1)}(\hat{\boldsymbol{\gamma}}^{(1)}), \pi_i^{(2)}(\hat{\boldsymbol{\gamma}}^{(2)}) - \hat{\zeta}^{(2)}(\hat{\boldsymbol{\gamma}}^{(2)})\}^T$ and $\hat{\zeta}^{(l)}(\hat{\boldsymbol{\gamma}}^{(l)}) = n^{-1} \sum_{i=1}^n \pi_i^{(l)}(\hat{\boldsymbol{\gamma}}^{(l)})$, $l = 1, 2$.
 Step 5. Calculate the weights

$$\hat{w}_i = \{n_{11}(1 + \hat{\boldsymbol{\rho}}_1^T \hat{\boldsymbol{g}}_{1i})\}^{-1}, \quad \{i : T_i = 1, R_{1i} = 1\},$$

$$\hat{v}_i = \{n_{01}(1 + \hat{\boldsymbol{\rho}}_0^T \hat{\boldsymbol{g}}_{0i})\}^{-1}, \quad \{i : T_i = 0, R_{0i} = 1\},$$

where $\hat{\boldsymbol{\rho}}_1$ and $\hat{\boldsymbol{\rho}}_0$ minimize, respectively,

$$F_1(\boldsymbol{\rho}_1) = -n^{-1} \sum_{i=1}^n T_i R_{1i} \log(1 + \boldsymbol{\rho}_1^T \hat{\boldsymbol{g}}_{1i}),$$

$$F_0(\boldsymbol{\rho}_0) = -n^{-1} \sum_{i=1}^n (1 - T_i) R_{0i} \log(1 + \boldsymbol{\rho}_0^T \hat{\boldsymbol{g}}_{0i}).$$

Here

$$\hat{\boldsymbol{\theta}}_1^{(j)}(\boldsymbol{\alpha}_1^{(j)}) = n^{-1} \sum_{i:T_i=1} \varpi_{1i}^{(j)}(\boldsymbol{\alpha}_1^{(j)}),$$

$$\hat{\boldsymbol{\theta}}_0^{(j)}(\boldsymbol{\alpha}_0^{(j)}) = n^{-1} \sum_{i:T_i=0} \varpi_{0i}^{(j)}(\boldsymbol{\alpha}_0^{(j)}),$$

$$\hat{\boldsymbol{\eta}}_1^{(k)}(\boldsymbol{\beta}_1^{(k)}) = \sum_{i:T_i=1} \hat{p}_i a_{1i}^{(k)}(\boldsymbol{\beta}_1^{(k)}),$$

$$\hat{\boldsymbol{\eta}}_0^{(k)}(\boldsymbol{\beta}_0^{(k)}) = \sum_{i:T_i=0} \hat{q}_i a_{0i}^{(k)}(\boldsymbol{\beta}_0^{(k)}),$$

$$\hat{\boldsymbol{g}}_{1i} = \left\{ \begin{array}{l} (n\hat{p}_i)^{-1} \varpi_{1i}^{(1)}(\hat{\boldsymbol{\alpha}}_1^{(1)}) - \hat{\boldsymbol{\theta}}_1^{(1)}(\hat{\boldsymbol{\alpha}}_1^{(1)}) \\ (n\hat{p}_i)^{-1} \varpi_{1i}^{(2)}(\hat{\boldsymbol{\alpha}}_1^{(2)}) - \hat{\boldsymbol{\theta}}_1^{(2)}(\hat{\boldsymbol{\alpha}}_1^{(2)}) \\ a_{1i}^{(1)}(\hat{\boldsymbol{\beta}}_1^{(1)}) - \hat{\boldsymbol{\eta}}_1^{(1)}(\hat{\boldsymbol{\beta}}_1^{(1)}) \\ a_{1i}^{(2)}(\hat{\boldsymbol{\beta}}_1^{(2)}) - \hat{\boldsymbol{\eta}}_1^{(2)}(\hat{\boldsymbol{\beta}}_1^{(2)}) \end{array} \right\},$$

$$\hat{\boldsymbol{g}}_{0i} = \left\{ \begin{array}{l} (n\hat{q}_i)^{-1} \varpi_{0i}^{(1)}(\hat{\boldsymbol{\alpha}}_0^{(1)}) - \hat{\boldsymbol{\theta}}_0^{(1)}(\hat{\boldsymbol{\alpha}}_0^{(1)}) \\ (n\hat{q}_i)^{-1} \varpi_{0i}^{(2)}(\hat{\boldsymbol{\alpha}}_0^{(2)}) - \hat{\boldsymbol{\theta}}_0^{(2)}(\hat{\boldsymbol{\alpha}}_0^{(2)}) \\ a_{0i}^{(1)}(\hat{\boldsymbol{\beta}}_0^{(1)}) - \hat{\boldsymbol{\eta}}_0^{(1)}(\hat{\boldsymbol{\beta}}_0^{(1)}) \\ a_{0i}^{(2)}(\hat{\boldsymbol{\beta}}_0^{(2)}) - \hat{\boldsymbol{\eta}}_0^{(2)}(\hat{\boldsymbol{\beta}}_0^{(2)}) \end{array} \right\}.$$

Step 6. The multiply robust estimator is calculated as in (4.1).

Step 7. Calculate the weights

$$\begin{aligned}\tilde{w}_i &= 2/\{(n_{11} + n_{01})(1 + \tilde{\boldsymbol{\rho}}^T \tilde{\boldsymbol{g}}_{1i})\}, & \{i : T_i = 1, R_{1i} = 1\}, \\ \tilde{v}_i &= 2/\{(n_{11} + n_{01})(1 + \tilde{\boldsymbol{\rho}}^T \tilde{\boldsymbol{g}}_{0i})\}, & \{i : T_i = 0, R_{0i} = 1\},\end{aligned}$$

where $\tilde{\boldsymbol{\rho}}$ minimizes

$$F(\boldsymbol{\rho}) = \sum_{i:T_i=1, R_{1i}=1} \log(1 + \boldsymbol{\rho}^T \tilde{\boldsymbol{g}}_{1i}) + \sum_{i:T_i=0, R_{0i}=1} \log(1 + \boldsymbol{\rho}^T \tilde{\boldsymbol{g}}_{0i}),$$

where

$$\begin{aligned}\tilde{\boldsymbol{g}}_{1i} &= (1/2, Y_{1i} - \delta_*/2, \hat{\boldsymbol{g}}_{1i}^T, \mathbf{0}_{J_0 \times 1}^T)^T, \\ \tilde{\boldsymbol{g}}_{0i} &= (-1/2, -Y_{0i} - \delta_*/2, \mathbf{0}_{J_1 \times 1}^T, \hat{\boldsymbol{g}}_{0i}^T)^T.\end{aligned}$$

Step 8. The empirical likelihood ratio statistic is calculated as in (3.5).

In the case of only one working model $\pi(\mathbf{Z}; \boldsymbol{\gamma})$ for the propensity score, one should skip *Step 4* and replace $(n\hat{p}_i)^{-1}$ and $(n\hat{q}_i)^{-1}$ in *Step 5* with $\pi(\mathbf{Z}_i; \hat{\boldsymbol{\gamma}})$ and $1 - \pi(\mathbf{Z}_i; \hat{\boldsymbol{\gamma}})$ respectively.

Results reported in Tables 2 and 3 are based on 1000 repeated simulation samples, and for each sample 1000 bootstrap samples were used to calculate the critical value of the empirical likelihood ratio test or the standard error of the point estimator.

Table 2 contains results on the performance of the empirical likelihood ratio test and the Wald test based on four different combinations of models for $\pi(Z)$ and $\varpi_t(Z, X_t)$. The outcome regression models for $E(Y_t | Z, X_t)$ were not used for these tests. It can be seen that the empirical likelihood ratio test has type I error close to 5% and the empirical likelihood ratio confidence interval has coverage probability close to 95%, and both have better performance than the methods based on the Wald statistic. The Wald test has higher power compared to the empirical likelihood ratio test. Another observation is that the two bootstrap procedures described in Section 3.3 have similar performance.

Table 3 summarizes the simulation results on point estimation for the treatment effect. We focused on the scenario with $\delta = 20.2$ and sample size $n = 400$. We considered four different scenarios for the propensity score: (I) $\pi(Z)$ is known; (II) $\pi(Z)$ is correctly modeled by $\pi^{(1)}(\boldsymbol{\gamma}^{(1)})$; (III) $\pi(Z)$ is incorrectly modeled by $\pi^{(2)}(\boldsymbol{\gamma}^{(2)})$; and (IV) both models $\pi^{(1)}(\boldsymbol{\gamma}^{(1)})$ and $\pi^{(2)}(\boldsymbol{\gamma}^{(2)})$ are used. Our proposed multiply robust estimator along with the inverse probability weighted estimator and the augmented inverse probability weighted estimator were included in the simulation. The working models used for the missingness probability and the outcome regression are indicated as part of the notation in the first column. For instance, the multiply robust estimator $\text{MR-}\varpi^{(1,2)}_a^{(1,2)}$ was computed with both models for the missingness probability and both models for the outcome regression. Performance of estimators is evaluated through

TABLE 2. Simulation results (in %) of tests on $H_0 : \delta = 0$ and confidence intervals. The Type I error and power are summarized based on significance level 5% and the nominal level for the confidence intervals is 95%. Models in brackets are the ones used.

δ	n	$\{\pi^{(1)}, \varpi^{(1)}\}$				$\{\pi^{(1)}, \varpi^{(1)}, \varpi^{(2)}\}$							
		ELR1		Wald		ELR2		ELR1		Wald		ELR2	
		Error-I	Cover	Error-I	Cover	Error-I	Cover	Error-I	Cover	Error-I	Cover	Error-I	Cover
0	400	7.6	92.4	12	88	7.7	92.3	6.4	93.6	10.7	89.3	6.2	93.8
	800	5.8	94.2	8.7	91.3	6.3	93.7	6.3	93.7	9.8	90.2	6.3	93.7
		Power	Cover	Power	Cover	Power	Cover	Power	Cover	Power	Cover	Power	Cover
8.2	400	46.5	93.8	62.1	91.5	46.7	93.9	73.8	94.4	79.8	93.3	74.2	94.6
	800	57.8	95	79	93.9	58.4	95.2	88.9	95.3	94.7	93.8	89.5	95.3
12.2	400	66.7	94	82.8	91.8	67.8	94	91	94.4	94.6	93.3	91.6	94.6
	800	80.3	95.1	94.7	94	81.3	95	97.9	95.3	99.2	93.8	98.1	95.3
16.2	400	79.1	93.8	92.6	92.5	80.2	93.5	96.6	94.4	98.5	93.3	96.7	94.6
	800	90.9	95.6	98.4	94.9	92.4	95.4	99.3	95.3	100	93.8	99.4	95.3
20.2	400	85.1	94.3	96.7	93.7	87	94.2	98.1	94.4	99.5	93.3	98.3	94.6
	800	95.9	95.4	99.5	95.2	96.3	95	99.8	95.3	100	93.8	99.8	95.3
		$\{\pi^{(1)}, \pi^{(2)}, \varpi^{(1)}\}$				$\{\pi^{(1)}, \pi^{(2)}, \varpi^{(1)}, \varpi^{(2)}\}$							
		ELR1		Wald		ELR2		ELR1		Wald		ELR2	
δ	n	Error-I	Cover	Error-I	Cover	Error-I	Cover	Error-I	Cover	Error-I	Cover	Error-I	Cover
0	400	6.4	93.6	8.8	91.2	5.9	94.1	6.4	93.6	9.3	90.7	6	94
	800	5.3	94.7	7.4	92.6	5.2	94.8	5.4	94.6	8.3	91.7	5.6	94.4
		Power	Cover	Power	Cover	Power	Cover	Power	Cover	Power	Cover	Power	Cover
8.2	400	77.5	95.4	83	93.8	78	95.4	75.9	95.2	82.1	93.9	76.4	95.3
	800	91.9	94.8	95.5	93.8	92.7	94.9	91.6	94.9	95.7	93.8	91.8	95
12.2	400	91.6	95.4	95.2	93.8	92.3	95.4	91.7	95.2	95.6	93.9	92.1	95.3
	800	97.8	94.8	99.5	93.8	98.2	94.9	97.9	94.9	99.5	93.8	98.1	95
16.2	400	96.5	95.4	98.8	93.8	97.1	95.4	97	95.2	98.8	93.9	97.1	95.3
	800	99.3	94.8	100	93.8	99.3	94.9	99.4	94.9	100	93.8	99.4	95
20.2	400	98.5	95.4	99.6	93.8	98.9	95.4	98.5	95.2	99.6	93.9	98.7	95.3
	800	99.8	94.8	100	93.8	99.8	94.9	99.8	94.9	100	93.8	99.8	95

Non-randomized pretest-posttest studies with missing data

ELR1: empirical likelihood ratio test with the bootstrap procedure Steps 1-5 as in Section 3.3. Wald: Wald test. ELR2: empirical likelihood ratio test with the bootstrap procedure Steps 1-3 and Steps 4'-5' as in Section 3.3.
 Error-I: type I error. Power: power of the test. Cover: coverage probability of the 95% confidence intervals.

relative bias and root mean squared error. To evaluate the performance of the bootstrap method, we also included under scenario (IV) the values of empirical standard error and the bootstrap-based standard error.

From Table 3, the proposed multiply robust estimator has negligible relative bias under the combination of multiple working models as outlined in Theorems 4.1 and 4.2. The inverse probability weighted estimator and the augmented inverse probability weighted estimator cannot accommodate multiple working models. Another interesting observation is that, compared to the inverse probability weighted estimator and the augmented inverse probability weighted estimator, the multiply robust estimator has smaller bias when the combination of working models does not satisfy the specification outlined in Theorems 4.1 and 4.2. The small bias of multiply robust estimators when they are not theoretically consistent has been previously documented in the missing data literature (Han 11, 12, Chen and Haziza 7), and is due to the calibration constraints used to derive the weights \hat{w}_i and \hat{v}_i . Han [12] contains more discussion on this intriguing observation. We also observe from Table 3 that the bootstrap standard error is very close to the empirical standard error, showing that the bootstrap method is reliable.

TABLE 3
Simulation results ($\times 10^2$) for point estimation ($\delta = 20.2$ and $n = 400$)

	I. $\{\pi(Z)\}$		II. $\{\pi^{(1)}\}$		III. $\{\pi^{(2)}\}$		IV. $\{\pi^{(1)}, \pi^{(2)}\}$			
	rBias	RMSE	rBias	RMSE	rBias	RMSE	rBias	RMSE	ESE	BSE
MR- $\varpi^{(1)}$	0	328	0	298	4	309	0	286	298	270
MR- $\varpi^{(2)}$	8	369	8	347	12	398	7	333	317	280
MR- $a^{(1)}$	0	357	0	310	-17	965	1	248	260	237
MR- $a^{(2)}$	11	400	11	386	-11	903	11	354	290	260
MR- $\varpi^{(1,2)}$	-1	331	0	298	4	304	-1	288	298	272
MR- $\varpi^{(1)}a^{(1)}$	0	358	0	312	-14	861	1	252	263	241
MR- $\varpi^{(1)}a^{(2)}$	0	336	1	315	-12	764	1	278	283	258
MR- $\varpi^{(2)}a^{(1)}$	0	357	0	311	-14	866	1	252	264	241
MR- $\varpi^{(2)}a^{(2)}$	-4	351	-4	316	-18	823	-4	272	274	248
MR- $a^{(1,2)}$	0	357	0	311	-9	689	1	250	259	239
MR- $\varpi^{(1,2)}a^{(1)}$	0	357	0	311	-12	799	1	252	263	241
MR- $\varpi^{(1,2)}a^{(2)}$	1	340	1	306	-11	741	2	264	273	249
MR- $\varpi^{(1)}a^{(1,2)}$	0	357	0	311	-8	632	1	253	264	242
MR- $\varpi^{(2)}a^{(1,2)}$	0	359	0	311	-7	626	1	254	266	242
MR- $\varpi^{(1,2)}a^{(1,2)}$	0	357	0	314	-6	590	1	261	265	243
IPW- $\varpi^{(1)}$	0	401	0	378	-212	34122				
IPW- $\varpi^{(2)}$	10	427	11	394	-198	33041				
AIPW- $\varpi^{(1)}a^{(1)}$	0	361	0	314	-202	31829				
AIPW- $\varpi^{(1)}a^{(2)}$	0	415	0	387	-202	31473				
AIPW- $\varpi^{(2)}a^{(1)}$	0	358	0	309	-203	32084				
AIPW- $\varpi^{(2)}a^{(2)}$	-8	396	-8	361	-213	31968				

MR: the multiply robust estimator. IPW: the inverse probability weighted estimator. AIPW: the augmented inverse probability weighted estimator. rBias: relative bias. RMSE: root mean square error. ESE: empirical standard error. BSE: bootstrap standard error.

6. Additional remarks

Pretest-posttest studies are commonly used by social science, medical and health researchers, where non-randomized treatment assignment and missing data are two important issues. Valid statistical analyses depend on proper handling of models for the propensity score of treatment assignment, the missingness probability and the outcome regression. Our proposed empirical likelihood approach to testing and estimation of the treatment effect provides a unified inference tool to incorporate models from different sources. Theoretical and simulation results presented in this paper show that the proposed approach is promising. It is worthwhile to point out that the proposed method can still be applied when there is only a single model for each quantity, and thus it indeed provides a powerful alternative to existing methods.

Based on our experience from simulation studies, the proposed method is very robust to high collinearity among the possibly multiple working models for the same quantity. In practice, when predictions from different working models are very close so that collinearity may be of an issue, an ad hoc solution would be to remove some of those models because they provide mostly redundant information. It will be of future research interest to develop a formal way to address potential collinearity problem.

This paper focuses on the robustness of estimation. In the missing data literature, estimation efficiency is another important issue (Robins et al. 21, Tsiatis 26). However, a theoretical investigation on efficiency in our setting with multiple working models is both challenging and lack of practical implication. The typical framework for semiparametric efficiency theory in the missing data literature (e.g., Tsiatis 26) assumes that both $\pi(\mathbf{Z})$ and $\varpi_t(\mathbf{Z}, \mathbf{X}_t)$ are correctly modeled and the correct models are identified. For non-randomized pretest-posttest studies such an assumption is often unrealistic. This was the major motivation that we considered multiple working models. The simulation results presented in this paper as well as findings from the existing missing data literature, such as Han [11], Han [12], Chen and Haziza [7], among others, show that multiply robust estimators typically have similar or higher efficiency compared to other estimators when the same working models are used.

7. Appendix

7.1. Expressions for the Scaling Constants in the Theorems

7.1.1. Expressions for Theorem 3.1

The scaling constant for the asymptotic distribution in Theorem 3.1 is given by

$$\sigma_1 = (D_1 + D_0)^{-1} \text{var} \left\{ \varphi - \mathbf{A}_1 E(\mathbf{S}_1 \mathbf{S}_1^T)^{-1} \mathbf{S}_1 + \mathbf{A}_0 E(\mathbf{S}_0 \mathbf{S}_0^T)^{-1} \mathbf{S}_0 \right\},$$

where

$$\varphi = \frac{TR_1 Y_1}{\pi(\mathbf{Z}) \varpi_1(\mathbf{Z}, \mathbf{X}_1)} - \frac{(1 - T) R_0 Y_0}{\{1 - \pi(\mathbf{Z})\} \varpi_0(\mathbf{Z}, \mathbf{X}_0)} - \delta_*,$$

$$\begin{aligned}
g_{1*} &= \frac{Y_1}{\pi(\mathbf{Z})\varpi_1(\mathbf{Z}, \mathbf{X}_1)} - \frac{\mu_0 + \delta_*}{\mathbb{P}(T = 1, R_1 = 1)}, \\
g_{0*} &= \frac{Y_0}{\{1 - \pi(\mathbf{Z})\}\varpi_0(\mathbf{Z}, \mathbf{X}_0)} - \frac{\mu_0}{\mathbb{P}(T = 0, R_0 = 1)}, \\
D_1 &= E\{TR_1g_{1*}^2\}, \quad D_0 = E\{(1 - T)R_0g_{0*}^2\}, \\
\mathbf{A}_1 &= E\left\{\frac{E(Y_1 | \mathbf{Z}, \mathbf{X}_1)}{\varpi_1(\mathbf{Z}, \mathbf{X}_1)} \frac{\partial \varpi_1(\boldsymbol{\alpha}_{1*})}{\partial \boldsymbol{\alpha}_1^T}\right\}, \\
\mathbf{A}_0 &= E\left\{\frac{E(Y_0 | \mathbf{Z}, \mathbf{X}_0)}{\varpi_0(\mathbf{Z}, \mathbf{X}_0)} \frac{\partial \varpi_0(\boldsymbol{\alpha}_{0*})}{\partial \boldsymbol{\alpha}_0^T}\right\},
\end{aligned}$$

and \mathbf{S}_t , $t = 0, 1$ is the score function of the binomial likelihood (3.1).

7.1.2. Expressions for Theorem 3.2

Without loss of generality, suppose $\varpi_1^{(1)}(\boldsymbol{\alpha}_1^{(1)})$ and $\varpi_0^{(1)}(\boldsymbol{\alpha}_0^{(1)})$ are correctly specified models for $\varpi_1(\mathbf{Z}, \mathbf{X}_1)$ and $\varpi_0(\mathbf{Z}, \mathbf{X}_0)$ respectively. Then the scaling constant for the asymptotic distribution in Theorem 3.2 is given by

$$\begin{aligned}
\sigma_2 &= (D_1^{\text{MR}} + D_0^{\text{MR}})^{-1} D_1^{\text{MR}} D_0^{\text{MR}} \\
&\quad \times \text{var} \left\{ \varphi^{\text{MR}} - \mathbf{A}_1^{\text{MR}} E(\mathbf{S}_1^{(1)} \mathbf{S}_1^{(1),T})^{-1} \mathbf{S}_1^{(1)} + \mathbf{A}_0^{\text{MR}} E(\mathbf{S}_0^{(1)} \mathbf{S}_0^{(1),T})^{-1} \mathbf{S}_0^{(1)} \right\},
\end{aligned}$$

where

$$\begin{aligned}
\theta_{1*}^{(j)} &= E\{\pi(\mathbf{Z})\varpi_1^{(j)}(\boldsymbol{\alpha}_1^{(j)})\}, \quad \theta_{0*}^{(j)} = E[\{1 - \pi(\mathbf{Z})\}\varpi_0^{(j)}(\boldsymbol{\alpha}_0^{(j)})], \\
\mathbf{g}_1(\boldsymbol{\alpha}_{1*}) &= \left\{ \pi(\mathbf{Z})\varpi_1^{(1)}(\boldsymbol{\alpha}_{1*}^{(1)}) - \theta_{1*}^{(1)}, \dots, \pi(\mathbf{Z})\varpi_1^{(J_1)}(\boldsymbol{\alpha}_{1*}^{(J_1)}) - \theta_{1*}^{(J_1)} \right\}^T, \\
\mathbf{g}_0(\boldsymbol{\alpha}_{0*}) &= \left[\{1 - \pi(\mathbf{Z})\}\varpi_0^{(1)}(\boldsymbol{\alpha}_{0*}^{(1)}) - \theta_{0*}^{(1)}, \dots, \right. \\
&\quad \left. \{1 - \pi(\mathbf{Z})\}\varpi_0^{(J_0)}(\boldsymbol{\alpha}_{0*}^{(J_0)}) - \theta_{0*}^{(J_0)} \right]^T, \\
\mathbf{h}_t(\boldsymbol{\alpha}_{t*}) &= \{\theta_{t*}^{(1)}(\boldsymbol{\alpha}_{t*}^{(1)}), \dots, \theta_{t*}^{(J_t)}(\boldsymbol{\alpha}_{t*}^{(J_t)})\}^T, \quad t = 0, 1, \\
\mathbf{B}_1 &= E\left[\frac{\{E(Y_1 | \mathbf{Z}, \mathbf{X}_1) - \mu_0 - \delta_*\}\mathbf{g}_1(\boldsymbol{\alpha}_{1*})}{\pi(\mathbf{Z})\varpi_1(\mathbf{Z}, \mathbf{X}_1)}\right], \\
\mathbf{B}_0 &= E\left[\frac{\{E(Y_0 | \mathbf{Z}, \mathbf{X}_0) - \mu_0\}\mathbf{g}_0(\boldsymbol{\alpha}_{0*})}{\{1 - \pi(\mathbf{Z})\}\varpi_0(\mathbf{Z}, \mathbf{X}_0)}\right], \\
\mathbf{G}_1 &= E\left\{\frac{\mathbf{g}_1(\boldsymbol{\alpha}_{1*})\mathbf{g}_1(\boldsymbol{\alpha}_{1*})^T}{\pi(\mathbf{Z})\varpi_1(\mathbf{Z}, \mathbf{X}_1)}\right\}, \quad \mathbf{G}_0 = E\left[\frac{\mathbf{g}_0(\boldsymbol{\alpha}_{0*})\mathbf{g}_0(\boldsymbol{\alpha}_{0*})^T}{\{1 - \pi(\mathbf{Z})\}\varpi_0(\mathbf{Z}, \mathbf{X}_0)}\right], \\
\mathbf{A}_1^{\text{MR}} &= E\left\{\frac{E(Y_1 | \mathbf{Z}, \mathbf{X}_1) - \mu_0 - \delta_* - \mathbf{B}_1^T \mathbf{G}_1^{-1} \mathbf{g}_1(\boldsymbol{\alpha}_{1*})}{\varpi_1(\mathbf{Z}, \mathbf{X}_1)} \frac{\partial \varpi_1^{(1)}(\boldsymbol{\alpha}_{1*}^{(1)})}{\partial \boldsymbol{\alpha}_1^{(1),T}}\right\}, \\
\mathbf{A}_0^{\text{MR}} &= E\left\{\frac{E(Y_0 | \mathbf{Z}, \mathbf{X}_0) - \mu_0 - \mathbf{B}_0^T \mathbf{G}_0^{-1} \mathbf{g}_0(\boldsymbol{\alpha}_{0*})}{\varpi_0(\mathbf{Z}, \mathbf{X}_0)} \frac{\partial \varpi_0^{(1)}(\boldsymbol{\alpha}_{0*}^{(1)})}{\partial \boldsymbol{\alpha}_0^{(1),T}}\right\},
\end{aligned}$$

$$\begin{aligned}
 D_1^{\text{MR}} &= \left[E \left\{ \frac{TR_1(Y_1 - \mu_0 - \delta_*)^2}{\pi(\mathbf{Z})^2 \varpi_1(\mathbf{Z}, \mathbf{X}_1)^2} \right\} - \mathbf{B}_1^T \mathbf{G}_1^{-1} \mathbf{B}_1 \right]^{-1}, \\
 D_0^{\text{MR}} &= \left(E \left[\frac{(1-T)R_0(Y_0 - \mu_0)^2}{\{1 - \pi(\mathbf{Z})\}^2 \varpi_0(\mathbf{Z}, \mathbf{X}_0)^2} \right] - \mathbf{B}_0^T \mathbf{G}_0^{-1} \mathbf{B}_0 \right)^{-1}, \\
 \varphi^{\text{MR}} &= \frac{TR_1 Y_1}{\pi(\mathbf{Z}) \varpi_1(\mathbf{Z}, \mathbf{X}_1)} - \frac{(1-T)R_0 Y_0}{\{1 - \pi(\mathbf{Z})\} \varpi_0(\mathbf{Z}, \mathbf{X}_0)} - \delta_* \\
 &\quad - \mathbf{B}_1^T \mathbf{G}_1^{-1} \left[\frac{T\{R_1 - \varpi_1(\mathbf{Z}, \mathbf{X}_1)\}}{\pi(\mathbf{Z}) \varpi_1(\mathbf{Z}, \mathbf{X}_1)} \mathbf{g}_1(\boldsymbol{\alpha}_{1*}) - \frac{T - \pi(\mathbf{Z})}{\pi(\mathbf{Z})} \mathbf{h}_1(\boldsymbol{\alpha}_{1*}) \right] \\
 &\quad + \frac{(1-T)\{R_0 - \varpi_0(\mathbf{Z}, \mathbf{X}_0)\}}{\{1 - \pi(\mathbf{Z})\} \varpi_0(\mathbf{Z}, \mathbf{X}_0)} \mathbf{B}_0^T \mathbf{G}_0^{-1} \mathbf{g}_0(\boldsymbol{\alpha}_{0*}) \\
 &\quad + \frac{T - \pi(\mathbf{Z})}{1 - \pi(\mathbf{Z})} \mathbf{B}_0^T \mathbf{G}_0^{-1} \mathbf{h}_0(\boldsymbol{\alpha}_{0*}),
 \end{aligned}$$

and $\mathbf{S}_t^{(1)}$, $t = 0, 1$, is the score function of (3.1) with $\varpi_t(\boldsymbol{\alpha}_t)$ replaced by $\varpi_t^{(1)}(\boldsymbol{\alpha}_t^{(1)})$.

7.1.3. Expressions for Theorem 3.3

The scaling constant for the asymptotic distribution in Theorem 3.3 is given by

$$\begin{aligned}
 \sigma_3 &= \text{var} \left\{ \varphi - \mathbf{A}_1 E(\mathbf{S}_1 \mathbf{S}_1^T)^{-1} \mathbf{S}_1 + \mathbf{A}_0 E(\mathbf{S}_0 \mathbf{S}_0^T)^{-1} \mathbf{S}_0 \right. \\
 &\quad \left. - (\mathbf{C}_1 + \mathbf{C}_0) E(\mathbf{S}_\gamma \mathbf{S}_\gamma^T)^{-1} \mathbf{S}_\gamma \right\},
 \end{aligned}$$

where

$$\mathbf{C}_1 = E \left\{ \frac{E(Y_1 | \mathbf{Z}, \mathbf{X}_1)}{\pi(\mathbf{Z})} \frac{\partial \pi(\gamma_*)}{\partial \gamma^T} \right\}, \quad \mathbf{C}_0 = E \left\{ \frac{E(Y_0 | \mathbf{Z}, \mathbf{X}_0)}{1 - \pi(\mathbf{Z})} \frac{\partial \pi(\gamma_*)}{\partial \gamma^T} \right\},$$

and \mathbf{S}_γ is the score function of $\prod_{i=1}^n \{\pi_i(\gamma)\}^{T_i} \{1 - \pi_i(\gamma)\}^{1-T_i}$.

7.1.4. Expressions for Theorem 3.4

Without loss of generality, suppose $\varpi_1^{(1)}(\boldsymbol{\alpha}_1^{(1)})$ and $\varpi_0^{(1)}(\boldsymbol{\alpha}_0^{(1)})$ are correctly specified models for $\varpi_1(\mathbf{Z}, \mathbf{X}_1)$ and $\varpi_0(\mathbf{Z}, \mathbf{X}_0)$ respectively. Then the scaling constant for the asymptotic distribution in Theorem 3.4 is given by

$$\begin{aligned}
 \sigma_4 &= (D_1^{\text{MR}} + D_0^{\text{MR}})^{-1} D_1^{\text{MR}} D_0^{\text{MR}} \text{var} \left\{ \varphi^{\text{MR}} - \mathbf{A}_1^{\text{MR}} E(\mathbf{S}_1^{(1)} \mathbf{S}_1^{(1),T})^{-1} \mathbf{S}_1^{(1)} \right. \\
 &\quad \left. + \mathbf{A}_0^{\text{MR}} E(\mathbf{S}_0^{(1)} \mathbf{S}_0^{(1),T})^{-1} \mathbf{S}_0^{(1)} - (\mathbf{C}_1^{\text{MR}} + \mathbf{C}_0^{\text{MR}}) E(\mathbf{S}_\gamma \mathbf{S}_\gamma^T)^{-1} \mathbf{S}_\gamma \right\},
 \end{aligned}$$

where

$$\mathbf{C}_1^{\text{MR}} = E \left\{ \frac{E(Y_1 | \mathbf{Z}, \mathbf{X}_1) - \mu_0 - \delta_* - \mathbf{B}_1^T \mathbf{G}_1^{-1} \mathbf{g}_1(\boldsymbol{\alpha}_{1*})}{\pi(\mathbf{Z})} \frac{\partial \pi(\gamma_*)}{\partial \gamma^T} \right\},$$

$$\mathbf{C}_0^{\text{MR}} = E \left\{ \frac{E(Y_0 | \mathbf{Z}, \mathbf{X}_0) - \mu_0 - \mathbf{B}_0^{\text{T}} \mathbf{G}_0^{-1} \mathbf{g}_0(\boldsymbol{\alpha}_{0*})}{1 - \pi(\mathbf{Z})} \frac{\partial \pi(\boldsymbol{\gamma}_*)}{\partial \boldsymbol{\gamma}^{\text{T}}} \right\}.$$

7.2. Proofs of the Theorems

7.2.1. Proof of Theorem 3.1

Following a similar argument to that of [27], let $\mu_0^{\text{nu}}i$ be a nuisance parameter such that $\mu_0^{\text{nu}}i = \mu_0 + O_p(n^{-1/2})$. The introduction of this nuisance parameter $\mu_0^{\text{nu}}i$ facilitates the derivation of the asymptotic distribution of the empirical likelihood ratio statistic $W(\delta_*)$ and will later be profiled. Then the constraint (3.4) can be replaced by

$$\begin{aligned} \sum_{i:T_i=1, R_{1i}=1} w_i \left\{ \frac{Y_{1i}}{\pi_i \varpi_{1i}(\hat{\boldsymbol{\alpha}}_1)} - \frac{n}{n_{11}} (\mu_0^{\text{nu}}i + \delta_*) \right\} &= 0, \\ \sum_{i:T_i=0, R_{0i}=1} v_i \left\{ \frac{Y_{0i}}{(1 - \pi_i) \varpi_{0i}(\hat{\boldsymbol{\alpha}}_0)} - \frac{n}{n_{01}} \mu_0^{\text{nu}}i \right\} &= 0. \end{aligned}$$

Maximizing $L(\mathbf{w}, \mathbf{v})$ subject to the normalization constraints (3.3) in addition to these two constraints gives

$$\tilde{w}_i = \frac{1}{n_{11}} \frac{1}{1 + \tilde{\lambda}_1 g_{1i}(\hat{\boldsymbol{\alpha}}_1)}, \quad \tilde{v}_i = \frac{1}{n_{01}} \frac{1}{1 + \tilde{\lambda}_0 g_{0i}(\hat{\boldsymbol{\alpha}}_0)}$$

where $\tilde{\lambda}_1$ and $\tilde{\lambda}_0$ are respectively solutions to

$$\sum_{i:T_i=1, R_{1i}=1} \frac{g_{1i}(\hat{\boldsymbol{\alpha}}_1)}{1 + \tilde{\lambda}_1 g_{1i}(\hat{\boldsymbol{\alpha}}_1)} = 0, \quad \sum_{i:T_i=0, R_{0i}=1} \frac{g_{0i}(\hat{\boldsymbol{\alpha}}_0)}{1 + \tilde{\lambda}_0 g_{0i}(\hat{\boldsymbol{\alpha}}_0)} = 0, \quad (7.1)$$

and

$$\begin{aligned} g_{1i}(\hat{\boldsymbol{\alpha}}_1) &= \frac{Y_{1i}}{\pi_i \varpi_{1i}(\hat{\boldsymbol{\alpha}}_1)} - \frac{n}{n_{11}} (\mu_0^{\text{nu}}i + \delta_*), \\ g_{0i}(\hat{\boldsymbol{\alpha}}_0) &= \frac{Y_{0i}}{(1 - \pi_i) \varpi_{0i}(\hat{\boldsymbol{\alpha}}_0)} - \frac{n}{n_{10}} \mu_0^{\text{nu}}i. \end{aligned}$$

Taylor expansions of (7.1) at $(\lambda_{1*} = 0, \boldsymbol{\alpha}_{1*}^{\text{T}})$ and $(\lambda_{0*} = 0, \boldsymbol{\alpha}_{0*}^{\text{T}})$ yield

$$\begin{aligned} n^{1/2} \tilde{\lambda}_1 &= D_1^{-1} \left\{ n^{-1/2} \sum_{i=1}^n T_i R_{1i} g_{1i}(\boldsymbol{\alpha}_{1*}) - \mathbf{A}_1 n^{1/2} (\hat{\boldsymbol{\alpha}}_1 - \boldsymbol{\alpha}_{1*}) \right\} + o_p(1), \\ n^{1/2} \tilde{\lambda}_0 &= D_0^{-1} \left\{ n^{-1/2} \sum_{i=1}^n (1 - T_i) R_{0i} g_{0i}(\boldsymbol{\alpha}_{0*}) - \mathbf{A}_0 n^{1/2} (\hat{\boldsymbol{\alpha}}_0 - \boldsymbol{\alpha}_{0*}) \right\} + o_p(1). \end{aligned}$$

The empirical likelihood ratio statistic $W(\delta_*)$ is then expressed as

$$W(\mu_0^{\text{nui}}, \delta_*) = 2 \left[\sum_{i:T_i=1, R_{1i}=1} \log\{1 + \tilde{\lambda}_1 g_{1i}(\hat{\alpha}_1)\} + \sum_{i:T_i=0, R_{0i}=1} \log\{1 + \tilde{\lambda}_0 g_{0i}(\hat{\alpha}_0)\} \right]. \tag{7.2}$$

We maximize (7.2) with respect to μ_0^{nui} by setting $\partial W(\mu_0^{\text{nui}}, \delta_*)/\partial \mu_0^{\text{nui}} = 0$, which implies $\tilde{\lambda}_1 + \tilde{\lambda}_0 = 0$. Therefore the maximizer is given by

$$\hat{\mu}_0^{\text{nui}} = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{D_0}{D_0 + D_1} \left(\frac{T_i R_{1i} Y_{1i}}{\pi_i \varpi_{1i}} - \delta_* \right) + \frac{D_1}{D_0 + D_1} \frac{(1 - T_i) R_{0i} Y_{0i}}{(1 - \pi_i) \varpi_{0i}} \right\} - \frac{D_0}{D_0 + D_1} \mathbf{A}_1(\hat{\alpha}_1 - \alpha_{1*}) - \frac{D_1}{D_0 + D_1} \mathbf{A}_0(\hat{\alpha}_0 - \alpha_{0*}) + o_p(n^{-1/2}).$$

A Taylor expansion of (7.2) at $(\lambda_{1*} = 0, \lambda_{0*} = 0, \alpha_{1*}^T, \alpha_{0*}^T)$ with μ_0^{nui} replaced by $\hat{\mu}_0^{\text{nui}}$ gives

$$W(\hat{\mu}_0^{\text{nui}}, \delta_*) = (D_0 + D_1)^{-1} \times \left[n^{-1/2} \sum_{i=1}^n \{ \varphi_i - \mathbf{A}_1 n^{1/2}(\hat{\alpha}_1 - \alpha_{1*}) + \mathbf{A}_0 n^{1/2}(\hat{\alpha}_0 - \alpha_{0*}) \} \right]^2 + o_p(1).$$

Then the desired result follows.

7.2.2. Proof of Theorem 3.2

Similar to the proof of Theorem 3.1, we introduce the nuisance parameter μ_0^{nui} such that $\mu_0^{\text{nui}} = \mu_0 + O_p(n^{-1/2})$. Then the constraint (3.9) can be replaced by $\sum_{i:T_i=1, R_{1i}=1} w_i Y_{1i} = \mu_0^{\text{nui}} + \delta_*$ and $\sum_{i:T_i=0, R_{0i}=1} v_i Y_{0i} = \mu_0^{\text{nui}}$. Maximizing $L(\mathbf{w}, \mathbf{v})$ subject to the constraints (3.3), (3.7), (3.8) and these two constraints yields

$$\tilde{w}_i = \frac{1}{n_{11}} \frac{1}{1 + \check{\rho}_1^T \check{\mathbf{g}}_{1i}(\hat{\alpha}_1)}, \quad \{i : T_i = 1, R_{1i} = 1\},$$

$$\tilde{v}_i = \frac{1}{n_{01}} \frac{1}{1 + \check{\rho}_0^T \check{\mathbf{g}}_{0i}(\hat{\alpha}_0)}, \quad \{i : T_i = 0, R_{0i} = 1\},$$

where $\check{\rho}_1$ and $\check{\rho}_0$ are respectively solutions to

$$\sum_{i:T_i=1, R_{1i}=1} \frac{\check{\mathbf{g}}_{1i}(\hat{\alpha}_1)}{1 + \check{\rho}_1^T \check{\mathbf{g}}_{1i}(\hat{\alpha}_1)} = \mathbf{0}, \quad \sum_{i:T_i=0, R_{0i}=1} \frac{\check{\mathbf{g}}_{0i}(\hat{\alpha}_0)}{1 + \check{\rho}_0^T \check{\mathbf{g}}_{0i}(\hat{\alpha}_0)} = \mathbf{0}$$

and $\check{\mathbf{g}}_{1i}(\hat{\boldsymbol{\alpha}}_1) = \{\hat{\mathbf{g}}_{1i}(\hat{\boldsymbol{\alpha}}_1)^T, Y_{1i} - \mu_0^{\text{nu}} - \delta_*\}^T$, $\check{\mathbf{g}}_{0i}(\hat{\boldsymbol{\alpha}}_0) = \{\hat{\mathbf{g}}_{0i}(\hat{\boldsymbol{\alpha}}_0)^T, Y_{0i} - \mu_0^{\text{nu}}\}^T$. A reparameterization similar to that of [14] gives

$$\begin{aligned}\tilde{w}_i &= \frac{\hat{\theta}_1^{(1)}(\hat{\boldsymbol{\alpha}}_1^{(1)})}{n_{11}} \frac{1}{\pi_i \varpi_{1i}^{(1)}(\hat{\boldsymbol{\alpha}}_1^{(1)}) + \check{\boldsymbol{\lambda}}_1^T \check{\mathbf{g}}_{1i}(\hat{\boldsymbol{\alpha}}_1)}, & \{i : T_i = 1, R_{1i} = 1\}, \\ \tilde{v}_i &= \frac{\hat{\theta}_0^{(1)}(\hat{\boldsymbol{\alpha}}_0^{(1)})}{n_{01}} \frac{1}{(1 - \pi_i) \varpi_{0i}^{(1)}(\hat{\boldsymbol{\alpha}}_0^{(1)}) + \check{\boldsymbol{\lambda}}_0^T \check{\mathbf{g}}_{0i}(\hat{\boldsymbol{\alpha}}_0)}, & \{i : T_i = 0, R_{0i} = 1\},\end{aligned}$$

where $\check{\boldsymbol{\lambda}}_1$ and $\check{\boldsymbol{\lambda}}_0$ are respectively solutions to

$$\begin{aligned}\sum_{i:T_i=1, R_{1i}=1} \frac{\check{\mathbf{g}}_{1i}(\hat{\boldsymbol{\alpha}}_1)}{\pi_i \varpi_{1i}^{(1)}(\hat{\boldsymbol{\alpha}}_1^{(1)}) + \check{\boldsymbol{\lambda}}_1^T \check{\mathbf{g}}_{1i}(\hat{\boldsymbol{\alpha}}_1)} &= \mathbf{0}, \\ \sum_{i:T_i=0, R_{0i}=1} \frac{\check{\mathbf{g}}_{0i}(\hat{\boldsymbol{\alpha}}_0)}{(1 - \pi_i) \varpi_{0i}^{(1)}(\hat{\boldsymbol{\alpha}}_0^{(1)}) + \check{\boldsymbol{\lambda}}_0^T \check{\mathbf{g}}_{0i}(\hat{\boldsymbol{\alpha}}_0)} &= \mathbf{0}.\end{aligned}\tag{7.3}$$

Taylor expansions of (7.3) at $(\boldsymbol{\lambda}_{1*} = \mathbf{0}, \boldsymbol{\lambda}_{0*} = \mathbf{0}, \boldsymbol{\alpha}_{1*}^T, \boldsymbol{\alpha}_{0*}^T)$ yield

$$\begin{aligned}n^{1/2} \check{\boldsymbol{\lambda}}_1 &= \tilde{\mathbf{G}}_1^{-1} \left\{ n^{-1/2} \sum_{i=1}^n \frac{T_i R_{1i} \check{\mathbf{g}}_{1i}(\boldsymbol{\alpha}_{1*})}{\pi_i \varpi_{1i}^{(1)}(\boldsymbol{\alpha}_{1*}^{(1)})} - \tilde{\mathbf{A}}_1 n^{1/2} (\hat{\boldsymbol{\alpha}}_1^{(1)} - \boldsymbol{\alpha}_{1*}^{(1)}) \right\} + o_p(1), \\ n^{1/2} \check{\boldsymbol{\lambda}}_0 &= \tilde{\mathbf{G}}_0^{-1} \left\{ n^{-1/2} \sum_{i=1}^n \frac{(1 - T_i) R_{0i} \check{\mathbf{g}}_{0i}(\boldsymbol{\alpha}_{0*})}{(1 - \pi_i) \varpi_{0i}^{(1)}(\boldsymbol{\alpha}_{0*}^{(1)})} - \tilde{\mathbf{A}}_0 n^{1/2} (\hat{\boldsymbol{\alpha}}_0^{(1)} - \boldsymbol{\alpha}_{0*}^{(1)}) \right\} \\ &\quad + o_p(1)\end{aligned}$$

where

$$\begin{aligned}\tilde{\mathbf{G}}_1 &= \begin{bmatrix} \mathbf{G}_1 & \mathbf{B}_1 \\ \mathbf{B}_1^T & E \left\{ \frac{TR_{1i}(Y_1 - \mu_0 - \delta_*)^2}{\pi(Z)^2 \varpi_1(Z, \mathbf{X}_1)^2} \right\} \end{bmatrix}, \\ \tilde{\mathbf{G}}_0 &= \begin{bmatrix} \mathbf{G}_0 & \mathbf{B}_0 \\ \mathbf{B}_0^T & E \left[\frac{(1-T)R_{0i}(Y_0 - \mu_0)^2}{\{1 - \pi(Z)\}^2 \varpi_0(Z, \mathbf{X}_0)^2} \right] \end{bmatrix}, \\ \tilde{\mathbf{A}}_1 &= E \left[\frac{\{\mathbf{g}_1(\boldsymbol{\alpha}_{1*})^T, E(Y_1 | \mathbf{Z}, \mathbf{X}_1) - \mu_0 - \delta_*\}^T}{\varpi_1(\mathbf{Z}, \mathbf{X}_1)} \frac{\partial \varpi_1^{(1)}(\boldsymbol{\alpha}_{1*}^{(1)})}{\partial \boldsymbol{\alpha}_{1*}^{(1),T}} \right], \\ \tilde{\mathbf{A}}_0 &= E \left[\frac{\{\mathbf{g}_0(\boldsymbol{\alpha}_{0*})^T, E(Y_0 | \mathbf{Z}, \mathbf{X}_0) - \mu_0\}^T}{\varpi_0(\mathbf{Z}, \mathbf{X}_0)} \frac{\partial \varpi_0^{(1)}(\boldsymbol{\alpha}_{0*}^{(1)})}{\partial \boldsymbol{\alpha}_{0*}^{(1),T}} \right].\end{aligned}$$

We maximize $W(\mu_0^{\text{nu}}, \delta_*)$ with respect to μ_0^{nu} by setting $\partial W(\mu_0^{\text{nu}}, \delta_*) / \partial \mu_0^{\text{nu}} = 0$, which gives

$$\frac{n_{11}}{\hat{\theta}_1^1(\hat{\boldsymbol{\alpha}}_1^1)} \check{\boldsymbol{\lambda}}_{1, J_1 + K_1 + 1} + \frac{n_{01}}{\hat{\theta}_0^1(\hat{\boldsymbol{\alpha}}_0^1)} \check{\boldsymbol{\lambda}}_{0, J_0 + K_0 + 1} = 0$$

where $\check{\lambda}_{1, J_1+K_1+1}$ and $\check{\lambda}_{0, J_0+K_0+1}$ are the last component of $\check{\lambda}_1$ and $\check{\lambda}_0$ respectively. Some calculations show that the maximizer is

$$\begin{aligned} \hat{\mu}_0^{\text{nu}} &= \frac{1}{n} \sum_{i=1}^n \left[\frac{D_1^{\text{MR}}}{D_0^{\text{MR}} + D_1^{\text{MR}}} \frac{T_i R_{1i}}{\pi_i \varpi_{1i}} \{Y_{1i} - \delta_* - \mathbf{B}_1^{\text{T}} \mathbf{G}_1^{-1} \hat{\mathbf{g}}_{1i}(\boldsymbol{\alpha}_{1*})\} \right. \\ &\quad \left. + \frac{D_0^{\text{MR}}}{D_0^{\text{MR}} + D_1^{\text{MR}}} \frac{(1 - T_i) R_{0i}}{(1 - \pi_i) \varpi_{0i}} \{Y_{0i} - \mathbf{B}_0^{\text{T}} \mathbf{G}_0^{-1} \hat{\mathbf{g}}_{0i}(\boldsymbol{\alpha}_{0*})\} \right] \\ &\quad + \frac{D_1^{\text{MR}}}{D_0^{\text{MR}} + D_1^{\text{MR}}} \mathbf{A}_1^{\text{MR}} (\hat{\boldsymbol{\alpha}}_1^{(1)} - \boldsymbol{\alpha}_{1*}^{(1)}) \\ &\quad + \frac{D_0^{\text{MR}}}{D_0^{\text{MR}} + D_1^{\text{MR}}} \mathbf{A}_0^{\text{MR}} (\hat{\boldsymbol{\alpha}}_0^{(1)} - \boldsymbol{\alpha}_{0*}^{(1)}) + o_p(n^{-1/2}). \end{aligned}$$

A Taylor expansions of $W(\hat{\mu}_0^{\text{nu}}, \delta_*)$ at $(\boldsymbol{\lambda}_{1*} = \mathbf{0}, \boldsymbol{\lambda}_{0*} = \mathbf{0}, \boldsymbol{\alpha}_{1*}^{\text{T}}, \boldsymbol{\alpha}_{0*}^{\text{T}})$ yields

$$\begin{aligned} W(\hat{\mu}_0^{\text{nu}}, \delta_*) &= \frac{D_1^{\text{MR}} D_0^{\text{MR}}}{D_0^{\text{MR}} + D_1^{\text{MR}}} \left[n^{-1/2} \sum_{i=1}^n \left\{ \varphi_i - \frac{T_i R_{1i}}{\pi_i \varpi_{1i}} \mathbf{B}_1^{\text{T}} \mathbf{G}_1^{-1} \hat{\mathbf{g}}_{1i}(\boldsymbol{\alpha}_{1*}) \right. \right. \\ &\quad \left. \left. + \frac{(1 - T_i) R_{0i}}{(1 - \pi_i) \varpi_{0i}} \mathbf{B}_0^{\text{T}} \mathbf{G}_0^{-1} \hat{\mathbf{g}}_{0i}(\boldsymbol{\alpha}_{0*}) \right. \right. \\ &\quad \left. \left. - \mathbf{A}_1^{\text{MR}} n^{1/2} (\hat{\boldsymbol{\alpha}}_1^{(1)} - \boldsymbol{\alpha}_{1*}^{(1)}) + \mathbf{A}_0^{\text{MR}} n^{1/2} (\hat{\boldsymbol{\alpha}}_0^{(1)} - \boldsymbol{\alpha}_{0*}^{(1)}) \right\} \right]^2 + o_p(1). \end{aligned}$$

Then the desired result follows by noticing that

$$\begin{aligned} n^{-1/2} \sum_{i=1}^n \frac{T_i R_{1i}}{\pi_i \varpi_{1i}} \hat{\mathbf{g}}_{1i}(\boldsymbol{\alpha}_{1*}) &= n^{-1/2} \sum_{i=1}^n \left\{ \frac{T_i (R_{1i} - \varpi_{1i})}{\pi_i \varpi_{1i}} \mathbf{g}_{1i}(\boldsymbol{\alpha}_{1*}) \right. \\ &\quad \left. - \frac{T_i - \pi_i}{\pi_i} \mathbf{h}_1(\boldsymbol{\alpha}_{1*}) \right\}, \\ n^{-1/2} \sum_{i=1}^n \frac{(1 - T_i) R_{0i}}{(1 - \pi_i) \varpi_{0i}} \hat{\mathbf{g}}_{0i}(\boldsymbol{\alpha}_{0*}) &= n^{-1/2} \sum_{i=1}^n \left\{ \frac{(1 - T_i) (R_{0i} - \varpi_{0i})}{(1 - \pi_i) \varpi_{0i}} \mathbf{g}_{0i}(\boldsymbol{\alpha}_{0*}) \right. \\ &\quad \left. + \frac{T_i - \pi_i}{1 - \pi_i} \mathbf{h}_0(\boldsymbol{\alpha}_{0*}) \right\}. \end{aligned}$$

7.2.3. Proof of Theorem 4.1

Here and after we will only show the consistency of $\hat{\mu}_{1\text{MR}} = \sum_{i: T_i=1, R_{1i}=1} \hat{w}_i Y_{1i}$ for the treatment group mean. A similar argument will then give the consistency for the control group mean. We assume $\pi(\mathbf{Z})$ is known and omit the proof for the case when $\pi(\mathbf{Z})$ is unknown but is correctly modeled by $\pi(\mathbf{Z}; \boldsymbol{\gamma})$ due to similarity. Without loss of generality, let $\varpi_1^{(1)}(\mathbf{Z}, \mathbf{X}_1; \boldsymbol{\alpha}_1^{(1)})$ be the correctly specified

model for $\varpi_1(\mathbf{Z}, \mathbf{X}_1)$. Let $\alpha_{1*}^{(j)}$ and $\beta_{1*}^{(k)}$ be the probability limits of $\hat{\alpha}_1^{(j)}$ and $\hat{\beta}_1^{(k)}$ respectively. Denote $(\alpha_{1*}, \beta_{1*}) = (\alpha_{1*}^{(1),T}, \dots, \alpha_{1*}^{(J_1),T}, \beta_{1*}^{(1),T}, \dots, \beta_{1*}^{(K_1),T})^T$ and d_j and u_k the dimensions of $\alpha_{1*}^{(j)}$ and $\beta_{1*}^{(k)}$ respectively. Similar to [14], we can reparameterize $\hat{\rho}_1 = (\hat{\rho}_{11}, \dots, \hat{\rho}_{1, J_1+K_1})$ as $\hat{\lambda}_1 = (\hat{\lambda}_{11}, \dots, \hat{\lambda}_{1, J_1+K_1})$ such that $\hat{\rho}_{11} = (\hat{\lambda}_{11} + 1)/\hat{\theta}_1^{(1)}(\hat{\alpha}_1^{(1)})$ and $\hat{\rho}_{1j} = \hat{\lambda}_{1j}/\hat{\theta}_1^{(1)}(\hat{\alpha}_1^{(1)})$, $j = 2, \dots, J_1 + K_1$, where $\hat{\lambda}_1$ satisfies

$$\sum_{i: T_i=1, R_{1i}=1} \frac{\hat{g}_{1i}(\hat{\alpha}_1, \hat{\beta}_1)}{\pi_i \varpi_{1i}^{(1)}(\hat{\alpha}_1^{(1)}) + \hat{\lambda}_1^T \hat{g}_{1i}(\hat{\alpha}_1, \hat{\beta}_1)} = \mathbf{0}, \tag{7.4}$$

and

$$\hat{g}_{1i}(\hat{\alpha}_1, \hat{\beta}_1) = \left\{ \pi_i \varpi_{1i}^{(1)}(\hat{\alpha}_1^{(1)}) - \hat{\theta}_1^{(1)}(\hat{\alpha}_1^{(1)}), \dots, \pi_i \varpi_{1i}^{(J_1)}(\hat{\alpha}_1^{(J_1)}) - \hat{\theta}_1^{(J_1)}(\hat{\alpha}_1^{(J_1)}) \right. \\ \left. a_{1i}^{(1)}(\hat{\beta}_1^{(1)}) - \hat{\eta}_1^{(1)}(\hat{\beta}_1^{(1)}), \dots, a_{1i}^{(K_1)}(\hat{\beta}_1^{(K_1)}) - \hat{\eta}_1^{(K_1)}(\hat{\beta}_1^{(K_1)}) \right\}^T.$$

Thus we have

$$\hat{w}_i = \frac{\hat{\theta}_1^{(1)}(\hat{\alpha}_1^{(1)})}{n_{11}} \frac{1}{\pi_i \varpi_{1i}^{(1)}(\hat{\alpha}_1^{(1)}) + \hat{\lambda}_1^T \hat{g}_{1i}(\hat{\alpha}_1, \hat{\beta}_1)}.$$

A Taylor expansion of (7.4) at $(\lambda_{1*}^T = \mathbf{0}, \alpha_{1*}^T, \beta_{1*}^T)$ yields

$$\begin{aligned} \mathbf{0} &= n^{-1/2} \sum_{i=1}^n \frac{T_i R_{1i} \hat{g}_{1i}(\hat{\alpha}_1, \hat{\beta}_1)}{\pi_i \varpi_{1i}^{(1)}(\hat{\alpha}_1^{(1)}) + \hat{\lambda}_1^T \hat{g}_{1i}(\hat{\alpha}_1, \hat{\beta}_1)} \\ &= n^{-1/2} \sum_{i=1}^n \frac{T_i R_{1i} \hat{g}_{1i}(\alpha_{1*}, \beta_{1*})}{\pi_i \varpi_{1i}^{(1)}(\hat{\alpha}_1^{(1)})} \\ &\quad - \left[\frac{1}{n} \sum_{i=1}^n \frac{T_i R_{1i} \hat{g}_{1i}(\alpha_{1*}, \beta_{1*}) \hat{g}_{1i}(\alpha_{1*}, \beta_{1*})^T}{\{\pi_i \varpi_{1i}^{(1)}(\hat{\alpha}_1^{(1)})\}^2} \right] n^{1/2} \hat{\lambda}_1 \\ &\quad + \left(\frac{1}{n} \sum_{i=1}^n \frac{T_i R_{1i}}{\{\pi_i \varpi_{1i}^{(1)}(\alpha_{1*}^{(1)})\}^2} \left[\left\{ \begin{array}{c} \pi_i \frac{\partial \varpi_{1i}^{(1)}(\alpha_{1*}^{(1)})}{\partial \alpha_{1*}^{(1),T}} - \frac{1}{n} \sum_{h=1}^n T_h \frac{\partial \varpi_{1h}^{(1)}(\alpha_{1*}^{(1)})}{\partial \alpha_{1*}^{(1),T}} \\ \mathbf{0}_{d_1(J_1+K_1-1)} \end{array} \right\} \right. \right. \\ &\quad \left. \left. \times \pi_i \varpi_{1i}^{(1)}(\alpha_{1*}^{(1)}) - \hat{g}_{1i}(\alpha_{1*}, \beta_{1*}) \pi_i \frac{\partial \varpi_{1i}^{(1)}(\alpha_{1*}^{(1)})}{\partial \alpha_{1*}^{(1),T}} \right] \right) n^{1/2} (\hat{\alpha}_1^{(1)} - \alpha_{1*}^{(1)}) \\ &\quad + \sum_{j=2}^{J_1} \left[\frac{1}{n} \sum_{i=1}^n \frac{T_i R_{1i}}{\pi_i \varpi_{1i}^{(1)}(\alpha_{1*}^{(1)})} \left\{ \begin{array}{c} \mathbf{0}_{d_j(j-1)} \\ \pi_i \frac{\partial \varpi_{1i}^{(j)}(\alpha_{1*}^{(j)})}{\partial \alpha_{1*}^{(j),T}} - \frac{1}{n} \sum_{h=1}^n T_h \frac{\partial \varpi_{1h}^{(j)}(\alpha_{1*}^{(j)})}{\partial \alpha_{1*}^{(j),T}} \\ \mathbf{0}_{d_j(J_1+K_1-j)} \end{array} \right\} \right] \\ &\quad \times n^{1/2} (\hat{\alpha}_1^{(j)} - \alpha_{1*}^{(j)}) \end{aligned}$$

$$\begin{aligned}
 & + \sum_{k=1}^{K_1} \left[\frac{1}{n} \sum_{i=1}^n \frac{T_i R_{1i}}{\pi_i \varpi_{1i}^{(1)}(\boldsymbol{\alpha}_{1*}^{(1)})} \left\{ \begin{array}{c} \mathbf{0}_{u_k(J_1+k-1)} \\ \frac{\partial a_{1i}^{(k)}(\boldsymbol{\beta}_{1*}^{(k)})}{\partial \boldsymbol{\beta}_{1*}^{(k),T}} - \frac{1}{n} \sum_{h=1}^n T_h \frac{\partial a_{1h}^{(k)}(\boldsymbol{\beta}_{1*}^{(k)})}{\partial \boldsymbol{\beta}_{1*}^{(k),T}} \\ \mathbf{0}_{u_k(K_1-k)} \end{array} \right\} \right] \\
 & \quad \times n^{1/2}(\hat{\boldsymbol{\beta}}_1^{(k)} - \boldsymbol{\beta}_{1*}^{(k)}) \\
 & + o_p(1).
 \end{aligned}$$

Solving for $n^{1/2}\hat{\boldsymbol{\lambda}}_1$ implies

$$\begin{aligned}
 n^{1/2}\hat{\boldsymbol{\lambda}}_1 & = \mathbf{G}_1^{-1} \left\{ n^{-1/2} \sum_{i=1}^n \frac{T_i R_{1i} \hat{\mathbf{g}}_{1i}(\hat{\boldsymbol{\alpha}}_1, \hat{\boldsymbol{\beta}}_1)}{\pi_i \varpi_{1i}} - \mathbf{A}_2 n^{1/2}(\hat{\boldsymbol{\alpha}}_1^{(1)} - \boldsymbol{\alpha}_{1*}^{(1)}) \right\} + o_p(1) \\
 & = \mathbf{G}_1^{-1} \left\{ n^{-1/2} \sum_{i=1}^n \frac{T_i(R_{1i} - \varpi_{1i})}{\pi_i \varpi_{1i}} \mathbf{g}_{1i}(\boldsymbol{\alpha}_{1*}, \boldsymbol{\beta}_{1*}) \right. \\
 & \quad \left. - \frac{T_i - \pi_i}{\pi_i} \mathbf{h}_1(\boldsymbol{\alpha}_{1*}, \boldsymbol{\beta}_{1*}) - \mathbf{A}_2 n^{1/2}(\hat{\boldsymbol{\alpha}}_1^{(1)} - \boldsymbol{\alpha}_{1*}^{(1)}) \right\} + o_p(1).
 \end{aligned}$$

where

$$\mathbf{A}_2 = E \left\{ \frac{\mathbf{g}_1(\boldsymbol{\alpha}_{1*}, \boldsymbol{\beta}_{1*}) \partial \varpi_1^{(1)}(\boldsymbol{\alpha}_{1*}^{(1)})}{\varpi_1(\mathbf{Z}, \mathbf{X}_1) \partial \boldsymbol{\alpha}_1^{(1),T}} \right\}$$

and

$$\begin{aligned}
 \eta_{t*}^{(k)}(\boldsymbol{\beta}_{t*}^{(k)}) & = E\{a_t^{(k)}(\boldsymbol{\beta}_{t*}^{(k)})\}, \quad t = 0, 1, \\
 \mathbf{g}_1(\boldsymbol{\alpha}_{1*}, \boldsymbol{\beta}_{1*}) & = \left\{ \pi \varpi_1^{(1)}(\boldsymbol{\alpha}_{1*}^{(1)}) - \theta_{1*}^{(1)}, \dots, \pi \varpi_1^{(J_1)}(\boldsymbol{\alpha}_{1*}^{(J_1)}) - \theta_{1*}^{(J_1)} \right. \\
 & \quad \left. a_1^{(1)}(\boldsymbol{\beta}_{1*}^{(1)}) - \eta_{1*}^{(1)}, \dots, a_1^{(K_1)}(\boldsymbol{\beta}_{1*}^{(K_1)}) - \eta_{1*}^{(K_1)} \right\}^T, \\
 \mathbf{h}_t(\boldsymbol{\alpha}_{t*}, \boldsymbol{\beta}_{t*}) & = \left\{ \theta_{t*}^{(1)}(\boldsymbol{\alpha}_{t*}^{(1)}), \dots, \theta_{t*}^{(J_t)}(\boldsymbol{\alpha}_{t*}^{(J_t)}), \eta_{t*}^{(1)}(\boldsymbol{\beta}_{t*}^{(1)}), \dots, \eta_{t*}^{(K_t)}(\boldsymbol{\beta}_{t*}^{(K_t)}) \right\}^T.
 \end{aligned}$$

Thus $\hat{\boldsymbol{\lambda}} = O_p(n^{-1/2}) \rightarrow \mathbf{0}$ in probability. Note that $n_{11}/n \rightarrow \theta_{1*}^{(1)} = \mathbb{P}(T = 1, R_1 = 1)$ and thus we have

$$\begin{aligned}
 \hat{\mu}_{1MR} & = \sum_{i:T_i=1, R_{1i}=1} \hat{w}_i Y_{1i} \\
 & = \frac{\hat{\theta}_1^{(1)}}{n_{11}} \sum_{i=1}^n \frac{T_i R_{1i} Y_{1i}}{\pi_i \varpi_{1i}^{(1)}(\hat{\boldsymbol{\alpha}}_1^{(1)}) + \hat{\boldsymbol{\lambda}}_1^T \hat{\mathbf{g}}_{1i}(\hat{\boldsymbol{\alpha}}_1, \hat{\boldsymbol{\beta}}_1)} \\
 & \rightarrow E \left\{ \frac{TR_1 Y_1}{\pi(\mathbf{Z}) \varpi_1(\mathbf{Z}, \mathbf{X}_1)} \right\} = \mu_1
 \end{aligned}$$

in probability. Therefore $\hat{\mu}_{1MR}$ is a consistent estimator of μ_1 when one of the multiple working models in \mathcal{P}_1 is correctly specified. Now suppose one of the

models in \mathcal{A}_1 is correctly specified for $E(Y_1 | \mathbf{Z}, \mathbf{X}_1)$, say $a_1^{(1)}(\boldsymbol{\beta}_1^{(1)})$. Let $\boldsymbol{\rho}_{1*}$ be the probability limit of $\hat{\boldsymbol{\rho}}_1$, then we have

$$\begin{aligned} \hat{\mu}_{1\text{MR}} &= \sum_{i:T_i=1, R_{1i}=1} \hat{w}_i Y_{1i} \\ &= \sum_{i:T_i=1, R_{1i}=1} \hat{w}_i \{Y_{1i} - a_{1i}^{(1)}(\hat{\boldsymbol{\beta}}_1^{(1)})\} + \frac{1}{n} \sum_{i:T_i=1} \frac{a_{1i}^{(1)}(\hat{\boldsymbol{\beta}}_1^{(1)})}{\pi_i} \\ &= \frac{1}{n_{11}} \sum_{i=1}^n \frac{T_i R_{1i} \{Y_{1i} - a_{1i}^{(1)}(\hat{\boldsymbol{\beta}}_1^{(1)})\}}{1 + \hat{\boldsymbol{\rho}}_1^T \hat{\mathbf{g}}_{1i}(\hat{\boldsymbol{\alpha}}_1, \hat{\boldsymbol{\beta}}_1)} + \frac{1}{n} \sum_{i:T_i=1} \frac{a_{1i}^{(1)}(\hat{\boldsymbol{\beta}}_1^{(1)})}{\pi_i} \\ &\rightarrow \frac{1}{\theta_{1*}^{(1)}} E \left[\frac{TR_1 \{Y_1 - a_1^{(1)}(\mathbf{Z}, \mathbf{X}_1; \boldsymbol{\beta}_{1*}^{(1)})\}}{1 + \boldsymbol{\rho}_{1*}^T \mathbf{g}_1(\boldsymbol{\alpha}_{1*}, \boldsymbol{\beta}_{1*})} \right] + E \{a_1^{(1)}(\boldsymbol{\beta}_{1*}^{(1)})\} \\ &= 0 + \mu_1 = \mu_1 \end{aligned}$$

in probability. Therefore $\hat{\mu}_{1\text{MR}}$ is a consistent estimator of μ_1 when one of the models in \mathcal{A}_1 is correctly specified.

7.2.4. Proof of Theorem 4.2

We assume, without loss of generality, $\pi^{(1)}(\boldsymbol{\gamma}^{(1)})$ is correctly specified for $\pi(\mathbf{Z})$. [14] showed that $(n\hat{p}_i)^{-1} \rightarrow \pi_i^{(1)}(\boldsymbol{\gamma}_*^{(1)}) = \pi(\mathbf{Z}_i)$ in probability. Suppose \mathcal{P}_1 contains a correctly specified model for $\varpi_1(\mathbf{Z}, \mathbf{X}_1)$, say $\varpi_1^{(1)}(\mathbf{Z}, \mathbf{X}_1; \boldsymbol{\alpha}_1^{(1)})$. A reparametrization yields

$$\hat{w}_i = \frac{\hat{\theta}_1^{(1)}}{n_{11}} \frac{1}{(n\hat{p}_i)^{-1} \varpi_{1i}^{(1)}(\hat{\boldsymbol{\alpha}}_1^{(1)}) + \hat{\boldsymbol{\lambda}}_1^T \hat{\mathbf{g}}_{1i}(\hat{\boldsymbol{\alpha}}_1, \hat{\boldsymbol{\beta}}_1)}$$

where $\hat{\boldsymbol{\lambda}}_1$ satisfies

$$\sum_{i:T_i=1, R_{1i}=1} \frac{\hat{\mathbf{g}}_{1i}(\hat{\boldsymbol{\alpha}}_1, \hat{\boldsymbol{\beta}}_1)}{(n\hat{p}_i)^{-1} \varpi_{1i}^{(1)}(\hat{\boldsymbol{\alpha}}_1^{(1)}) + \hat{\boldsymbol{\lambda}}_1^T \hat{\mathbf{g}}_{1i}(\hat{\boldsymbol{\alpha}}_1, \hat{\boldsymbol{\beta}}_1)} = \mathbf{0}.$$

Empirical likelihood theory (Qin and Lawless 19) gives $\hat{\boldsymbol{\lambda}}_1 = O_p(n^{-1/2}) \rightarrow \mathbf{0}$ in probability. Thus

$$\begin{aligned} \hat{\mu}_{1\text{MR}} &= \sum_{i:T_i=1, R_{1i}=1} \hat{w}_i Y_{1i} \\ &= \frac{\hat{\theta}_1^{(1)}}{n_{11}} \sum_{i=1}^n \frac{T_i R_{1i} Y_{1i}}{(n\hat{p}_i)^{-1} \varpi_{1i}^{(1)}(\hat{\boldsymbol{\alpha}}_1^{(1)}) + \hat{\boldsymbol{\lambda}}_1^T \hat{\mathbf{g}}_{1i}(\hat{\boldsymbol{\alpha}}_1, \hat{\boldsymbol{\beta}}_1)} \\ &\rightarrow E \left\{ \frac{TR_1 Y_1}{\pi(\mathbf{Z}) \varpi_1(\mathbf{Z}, \mathbf{X}_1)} \right\} = \mu_1 \end{aligned}$$

in probability.

Now suppose \mathcal{A}_1 contains a correctly specified model $a_1^{(1)}(\beta_1^{(1)})$ for $E(Y_1 | \mathbf{Z}, \mathbf{X}_1)$. A similar argument to that in the proof of Theorem 4.1 gives the consistency of $\hat{\mu}_{1MR}$ by noticing that $\sum_{i:T_i=1} \hat{p}_i a_{1i}^{(1)}(\hat{\beta}_1^{(1)}) \rightarrow E\{a_1^{(1)}(\beta_{1*}^{(1)})\} = \mu_1$ in probability.

References

- [1] Bonate, P. L. (2000). *Analysis of Pretest-Posttest Designs*. Chapman & Hall/CRC.
- [2] Brogan, D. and Kutner, M. (1980). Comparative analyses of pretest-posttest research designs. *The American Statistician*, 34(4):229–232.
- [3] Chan, K. C. G. and Yam, S. C. P. (2014). Oracle, multiple robust and multipurpose calibration in a missing response problem. *Statistical Science*, 29(3):380–396. [MR3264551](#)
- [4] Chen, J., Sitter, R. R., and Wu, C. (2002). Using empirical likelihood methods to obtain range restricted weights in regression estimators for surveys. *Biometrika*, 89(1):230–237. [MR1888365](#)
- [5] Chen, M., Wu, C., and Thompson, M. (2015). An imputation based empirical likelihood approach to pretest-posttest studies. *The Canadian Journal of Statistics*, 43(3):378–402. [MR3388323](#)
- [6] Chen, M., Wu, C., and Thompson, M. (2016). Mann-Whitney test with empirical likelihood methods for pretest-posttest studies. *Journal of Non-parametric Statistics*, 28(2):360–374. [MR3488603](#)
- [7] Chen, S. and Haziza, D. (2017). Multiply robust imputation procedures for the treatment of item nonresponse in surveys. *Biometrika*, 104(2):439–453. [MR3698264](#)
- [8] Davidian, M., Tsiatis, A., and Leon, S. (2005). Semiparametric estimation of treatment effect in a pretest-posttest study with missing data. *Statistical Science*, 20(3):261–301. [MR2189002](#)
- [9] Follmann, D. (1991). The effect of screening on some pretest-posttest test variances. *Biometrics*, 47(2):763–771.
- [10] Han, P. (2014a). A further study of the multiply robust estimator in missing data analysis. *Journal of Statistical Planning and Inference*, 148:101–110. [MR3174151](#)
- [11] Han, P. (2014b). Multiply robust estimation in regression analysis with missing data. *Journal of the American Statistical Association*, 109(507):1159–1173. [MR3265688](#)
- [12] Han, P. (2016a). Combining inverse probability weighting and multiple imputation to improve robustness of estimation. *Scandinavian Journal of Statistics*, 43:246–260. [MR3467005](#)
- [13] Han, P. (2016b). Intrinsic efficiency and multiple robustness in longitudinal studies with dropout. *Biometrika*, 103:683–700. [MR3551792](#)
- [14] Han, P. and Wang, L. (2013). Estimation with missing data: beyond double robustness. *Biometrika*, 100(2):417–430. [MR3068443](#)
- [15] Huang, C.-Y., Qin, J., and Follmann, D. (2008). Empirical likelihood-based

- estimation of the treatment effect in a pretest-posttest study. *Journal of the American Statistical Association*, 103(483):1270–1280. [MR2462898](#)
- [16] Laird, N. (1983). Further comparative analyses of pretest-posttest research designs. *The American Statistician*, 37(4a):329–330.
- [17] Leon, S., Tsiatis, A., and Davidian, M. (2003). Semiparametric estimation of treatment effect in a pretest-posttest study. *Biometrics*, 59(4):1046–1055. [MR2025129](#)
- [18] Owen, A. (2001). *Empirical likelihood*. Chapman & Hall/CRC Press, New York.
- [19] Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *Annals of Statistics*, 22(1):300–325. [MR1272085](#)
- [20] Qin, J. and Zhang, B. (2007). Empirical-likelihood-based inference in missing response problems and its application in observational studies. *Journal of the Royal Statistical Society: Series B*, 69(1):101–122. [MR2301502](#)
- [21] Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89(427):846–866. [MR1294730](#)
- [22] Rosenbaum, P. and Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55. [MR0742974](#)
- [23] Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3):581–592. [MR0455196](#)
- [24] Singer, J. M. and Andrade, D. F. (1997). Regression models for the analysis of pretest-posttest data. *Biometrics*, 53:729–735.
- [25] Stanek, E. (1988). Choosing a pretest-posttest analysis. *The American Statistician*, 42(3):178–183.
- [26] Tsiatis, A. (2006). *Semiparametric theory and missing data*. Springer, New York, 1 edition. [MR2233926](#)
- [27] Wu, C. and Yan, Y. (2012). Empirical likelihood inference for two-sample problems. *Statistics and Its Interface*, 5:345–354. [MR2979214](#)
- [28] Yang, L. and Tsiatis, A. (2001). Efficiency study of estimators for a treatment effect in a pretest-posttest trial. *The American Statistician*, 55(4):314–321. [MR1943328](#)