Stable inverse probability weighting estimation for longitudinal studies

Vahe Avagyan[1] | Stijn Vansteelandt[2,3]

[1]Biometris, Mathematical and Statistical Methods Group, Wageningen University and Research, Wageningen, The Netherlands

[2]Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Ghent, Belgium

[3]Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, UK

Correspondence
Vahe Avagyan, Mathematical and Statistical Methods Group, Wageningen University and Research, Droevendaalsesteeg 1, 6708 PB, Wageningen, The Netherlands.
Email: Vahe.Avagyan@wur.nl

Abstract

We consider estimation of the average effect of time-varying dichotomous exposure on outcome using inverse probability weighting (IPW) under the assumption that there is no unmeasured confounding of the exposure–outcome association at each time point. Despite the popularity of IPW, its performance is often poor due to instability of the estimated weights. We develop an estimating equation-based strategy for the nuisance parameters indexing the weights at each time point, aimed at preventing highly volatile weights and ensuring the stability of IPW estimation. Our proposed approach targets the estimation of the counterfactual mean under a chosen treatment regime and requires fitting a separate propensity score model at each time point. We discuss and examine extensions to enable the fitting of marginal structural models using one propensity score model across all time points. Extensive simulation studies demonstrate adequate performance of our approach compared with the maximum likelihood propensity score estimator and the covariate balancing propensity score estimator.

KEYWORDS

calibration estimation, covariate balancing, inverse probability weighting, propensity score

---

1 | INTRODUCTION

Inverse probability weighting (IPW) has become very popular as a method to adjust statistical analyses for bias due to confounding or selection in cross-sectional and longitudinal observational studies. It is widely used in causal inference (Hernán & Robins, 2006), incomplete data problems (Seaman & White, 2013), sample surveys (Kott & Liao, 2012), among others. The reason for the increasing popularity of IPW is that it is relatively simple and can handle complex problems of confounding and selection bias. Moreover, IPW does not require a lot of modeling. In particular, this method demands a propensity score model, which is defined as the conditional probability of a subject being exposed (or assigned to certain treatment) given a set of pretreatment covariates (Rosenbaum & Rubin, 1983). Correction for confounding bias or selection bias works by weighting each subject differently in the analysis, that is, by assigning a weight equal to the inverse of the conditional probability of the observed exposure. This sole reliance on a propensity score model is useful as it helps to ensure robustness against outcome model misspecification, which is difficult to diagnose when the exposed and unexposed subjects have fairly different covariate or exposure history.

Unfortunately, despite its widespread popularity, inverse weighting has several important drawbacks. This method can be very sensitive to the choice of estimation method for the inverse probability weights (Kang & Schafer, 2007; Ridgeway & McCaffrey, 2007; Zubizarreta, 2015). These are commonly estimated based on maximum likelihood estimation under standard logistic models. The resulting estimated weights can be highly volatile and unstable, which can substantially increase the finite sample bias and variance of the weighted estimators leading to unstable results (Robins et al., 2007; Seaman & White, 2013). For instance, this could happen when for some treated subjects the estimated probability of receiving treatment is low based on their observed covariates. This can happen because maximum likelihood estimators do not guarantee stability of the estimated weights. Obtaining stable inverse weights can be especially problematic in longitudinal studies, where inverse probability weights are often accumulated as a product of inverse weights across time.

A number of methods have been proposed to attain stable weights in fixed time period studies (see Hainmueller, 2012; Imai & Ratkovic, 2014; Yiu & Su, 2018; Zubizarreta, 2015, among others). Relatively few have been designed for longitudinal studies (e.g., Han, 2016; Kallus & Santacatterina, 2018). In view of this, we introduce a new approach to estimate the inverse probability weights needed for estimation of the effects of a time-varying exposure on an outcome. We propose a system of estimating equations in the fixed time period settings and further extend it to the longitudinal settings. Our proposal relies on so-called calibration equations which endorse covariate balancing conditions between the whole sample and the subsample corresponding to the considered regime. A key feature of the proposed approach, unlike others, is that it adapts to the estimand. Through extensive simulation results, we demonstrate that, under a wide variety of settings, our proposed methodology performs well in terms of several statistical measures compared with the traditional maximum likelihood propensity score estimator (MLE), the covariate balancing propensity score (CBPS) estimator (Imai & Ratkovic, 2014), the stable balancing weights (SBW) estimator (only for the fixed time period setting) (Zubizarreta, 2015) and the adaptation of the calibration approach proposed by Han (2016).

The rest of the article is organized as follows. Section 2 describes the IPW estimation method. In Section 3, we provide our proposed methodology. Section 4 gives a thorough literature review where we discuss the limitations of the existing methods and their relation to our proposed method. In Section 5, we conduct a simulation study to evaluate the performance of our proposal.

In Section 6, we demonstrate the applicability of our proposed approach in the estimation of marginal structural models (MSM). In Section 7, we present an empirical application which aims at estimating the impact of negative advertisement on election outcomes. Section 8 concludes with discussions.

2 | IPW ESTIMATION

We start with the problem of estimating the mean outcome under a certain point treatment regime. For each subject $i = 1, \ldots, n$ we denote R_i as the treatment variable, X_i as the p-dimensional vector of observed pretreatment covariates, and Y_i as the outcome of interest. For simplicity we assume that the treatment variable is binary, that is, $R_i = 1$ if subject i received treatment and $R_i = 0$ if subject i received control. Let Y^r denote the potential outcome under treatment r. It is assumed that there is no unmeasured confounding (Rosenbaum & Rubin, 1983), that is, the potential outcome is conditionally independent of the treatment indicator given the covariates ($Y^r \perp\!\!\!\perp R \,|\, X$ for $r \in \{0; 1\}$). No other restrictions are imposed on the outcome and the covariates. Let the propensity score model be a parametric working model for the probability of receiving treatment: $P(R = 1 \,|\, X) = \pi(\gamma; X)$, where we assume that $\pi()$ is known, bounded away from zero and one with probability one (i.e., positivity assumption), and sufficiently smooth in the unknown nuisance parameter γ. For example, it is common to assume that the propensity score obeys a logistic regression model, that is, $\pi(\gamma, X) = \text{expit}(\gamma'(1, X')')$, where γ is the nuisance parameter. For a consistent estimator $\hat{\gamma}$, the IPW estimator of $\mu = E(Y^r)$ is defined as follows (Horvitz & Thompson, 1952):

$$\hat{\mu}_{\text{IPW}}(\hat{\gamma}) = \frac{1}{n} \sum_{i=1}^{n} \frac{I(R_i = r)Y_i}{\pi^r(\hat{\gamma}, X_i)(1 - \pi(\hat{\gamma}, X_i))^{1-r}}. \tag{1}$$

For instance, the IPW estimator of $E(Y^1)$ is defined as $\frac{1}{n}\sum_{i=1}^{n} \frac{R_i Y_i}{\pi(\hat{\gamma}, X_i)}$ and the IPW estimator of $E(Y^0)$ is defined as $\frac{1}{n}\sum_{i=1}^{n} \frac{(1-R_i)Y_i}{1-\pi(\hat{\gamma}, X_i)}$. The IPW estimator of μ is consistent if the propensity score model is correctly specified (Lunceford & Davidian, 2004).

Consider next the problem of estimating the effect of a time-varying treatment on an end-of-study outcome. For each subject $i = 1, \ldots, n$ we denote $R_{t,i}$ as the treatment variable at time $t \in \{0, \ldots, T\}$, $X_{t,i}$ as the p-dimensional vector of observed covariates at time point t and Y_i as the outcome of interest at time T. Again we assume that the treatment variables are binary, that is, $R_{t,i} = 1$ if subject i received treatment at time t and $R_{t,i} = 0$ if subject i received control at time t. We denote $\overline{X}_t = (X_0, \ldots, X_t)$, $\overline{R}_t = (R_0, \ldots, R_t)$ and $Y^{\overline{r}}$ as the potential outcome under treatment $\overline{r} = (r_0, \ldots, r_T)$. We assume that at each time point t the potential outcome at the end of the study is conditionally independent of the treatment indicator at period t given the history of covariates and exposures ($Y^{\overline{r}} \perp\!\!\!\perp R_t \,|\, \overline{X}_t, \overline{R}_{t-1} = \overline{r}_{t-1}$ for $\overline{r}_t = (r_0, \ldots, r_t)$ where $r_j \in \{0; 1\}$ for $j = 0, \ldots, t$), that is, the assumption of sequential randomization holds (Robins, 1986, 1999). As in the fixed time period case, we do not impose any other restrictions on the outcome and the covariates. We define parametric propensity score models at each time point:

$$P(R_0 = 1 \,|\, X_0) = \pi_0(\gamma_0, X_0) \in (0, 1),$$

$$\ldots$$

$$P(R_t = 1 \,|\, \overline{X}_T, \overline{R}_{T-1}) = \pi_T(\gamma_T, \overline{X}_T) \in (0, 1),$$

where each γ_t is an unknown nuisance parameter and functions $\pi_0, \dots, \pi_T$ obey similar restrictions as in the fixed time period setting. For consistent estimators $\hat{\eta} = (\hat{\gamma}_0, \dots, \hat{\gamma}_T)$, the IPW estimator of the counterfactual mean $\phi = E(Y^{\bar{r}})$ is defined as

$$\hat{\phi}_{\mathrm{IPW}}(\hat{\eta}) = \frac{1}{n}\sum_{i=1}^{n}\frac{I(R_{0,i}=r_0)\ \dots\ I(R_{T,i}=r_T)Y_i}{\hat{\pi}_{0,i}^{r_0}(1-\hat{\pi}_{0,i})^{1-r_0}\ \dots\ \hat{\pi}_{T,i}^{r_T}(1-\hat{\pi}_{T,i})^{1-r_T}}, \tag{2}$$

where we denote $\hat{\pi}_t \equiv \pi(\hat{\gamma}_t, \overline{X}_t)$ for $t = 0, \dots, T$. In particular, the IPW estimator of $E(Y^{(1,\dots,1)})$ is defined as $\frac{1}{n}\sum_{i=1}^{n}\frac{R_{0,i}\dots R_{T,i}Y_i}{\hat{\pi}_{0,i}\dots\hat{\pi}_{T,i}}$.

The choice of estimation method for the nuisance parameters can drastically affect the performance of $\hat{\mu}_{\mathrm{IPW}}(\hat{\gamma})$ and $\hat{\phi}_{\mathrm{IPW}}(\hat{\eta})$. For example, in the fixed time period settings, estimated propensity scores close to zero or one may lead to poor performance of the target estimator. In practice, the nuisance parameter is estimated through standard MLE assuming that the propensity score obeys a logistic regression model. The MLE does not guarantee the stability of the estimated inverse weights since even minor changes in the regression coefficients may drastically change the inverse weights for treated subjects who were unlikely treated. The choice of estimation method for the nuisance parameters is especially important in the context of longitudinal studies, where the inverse weights accumulate as a product of inverse probability weights across time.

3 | PROPOSED METHODOLOGY

3.1 | The fixed time period case

Before proceeding with our proposed methodology for longitudinal studies, we first consider the problem of stable inverse probability weight estimation in point treatment studies. For the sake of simplicity we consider a logistic regression model for the propensity score, that is, $\pi(\gamma; X) = \mathrm{expit}(\gamma'(1, X')')$, for $r = 0, 1$. We propose to estimate γ by solving the following system of $p + 1$ estimating equations:

$$0 = \sum_{i=1}^{n}\left(1 - \frac{I(R_i = r)}{\pi^r(\gamma, X_i)(1 - \pi(\gamma, X_i))^{1-r}}\right)(1, X_i')'. \tag{3}$$

For instance, if one aims at estimating the mean $E(Y^1)$, the following system of $p + 1$ estimating equations will be considered:

$$0 = \sum_{i=1}^{n}\left(1 - \frac{R_i}{\pi(\gamma, X_i)}\right)(1, X_i')'. \tag{4}$$

Otherwise, if interest lies in the counterfactual mean $E(Y^0)$, then the following system of $p + 1$ estimating equations will be considered:

$$0 = \sum_{i=1}^{n}\left(1 - \frac{1 - R_i}{1 - \pi(\gamma, X_i)}\right)(1, X_i')'. \tag{5}$$

The system of estimating equations (3) is popular in the survey sampling literature, where they are known as calibration weighting or calibration equations. In this context, they are designed to provide less volatile inverse weights when estimating the population mean (see Deville & Särndal, 1992; Kott & Liao, 2012; Wu, 2003). This choice of estimating equations has several attractive characteristics:

1. The estimating equations (3) ensure efficiency and robustness of the IPW estimator, when the outcome mean is linear in X_i. Under this assumption (i.e., $E(Y \mid R = r, X) = m(\beta_r, X) = \beta_r'(1, X')'$, where β_r is $p + 1$-dimensional nuisance parameter of the outcome model), the IPW estimator is the efficient estimator of $E(Y^r)$ under the model defined by the restrictions on the propensity score. This estimator can indeed be seen to equal the augmented IPW (AIPW) (Robins et al., 1994) because it can be rewritten as

$$\hat{\mu}_{\text{AIPW}}(\hat{\gamma}, \hat{\beta}) = \hat{\mu}_{\text{IPW}}(\hat{\gamma}) + \frac{1}{n} \sum_{i=1}^{n} \left(1 - \frac{I(R_i = r)}{\hat{\pi}_i^r (1 - \hat{\pi}_i)^{1-r}} \right) m(\beta_r, X_i).$$

   Note that the second term in the right-hand side of the equation above becomes zero when $m(X; \beta_r) = \beta_r'(1, X')'$. The AIPW estimator is moreover double robust (DR), in the sense that it is consistent for μ if either the propensity score model π or the linear outcome model m is correctly specified, but not necessarily both (Scharfstein et al., 1999).

2. The estimating equations (4) (and more generally (3)) ensure stability of the inverse weights by imposing the following two constraints. First, they guarantee that the sum of the inverse weights for treated subjects equals the sample size n (e.g., from (4) it follows that $\sum_{i=1}^{n} \frac{R_i}{\pi(\hat{\gamma}, X_i)} = n$). Second, they induce a certain covariate balancing condition, that is, they make the weighted average of the covariates for treated subjects equal to the average of the corresponding covariates in the whole sample (e.g., from (4) it follows that $\sum_{i=1}^{n} \frac{R_i}{\pi(\hat{\gamma}, X_i)} X_i = \sum_{i=1}^{n} X_i$).

3. In contrast to the existing approaches for estimating inverse probability weights, the proposed approach adapts to the estimand. For example, depending on the target of interest $E(Y^1)$ or $E(Y^0)$, the nuisance parameter γ is obtained by solving the estimating equations (4) or (5), respectively. The use of a different estimator of γ, depending on whether $E(Y^1)$ or $E(Y^0)$ is estimated, is attractive. It ensures stability of the considered weights, which differ when estimating $E(Y^1)$ versus $E(Y^0)$.

3.2 | The longitudinal case

In this article, we will extend the previous proposal to longitudinal studies. We consider logistic regression models for the propensity scores, that is, $\pi_t \equiv \pi_t(\gamma_t; \overline{X}_t) = \text{expit}(\gamma_t'(1, \overline{X}_t')')$, with variation-independent parameters γ_t, for $t = 0, \dots, T$ and we focus on the estimation of the mean $\phi = E(Y^r)$. For simplicity, we include all historical covariates in the set of covariates at any time period $t = 1, \dots, T$ (we will allow for more generality later). To guarantee the stability of the inverse weights in each time period, we propose to estimate the nuisance parameters $\gamma_0, \dots, \gamma_T$ using the following system of estimating equations:

$$0 = \sum_{i=1}^{n} \left(1 - \frac{I(R_{0,i} = r_0)}{\pi_{0,i}^{r_0} (1 - \pi_{0,i})^{1-r_0}} \right) (1, X_{0,i}')',$$

$$0 = \sum_{i=1}^{n} \frac{I(R_{0,i} = r_0)}{\pi_{0,i}^{r_0}(1 - \pi_{0,i})^{1-r_0}} \left(1 - \frac{I(R_{1,i} = r_1)}{\pi_{1,i}^{r_1}(1 - \pi_{1,i})^{1-r_1}}\right) \left(1, \overline{X}_{1,i}'\right)',$$

$$\cdots$$

$$0 = \sum_{i=1}^{n} \frac{I(R_{0,i} = r_0)}{\pi_{0,i}^{r_0}(1 - \pi_{0,i})^{1-r_0}} \cdots \frac{I(R_{T-1,i} = r_{T-1})}{\pi_{T-1,i}^{r_{T-1}}(1 - \pi_{T-1,i})^{1-r_{T-1}}} \left(1 - \frac{I(R_{T,i} = r_T)}{\pi_{T,i}^{r_T}(1 - \pi_{T,i})^{1-r_T}}\right) \left(1, \overline{X}_{T,i}'\right)', \quad (6)$$

We will refer to our proposed method as Stable Estimand AdaptiVE IPW estimation (hereafter SEAVE). The proposed estimating equations deliver IPW estimators with similar properties as before. First, the IPW estimator $\hat{\phi}_{IPW}(\hat{\eta})$ equals the AIPW estimator when $\hat{\eta} = (\hat{\gamma}_0, \ldots, \hat{\gamma}_T)$ is the solution of (6) with $E(Y \mid \overline{R}_T = \overline{r}, \overline{X}_T) = m_T(\beta_{T,\overline{r}}, \overline{X}_T) = \beta_{T,\overline{r}}'(1, \overline{X}_T')'$ and $E(m_{t+1} \mid \overline{R}_t = \overline{r}_t, \overline{X}_t) = m_t(\beta_{t,\overline{r}_t}, \overline{X}_t) = \beta_{t,\overline{r}_t}'(1, \overline{X}_t')'$ for each $t = 0, \ldots, T - 1$, where $\beta_{t,\overline{r}_t}$ is the nuisance parameter of the outcome model at time t corresponding to the treatment regime $\overline{r}_t$. Indeed, the AIPW estimator is defined as (Robins & Rotnitzky, 1995; van der Laan & Robins, 2003)

$$\hat{\phi}_{AIPW}(\hat{\eta}, \hat{\beta}_0, \ldots, \hat{\beta}_T) = \hat{\phi}_{IPW}(\hat{\eta}) + \frac{1}{n}\sum_{i=1}^{n} \left(1 - \frac{I(R_{0,i} = r_0)}{\hat{\pi}_{0,i}^{r_0}(1 - \hat{\pi}_{0,i})^{1-r_0}}\right) m_0(\hat{\beta}_{0,r_0}, X_{0,i})$$

$$+ \frac{1}{n}\sum_{i=1}^{n} \frac{I(R_{0,i} = r_0)}{\hat{\pi}_{0,i}^{r_0}(1 - \hat{\pi}_{0,i})^{1-r_0}} \left(1 - \frac{I(R_{1,i} = r_1)}{\hat{\pi}_{1,i}^{r_1}(1 - \hat{\pi}_{1,i})^{1-r_1}}\right) m_1(\hat{\beta}_{1,\overline{r}_1}, \overline{X}_{1,i})$$

$$\cdots$$

$$+ \frac{1}{n}\sum_{i=1}^{n} \frac{I(R_{0,i} = r_0)}{\hat{\pi}_{0,i}^{r_0}(1 - \hat{\pi}_{0,i})^{1-r_0}} \cdots \frac{I(R_{T-1,i} = r_{T-1})}{\hat{\pi}_{T-1,i}^{r_{T-1}}(1 - \hat{\pi}_{T-1,i})^{1-r_{T-1}}}$$

$$\times \left(1 - \frac{I(R_{T,i} = r_T)}{\hat{\pi}_{T,i}^{r_T}(1 - \hat{\pi}_{T,i})^{1-r_T}}\right) m_T(\hat{\beta}_{T,\overline{r}}, \overline{X}_{T,i}),$$

where all terms (except $\hat{\phi}_{IPW}(\hat{\eta})$) equal 0 by (6). This is attractive because the AIPW estimator is locally efficient (when the outcome model is correctly specified at each time t) and DR.

The estimating equations (6) moreover ensure stability of the inverse weights by inducing the following constraints. For the sake of simplicity, we consider the regime $\overline{r} = (1, \ldots, 1)$. At any time point $t = 0, \ldots, T$, the estimating equations guarantee that the sum of the inverse weights equals the sample size, that is,

$$\sum_{i=1}^{n} \frac{R_{0,i} \cdots R_{t,i}}{\hat{\pi}_{0,i} \cdots \hat{\pi}_{t,i}} = n, \quad (7)$$

which is a desirable property. Next, at any time point $t > 0$, it makes the following covariate balancing conditions hold

$$\sum_{i=1}^{n} \frac{R_{0,i} \cdots R_{t,i}}{\hat{\pi}_{0,i} \cdots \hat{\pi}_{t,i}} X_{k,i} = \sum_{i=1}^{n} \frac{R_{0,i} \cdots R_{t-1,i}}{\hat{\pi}_{0,i} \cdots \hat{\pi}_{t-1,i}} X_{k,i} \quad (8)$$

for the set of covariates X_k at time point $k = 0, \ldots, t$. By imposing these constraints, the considered estimated inverse probability weights target the estimand.

Remark 1. It is likely that some of the historical covariates cannot be included in the model when the number of time periods is large. In this case, we can define the following propensity score function at time t:

$$P(R_t = 1 \mid W_t, \overline{R}_{t-1}) = \pi_t^*(\gamma_t^*, W_t), \tag{9}$$

where W_t is the set of covariates which contains X_t and a subset of the covariates in $\overline{X}_{t-1}$. For example, by considering the propensity score $\pi_1^*(\gamma_1^*, W_1) = \text{expit}(\gamma^{*'}(1, X_1')')$ at $t = 1$, we define $W_1 = X_1$ and omit the baseline covariate X_0. The function π_t^* obeys similar restrictions as π_t and γ_t^* is the corresponding nuisance parameter. Furthermore, we can reformulate our proposed estimating equations (6) by replacing π_t and $\overline{X}_t$ with π_t^* and W_t, respectively.

Note from (8) that by including lagged covariates $\overline{X}_{t-1}$ in the set of covariates W_t, we obtain additional balancing conditions. For example, if the baseline covariate X_0 is included in all covariate sets $\overline{X}_t$ for $t = 1, \ldots, T$, then we will obtain the following sequence of balancing conditions for X_0:

$$\sum_{i=1}^{n} \frac{R_{0,i} \ldots R_{T,i}}{\hat{\pi}_{0,i} \ldots \hat{\pi}_{T,i}} X_{0,i} = \sum_{i=1}^{n} \frac{R_{0,i} \ldots R_{T-1,i}}{\hat{\pi}_{0,i} \ldots \hat{\pi}_{T-1,i}} X_{0,i} = \cdots = \sum_{i=1}^{n} \frac{R_{0,i}}{\hat{\pi}_{0,i}} X_{0,i} = \sum_{i=1}^{n} X_{0,i}. \tag{10}$$

Our numerical results confirm that additional balancing conditions lead to better estimation performance when the number of time periods is not large, but not generally otherwise.

4 | RELATION TO THE EXISTING LITERATURE

As mentioned in Section 1, several methods have been proposed to achieve stable inverse weights in fixed time period studies. In the literature related to causal inference and missing data problems, the criterion (3) is closely related to the covariate balancing framework which has been widely studied by several authors (see, e.g., Athey et al., 2018; Chan et al., 2016; Fan et al., 2016; Fong et al., 2018; Imai & Ratkovic, 2014; Zubizarreta, 2015). Most prominently, Imai and Ratkovic (2014) developed an approach known as covariate balancing propensity score (CBPS) aimed at estimating the propensity scores such that resulting covariate balancing is optimized. This method aims to improve covariate balance in the observed data and may potentially provide robust and more accurate IPW estimates of the counterfactual means compared with IPW estimates based on traditional maximum likelihood propensity scores. In general, CBPS balances a particular function of covariates X, that is,

$$0 = \sum_{i=1}^{n} \left(\frac{R_i}{\pi(\gamma, X_i)} - \frac{1 - R_i}{(1 - \pi(\gamma, X_i))} \right) f(X_i), \tag{11}$$

where function $f()$ is specified by the researcher. For instance, if $f(X) = (1, X')'$ then the first moment of each covariate will be balanced (known as "exactly identified" CBPS). In this case, CBPS ensures that the weighted average of the covariates for treated and control subjects are equal. Our proposed approach additionally ensures that these averages equal the sample average of the corresponding covariates.

Recently, several authors (see, for instance, Hainmueller, 2012; Yiu & Su, 2018; Zubizarreta, 2015, among others) have focused on estimating the inverse probability weights

directly without involving a propensity score model. In particular, Zubizarreta (2015) designed a method based on convex optimization that is referred to as SBW which aims at directly estimating the inverse weights in the context of a missing data problem. The SBW method simultaneously minimizes the variance of the inverse weights and balances the covariates under a tolerance level which is specified by the researcher based on data. In addition, Wang and Zubizarreta (2020) generalized this approach, aimed at minimizing the sum of a convex function applied on the weights and balancing basis function of the covariates. In Appendix A (see the supplementary materials), we conduct a simulation study in fixed time period settings and evaluate a comparison of IPW estimators when the inverse probability weights are estimated using MLE, SBW, CBPS, and our proposed approach (3).

Although considerable research has been devoted to estimating stable inverse probability weights in one time period studies, less attention has been paid to this problem in longitudinal studies. Imai and Ratkovic (2015) generalized the CBPS framework to longitudinal studies for estimating the inverse probability weights for MSM (Robins, 2000; Robins et al., 2000). While our proposed method in particular ensures the balancing conditions of CBPS in fixed time point studies for $f(x) = (1, x')'$, our numerical simulations demonstrate that, under certain settings CBPS, may fail to provide desirable stability for the estimated inverse weights.

Han (2016) considered the problem of estimating the propensity score in longitudinal studies with dropout using DR estimation. The proposed approach is based on the following calibrated estimating equations:

$$0 = \sum_{i=1}^{n} \left(\frac{R_{1,i}}{\hat{\pi}_{1,i}(\lambda_1)} - 1 \right) \hat{g}_{0,i},$$

$$0 = \sum_{i=1}^{n} \frac{R_{1,i}}{\hat{\pi}_{1,i}(\hat{\lambda}_1)} \left(\frac{R_{2,i}}{\hat{\pi}_{2,i}(\lambda_2)} - 1 \right) \hat{g}_{1,i},$$

$$\dots$$

$$0 = \sum_{i=1}^{n} \frac{R_{T-1,i}}{\hat{\pi}_{1,i}(\hat{\lambda}_1) \ \dots \ \hat{\pi}_{T-1,i}(\hat{\lambda}_{T-1})} \left(\frac{R_{T,i}}{\hat{\pi}_{T,i}(\lambda_T)} - 1 \right) \hat{g}_{T-1,i}, \tag{12}$$

where R_t denotes the indicator of observing a subject at period $t = 1, \dots, T$, that is, $R_t = 1$ if the subject is observed or is still in the study at period t and $R_t = 0$, otherwise. Here, $\hat{\pi}_t(\lambda_t)$ is the calibration of the propensity score $\pi_t = P(R_t = 1 \mid \overline{R}_{t-1}, \overline{X}_t)$, $\lambda = (\lambda_1, \dots, \lambda_T)$ is a calibration parameter and is estimated sequentially by solving (12). Two calibration techniques considered, multiplicative and additive, are defined as

$$\hat{\pi}(\lambda_t) = \hat{\pi}_t \exp\left(\lambda_t' \hat{g}_{t-1} \frac{1 - \hat{\pi}_t}{\hat{\pi}_1 \ \dots \ \hat{\pi}_t} \right),$$

$$\hat{\pi}(\lambda_t) = \hat{\pi}_t + \lambda_t' \hat{g}_{t-1} \frac{1 - \hat{\pi}_t}{\hat{\pi}_1 \ \dots \ \hat{\pi}_t},$$

respectively, where the noncalibrated $\hat{\pi}_t$ is obtained by maximizing the partial likelihood. Finally, $\hat{g}_{t-1}$ is a function of $\overline{X}_{t-1}$ which includes 1, $\hat{\pi}_t$ and the estimate of $E(Y \mid \overline{X}_{t-1})$. Although the proposed system of estimating equations is similar to (6), the main goal of Han (2016) is to achieve intrinsic efficiency of the AIPW estimator and not directly stability of the inverse probability weights. Intrinsic efficiency guarantees that, when the missingness model is correctly specified, the estimator $\hat{\phi}$ attains the minimum asymptotic variance among the class of AIPW estimators

regardless of whether the considered model for the outcome is correctly specified. The calibration of the propensity score also requires estimation of the noncalibrated propensity scores $\hat{\pi}_t$ using maximum likelihood. Moreover, the calibrated propensity scores are not guaranteed to lie between 0 and 1. Finally, in contrast to our method, the proposal by Han (2016) requires the estimation of $E(Y | \overline{X}_{t-1})$. Although the proposed calibration approach focuses on studies with dropouts, we include its adaptation in the numerical study provided in Appendix A. In our study, we use the additive calibration technique since it was recommended by the author.

More recently, Kallus and Santacatterina (2018) introduced a method called kernel optimal weighting (KOW) which is based on a quadratic optimization problem. Their method directly estimates the inverse weights by balancing the time-dependent confounders (by using kernels) and simultaneously controlling the precision of the MSM. Specifically, this method optimizes the trade-off between the imbalance (defined as the sum of discrepancies between the weighted observed data and the counterfactual over all treatment regimes) and the precision (defined as the squared distance of the weights from uniform weights).

5 | SIMULATION STUDY

In this section, we perform a simulation study to compare the performance of the proposed estimator SEAVE with that of different estimators of a counterfactual mean ϕ. Specifically, in Subsection 5.1, we detail the considered estimators. In Subsection 5.2, we describe the simulation scenarios for the models. We provide additional simulation results in Appendix C (see the supplementary materials).

5.1 | Considered estimators and study layout

In our numerical study, we consider the following weighted estimators for given nuisance parameter estimates $\hat{\eta} = (\hat{\gamma}_0, \dots, \hat{\gamma}_T)$: the Horvitz–Thompson estimator $\hat{\phi}_{\mathrm{IPW}}(\hat{\eta})$ defined in (2) and the sample-bounded IPW (SB-IPW) estimator with normalized weights (Hirano & Imbens, 2001; Hirano et al., 2003):

$$
\hat{\phi}_{\mathrm{SB-IPW}}(\hat{\eta}) = \left(\sum_{i=1}^{n} \frac{I(R_{0,i} = r_0) \dots I(R_{T,i} = r_T) Y_i}{\hat{\pi}_{0,i}^{r_0}(1 - \hat{\pi}_{0,i})^{1-r_0} \dots \hat{\pi}_{T,i}^{r_T}(1 - \hat{\pi}_{T,i})^{1-r_T}} \right)
$$
$$
\times \left(\sum_{i=1}^{n} \frac{I(R_{0,i} = r_0) \dots I(R_{T,i} = r_T)}{\hat{\pi}_{0,i}^{r_0}(1 - \hat{\pi}_{0,i})^{1-r_0} \dots \hat{\pi}_{T,i}^{r_T}(1 - \hat{\pi}_{T,i})^{1-r_T}} \right)^{-1}. \tag{13}
$$

This estimator can be obtained by solving the following score equations:

$$
\frac{1}{n} \sum_{i=1}^{n} U_i(\hat{\phi}) = 0,
$$

where

$$
U_i(\phi) = I(R_{0,i} = r_0) \dots I(R_{T,i} = r_T) \left(\hat{\pi}_{0,i}^{r_0}(1 - \hat{\pi}_{0,i})^{1-r_0} \dots \hat{\pi}_{T,i}^{r_T}(1 - \hat{\pi}_{T,i})^{1-r_T} \right)^{-1} (Y_i - \phi)
$$

for $i = 1, \ldots, n$. The considered estimator (13) is sample-bounded in the sense that $\hat{\phi}_{\text{SB-IPW}}(\hat{\eta})$ lies in the range of the outcome Y_i. Due to the condition (7) both SB-IPW and IPW estimators are equal if the nuisance parameters are estimated using our proposed approach (6).

We examine the numerical performance of these estimators when parameter γ_t is estimated for each period t using the following methods:

1. Maximum likelihood estimator, that is, using standard logistic regression (MLE).
2. Covariate balancing propensity score (CBPS),
3. Our proposed estimator (SEAVE).

In our study, we use the R commands `glm` and `CBPS` to obtain MLE and CBPS, respectively. For the CBPS approach we use the argument `exact` aimed at obtaining "exactly identified" CBPS. We additionally consider the argument `over` (not reported) which provides "overidentified" CBPS by employing both covariate balancing and score conditions, but this may provide poor finite sample performance (Imai & Ratkovic, 2014), which was confirmed by our simulation results. Finally, we solve the estimating equations of SEAVE by considering the corresponding optimization problems (see Fu, 1998) which are solved using the command `nlm`. In Appendix E (see the supplementary materials), we provide R-code for a particular simulation setup (e.g., Scenario 1 described below).

Remark 2. If no convergence is attained for solving certain estimating equations of the proposed SEAVE approach using the command `nlm`, we recommend penalizing the corresponding estimating function using a ridge penalty $\epsilon\gamma_t$, where the penalty parameter ϵ has a small positive value (e.g., $\epsilon = 0.001$).

To evaluate the performance of a given estimator $\hat{\phi}$, we consider the following measures:

1. Monte Carlo bias (Bias): $\frac{1}{M}\sum_{m=1}^{M}\tilde{\phi}_m - \phi$
2. Root mean square error (RMSE): $\left[\frac{1}{M}\sum_{m=1}^{M}(\tilde{\phi}_m - \phi)^2\right]^{1/2}$
3. Median of absolute errors (MAE): $\text{med}|\tilde{\phi} - \phi|$
4. Monte Carlo standard deviation (MCSD): $\text{sd}(\tilde{\phi})$,

where M is the number of replications, $\tilde{\phi} = (\hat{\phi}_1, \ldots, \hat{\phi}_M)$ is a $1 \times M$ vector of estimators $\hat{\phi}$ based on each Monte Carlo replication. In our numerical evaluation, we also consider the average of the sandwich standard errors (ASSE) and the Monte Carlo coverage of 95% confidence intervals (COV). For the IPW estimator, the standard error $\hat{\sigma}/\sqrt{n}$ is calculated as $\hat{\sigma} = \text{sd}_n(U)$, where sd_n is the empirical standard deviation and U is the score function of the IPW estimator. The 95% confidence intervals are calculated as

$$\text{CI}(\phi) = \left(\hat{\phi} \pm 1.96\hat{\sigma}/\sqrt{n}\right).$$

Furthermore, in order to calculate the confidence intervals for the SB-IPW estimator, we consider the estimated standard error $\hat{\sigma}/\sqrt{n}$ with

$$\hat{\sigma} = \text{sd}_n(U)\text{E}_n^{-1}\left[\prod_{t=0}^{T}\frac{I(R_{t,i} = r)}{\hat{\pi}_{t,i}^{r_t}(1 - \hat{\pi}_{t,i})^{1-r_t}}\right], \tag{14}$$

where U is the score function of SB-IPW estimator defined earlier and E_n is the empirical expected value. We also consider the corrected estimators of the standard errors (accounting for the uncertainty in the estimated propensity scores) by calculating the standard deviations using the influence function (Tsiatis, 2007):

$$\hat{\sigma} = \mathrm{sd}_n \left(U - E \left(\frac{\partial U}{\partial \eta} \right) E^{-1} \left(\frac{\partial U_\eta}{\partial \eta} \right) U_\eta \right) E_n^{-1} \left[\prod_{t=0}^{T} \frac{I(R_{t,i} = r)}{\hat{\pi}_{t,i}^{r_t}(1 - \hat{\pi}_{t,i})^{1-r_t}} \right], \tag{15}$$

where $U_\eta = U_\eta(\overline{R}_T, \overline{X}_T)$ is the estimating function of the considered nuisance parameter estimator, as defined in Appendix B (see the supplementary materials) for each estimation method.

5.2 | Simulation scenarios

5.2.1 | Scenario 1

In the first simulation scenario, we consider a simple data-generating mechanism where for each $i = 1, \dots, n$ we generate $Z_i = (X_{0,i}, R_{0,i}, X_{1,i}, R_{1,i}, Y_i)$ as

$$\begin{aligned}
X_{0,i} &\sim N(0, 1), \\
R_{0,i} &\sim \mathrm{Ber}(\mathrm{expit}(1 + X_{0,i})), \\
X_{1,i} &\sim N(0.5X_{0,i} + 0.5R_{0,i}, 1), \\
R_{1,i} &\sim \mathrm{Ber}(\mathrm{expit}(X_{1,i} + X_{0,i} + 0.5R_{0,i})), \\
Y_i &\sim N(1 + 0.5R_{0,i} + 0.5R_{1,i} + X_{1,i} + X_{0,i}, 1).
\end{aligned}$$

We focus on estimating $E(Y^{(1,1)})$ at sample sizes $n = 500$, 1000, and 2000. The following working models are considered: $\pi_0(\gamma_0; X_0) = \mathrm{expit}(\gamma_0'\left(1, X_0'\right)')$ and $\pi_1(\gamma_1; \overline{X}_1) = \mathrm{expit}(\gamma_1'\left(1, \overline{X}_1'\right)')$.

5.2.2 | Scenario 2

In the second scenario, we consider three time periods. For each $i = 1, \dots, n$ we generate $Z_i = (X_{0,i}, X_{1,i}, X_{2,i}, R_{0,i}, R_{1,i}, R_{2,i}, Y_i)$, where

$$\begin{aligned}
X_{0,i} &\sim N(0, 1), \\
R_{0,i} &\sim \mathrm{Ber}(\mathrm{expit}(1 + X_{0,i})), \\
X_{1,i} &\sim N(0.5X_{0,i} + 0.5R_{0,i}, 1), \\
R_{1,i} &\sim \mathrm{Ber}(\mathrm{expit}(X_{1,i} + 0.5X_{0,i} + 0.5R_{0,i})), \\
X_{2,i} &\sim N(X_{1,i} + 0.5R_{1,i}, 1), \\
R_{2,i} &\sim \mathrm{Ber}(\mathrm{expit}(X_{2,i} + 0.5X_{1,i} + 0.25X_{0,i} + 0.5R_{1,i})), \\
Y_i &\sim N(1 + 0.5R_{0,i} + 0.5R_{1,i} + 0.5R_{2,i} + X_{2,i} + X_{1,i} + X_{0,i}, 1).
\end{aligned}$$

We focus on estimating $E(Y^{(1,1,1)})$ at sample size $n = 500$, 1000, and 2000. The following working models are considered $\pi_0(\gamma_0; X_0) = \mathrm{expit}(\gamma_0'\left(1, X_0'\right)')$, $\pi_1(\gamma_1; \overline{X}_1) = \mathrm{expit}(\gamma_1'\left(1, \overline{X}_1'\right)')$, and $\pi_2(\gamma_2; \overline{X}_2) = \mathrm{expit}(\gamma_2'\left(1, \overline{X}_2'\right)')$.

5.2.3 | Scenario 3

In the third scenario, we consider a data-generating mechanism where for each $i = 1, \ldots, n$, we generate $Z_i = (X_{0,i}, R_{0,i}, X_{1,i}, R_{0,i}, Y_i)$, $X_{0,i} = (X_{01i}, X_{02i}, X_{03i}, X_{04i})$, $X_{1,i} = (X_{11i}, X_{12i}, X_{13i}, X_{14i})$, as

$$X_{t1i} = U_{it}W_{t1i}, \quad X_{t2i} = U_{ti}W_{t2i}, \quad X_{t3i} = |U_{ti}W_{t3i}|, \quad X_{t4i} = |U_{ti}W_{t4i}|,$$

$$W_{tki} \sim N(0,1), \quad \text{for } k = 1, \ldots, 4, \text{ and } U_{0i} = 1,$$

$$R_{0,i} \sim \text{Ber}(\text{expit}(\gamma'X_{0,i} - 0.5)), \quad U_{1i} = 2 + (2R_{0,i} - 1)/3,$$

$$R_{1,i} \sim \text{Ber}(\text{expit}(-R_{0,i} + \gamma'X_{1,i} + 0.25)),$$

$$Y_i \sim N(250 - 10R_{0,i} - 10R_{1,i} + \delta'X_{0,i} + \delta'X_{1,i}, 1))$$

for $t = 0, 1$, where $\gamma = (1, -0.5, 0.25, 0.1)'$ and $\delta = (27.4, 13.7, 13.7, 13.7)'$. This mechanism was previously used by Imai and Ratkovic (2015) with three time periods. In our study, we consider only two time periods, because the additional time period leads to fewer observations required for the convergence of the proposed SEAVE approach (considering the presence of the instrumental variables).

We focus on estimating $E(Y^{(1,1)})$ at sample sizes $n = 3000$ and 6000. We note that for this simulation setting relatively larger n, since smaller sample sizes lead to unidentified estimates of the proposed method. Following the data generating mechanism, we consider the following working models: $\pi_0(\gamma_0; X_0) = \text{expit}(\gamma_0'(1, X_0')')$ and $\pi_1(\gamma_1; \overline{X}_1) = \text{expit}(\gamma_1'(1, X_1')')$. We additionally consider the following working models: $\pi_0(\gamma_0; X_0) = \text{expit}(\gamma_0'(1, X_0')')$ and $\pi_1(\gamma_1; \overline{X}_1) = \text{expit}(\gamma_1'(1, \overline{X}_1')')$, thus, including the baseline covariates in the propensity score model π_1.

5.3 | Discussion of results

Tables 1–3 report the simulation results based on 1000 replications. We observe that SEAVE outperforms other methods in terms of all statistical measures. More specifically, our proposed methodology provides lower RMSE and MAE than MLE and CBPS. Moreover, we observe that for most of the considered settings, the confidence interval coverage (based on the corrected standard errors) obtained using the proposed SEAVE approach is closer to the nominal level than that calculated using MLE or CBPS methods.

We observe that for Scenario 3 (see Table 3) MLE and CBPS methods provide the lowest bias for IPW and SB-IPW estimators, respectively. However, both methods provide high RMSE and MAE as a result of high variability in the estimated inverse weights. Note that in this scenario the covariates X_0 do not affect the treatment variable A_1. We repeated our numerical study by including X_0 in the estimation procedure of the propensity score π_1 (see Table 4). By doing so, we induce an additional balancing condition for the covariate X_0. Table 4 shows that by including the baseline covariates X_0, we significantly improve the statistical performance of SEAVE in terms of all measures compared with those in Table 2.

Furthermore, we provide additional results in Appendix C (see the supplementary materials) for the counterfactual mean $E(Y^{(1,0)})$ for Scenarios 1 and 3. We observe that the proposed method again provides stable results. Moreover, similar results are obtained for dichotomous outcome, where SEAVE provides desirable results compared with other methods (see Appendix C for more details).

TABLE 1 Estimation of $E(Y^{(1,1)})$. Simulation results based on 1000 replications: Scenario 1

Estimator	Bias	RMSE	MAE	MCSD	ASSE	COV
$n = 500$						
$\hat{\phi}_{\text{IPW}}(\hat{\eta}_{\text{MLE}})$	−0.011	0.551	0.092	0.552	0.195	0.984
					0.159	0.929
$\hat{\phi}_{\text{IPW}}(\hat{\eta}_{\text{CBPS}})$	0.012	0.224	0.119	0.224	0.189	0.970
					0.158	0.889
$\hat{\phi}_{\text{IPW}}(\hat{\eta}_{\text{SEAVE}})$	0.0009	0.129	0.084	0.129	0.164	0.988
					0.121	0.926
$\hat{\phi}_{\text{SB−IPW}}(\hat{\eta}_{\text{MLE}})$	0.023	0.339	0.158	0.338	0.253	0.901
					0.199	0.819
$\hat{\phi}_{\text{SB−IPW}}(\hat{\eta}_{\text{CBPS}})$	−0.074	0.298	0.153	0.289	0.276	0.946
					0.193	0.848
$\hat{\phi}_{\text{SB−IPW}}(\hat{\eta}_{\text{SEAVE}})$	0.0009	0.129	0.084	0.129	0.246	0.997
					0.121	0.926
$n = 1000$						
$\hat{\phi}_{\text{IPW}}(\hat{\eta}_{\text{MLE}})$	0.009	0.148	0.069	0.148	0.129	0.973
					0.107	0.914
$\hat{\phi}_{\text{IPW}}(\hat{\eta}_{\text{CBPS}})$	0.009	0.160	0.087	0.159	0.133	0.944
					0.116	0.885
$\hat{\phi}_{\text{IPW}}(\hat{\eta}_{\text{SEAVE}})$	0.002	0.091	0.058	0.091	0.117	0.986
					0.087	0.933
$\hat{\phi}_{\text{SB−IPW}}(\hat{\eta}_{\text{MLE}})$	0.028	0.230	0.135	0.228	0.193	0.900
					0.156	0.818
$\hat{\phi}_{\text{SB−IPW}}(\hat{\eta}_{\text{CBPS}})$	−0.036	0.240	0.142	0.237	0.206	0.903
					0.156	0.847
$\hat{\phi}_{\text{SB−IPW}}(\hat{\eta}_{\text{SEAVE}})$	0.002	0.091	0.058	0.091	0.186	1
					0.087	0.933
$n = 2000$						
$\hat{\phi}_{\text{IPW}}(\hat{\eta}_{\text{MLE}})$	0.004	0.120	0.053	0.120	0.096	0.963
					0.082	0.916
$\hat{\phi}_{\text{IPW}}(\hat{\eta}_{\text{CBPS}})$	0.0002	0.131	0.064	0.132	0.100	0.935
					0.090	0.894
$\hat{\phi}_{\text{IPW}}(\hat{\eta}_{\text{SEAVE}})$	−0.0001	0.066	0.042	0.066	0.087	0.984
					0.063	0.931
$\hat{\phi}_{\text{SB−IPW}}(\hat{\eta}_{\text{MLE}})$	0.015	0.177	0.093	0.176	0.151	0.917
					0.126	0.852
$\hat{\phi}_{\text{SB−IPW}}(\hat{\eta}_{\text{CBPS}})$	−0.035	0.194	0.100	0.191	0.162	0.921
					0.129	0.887
$\hat{\phi}_{\text{SB−IPW}}(\hat{\eta}_{\text{SEAVE}})$	−0.0001	0.066	0.042	0.066	0.144	1
					0.063	0.931

Note: The top/bottom standard errors (and the corresponding confidence intervals) are obtained using the uncorrected/corrected standard deviations defined in (14) and (15), respectively (see Section 5.1). The corrected standard errors are considered to account for the uncertainty in the estimated propensity scores.

Abbreviations: ASSE, average of sandwich standard errors; bias, Monte Carlo bias; CBPS, covariate balancing propensity score; COV, coverage of 95% confidence intervals; IPW, inverse probability weighting; MAE, median of absolute errors; MCSD, Monte Carlo standard deviation; MLE, maximum likelihood propensity score estimator; RMSE, root mean square error; SB-IPW, sample-bounded IPW; SEAVE, Stable Estimand AdaptiVE IPW estimation.

TABLE 2 Estimation of $E(Y^{(1,1,1)})$. Simulation results based on 1000 replications: Scenario 2

Estimator	Bias	RMSE	MAE	MCSD	ASSE	COV
$n = 500^{a}$						
$\hat{\phi}_{\mathrm{IPW}}(\hat{\eta}_{\mathrm{MLE}})$	0.039	0.253	0.131	0.250	0.262	0.994
					0.198	0.947
$\hat{\phi}_{\mathrm{IPW}}(\hat{\eta}_{\mathrm{CBPS}})$	0.114	0.465	0.220	0.451	0.325	0.959
					0.253	0.833
$\hat{\phi}_{\mathrm{IPW}}(\hat{\eta}_{\mathrm{SEAVE}})$	0.010	0.192	0.121	0.192	0.260	0.993
					0.174	0.919
$\hat{\phi}_{\mathrm{SB-IPW}}(\hat{\eta}_{\mathrm{MLE}})$	0.116	0.424	0.275	0.408	0.359	0.861
					0.311	0.791
$\hat{\phi}_{\mathrm{SB-IPW}}(\hat{\eta}_{\mathrm{CBPS}})$	−0.219	0.651	0.290	0.613	0.432	0.903
					0.346	0.789
$\hat{\phi}_{\mathrm{SB-IPW}}(\hat{\eta}_{\mathrm{SEAVE}})$	0.010	0.192	0.121	0.192	0.374	0.996
					0.174	0.919
$n = 1000$						
$\hat{\phi}_{\mathrm{IPW}}(\hat{\eta}_{\mathrm{MLE}})$	0.015	0.337	0.100	0.336	0.205	0.996
					0.160	0.942
$\hat{\phi}_{\mathrm{IPW}}(\hat{\eta}_{\mathrm{CBPS}})$	0.042	0.298	0.144	0.295	0.230	0.964
					0.188	0.867
$\hat{\phi}_{\mathrm{IPW}}(\hat{\eta}_{\mathrm{SEAVE}})$	0.005	0.126	0.089	0.126	0.182	0.995
					0.122	0.953
$\hat{\phi}_{\mathrm{SB-IPW}}(\hat{\eta}_{\mathrm{MLE}})$	0.058	0.411	0.216	0.407	0.302	0.856
					0.264	0.803
$\hat{\phi}_{\mathrm{SB-IPW}}(\hat{\eta}_{\mathrm{CBPS}})$	−0.158	0.545	0.241	0.522	0.354	0.897
					0.308	0.838
$\hat{\phi}_{\mathrm{SB-IPW}}(\hat{\eta}_{\mathrm{SEAVE}})$	0.005	0.126	0.089	0.126	0.290	1
					0.122	0.953
$n = 2000$						
$\hat{\phi}_{\mathrm{IPW}}(\hat{\eta}_{\mathrm{MLE}})$	0.003	0.190	0.074	0.190	0.152	0.986
					0.123	0.920
$\hat{\phi}_{\mathrm{IPW}}(\hat{\eta}_{\mathrm{CBPS}})$	0.006	0.265	0.100	0.265	0.173	0.954
					0.145	0.882
$\hat{\phi}_{\mathrm{IPW}}(\hat{\eta}_{\mathrm{SEAVE}})$	−0.003	0.095	0.062	0.095	0.132	0.995
					0.088	0.927
$\hat{\phi}_{\mathrm{SB-IPW}}(\hat{\eta}_{\mathrm{MLE}})$	0.028	0.330	0.176	0.329	0.249	0.866
					0.222	0.804
$\hat{\phi}_{\mathrm{SB-IPW}}(\hat{\eta}_{\mathrm{CBPS}})$	−0.114	0.455	0.198	0.441	0.281	0.880
					0.263	0.849
$\hat{\phi}_{\mathrm{SB-IPW}}(\hat{\eta}_{\mathrm{SEAVE}})$	−0.003	0.095	0.062	0.095	0.229	1
					0.088	0.927

Note: The top/bottom standard errors (and the corresponding confidence intervals) are obtained using the uncorrected/corrected standard deviations defined in (14) and (15), respectively (see Section 5.1). The corrected standard errors are considered to account for the uncertainty in the estimated propensity scores.
Abbreviations: ASSE, average of sandwich standard errors; bias, Monte Carlo bias; CBPS, covariate balancing propensity score; COV, coverage of 95% confidence intervals; IPW, inverse probability weighting; MAE, median of absolute errors; MCSD, Monte Carlo standard deviation; MLE, maximum likelihood propensity score estimator; RMSE, root mean square error; SB-IPW, sample-bounded IPW; SEAVE, Stable Estimand AdaptiVE IPW estimation.
[a] No convergence was attained for SEAVE in 24 runs out of 1000.

TABLE 3 Estimation of $E(Y^{(1,1)})$. Simulation results based on 1000 replications: Scenario 3

Estimator	Bias	RMSE	MAE	MCSD	ASSE	COV
$n = 3000$						
$\hat{\phi}_{\text{IPW}}(\hat{\eta}_{\text{MLE}})$	−0.177	34.45	13.85	34.47	26.80	0.965
					19.35	0.823
$\hat{\phi}_{\text{IPW}}(\hat{\eta}_{\text{CBPS}})$	9.662	20.23	11.71	17.78	26.83	0.999
					12.01	0.814
$\hat{\phi}_{\text{IPW}}(\hat{\eta}_{\text{SEAVE}})$	0.995	4.457	3.074	4.347	25.88	1
					3.412	0.845
$\hat{\phi}_{\text{SB−IPW}}(\hat{\eta}_{\text{MLE}})$	2.161	11.80	7.208	11.60	8.637	0.809
					7.682	0.758
$\hat{\phi}_{\text{SB−IPW}}(\hat{\eta}_{\text{CBPS}})$	0.710	8.817	5.710	8.793	8.867	0.913
					5.988	0.803
$\hat{\phi}_{\text{SB−IPW}}(\hat{\eta}_{\text{SEAVE}})$	0.995	4.457	3.074	4.347	8.798	0.993
					3.412	0.845
$n = 6000$						
$\hat{\phi}_{\text{IPW}}(\hat{\eta}_{\text{MLE}})$	0.291	21.83	10.35	21.84	20.12	0.950
					15.25	0.860
$\hat{\phi}_{\text{IPW}}(\hat{\eta}_{\text{CBPS}})$	5.699	14.03	8.427	12.83	19.59	1
					9.213	0.859
$\hat{\phi}_{\text{IPW}}(\hat{\eta}_{\text{SEAVE}})$	0.510	3.064	2.059	3.023	18.87	1
					2.596	0.892
$\hat{\phi}_{\text{SB−IPW}}(\hat{\eta}_{\text{MLE}})$	1.007	9.910	5.647	9.863	7.322	0.829
					6.694	0.773
$\hat{\phi}_{\text{SB−IPW}}(\hat{\eta}_{\text{CBPS}})$	0.197	7.334	4.125	7.335	7.275	0.920
					4.962	0.840
$\hat{\phi}_{\text{SB−IPW}}(\hat{\eta}_{\text{SEAVE}})$	0.510	3.064	2.059	3.023	6.989	0.997
					2.596	0.892

Note: The top/bottom standard errors (and the corresponding confidence intervals) are obtained using the uncorrected/corrected standard deviations defined in (14) and (15), respectively (see Section 5.1). The corrected standard errors are considered to account for the uncertainty in the estimated propensity scores.
Abbreviations: ASSE, average of sandwich standard errors; bias, Monte Carlo bias; CBPS, covariate balancing propensity score; COV, coverage of 95% confidence intervals; IPW, inverse probability weighting; MAE, median of absolute errors; MCSD, Monte Carlo standard deviation; MLE, maximum likelihood propensity score estimator; RMSE, root mean square error; SB-IPW, sample-bounded IPW; SEAVE, Stable Estimand AdaptiVE IPW estimation.

6 | MARGINAL STRUCTURAL MODELS

In this section, we demonstrate the performance of the proposed SEAVE method for estimating the coefficients of an MSM. This is a model for the causal effect of (time-varying) exposure on the population mean of an outcome, and is defined as, for example,

$$E(Y^{\bar{r}}) = \rho_0 + \sum_{t=0}^{T} \rho_{t+1} r_t,$$

TABLE 4 Estimation of $E(Y^{(1,1)})$. Simulation results based on 1000 replications: Scenario 3, including additional balancing conditions

Estimator	Bias	RMSE	MAE	MCSD	ASSE	COV
$n = 3000$[a]						
$\hat{\phi}_{\text{IPW}}(\hat{\eta}_{\text{MLE}})$	−0.059	35.50	13.54	35.52	27.08	0.972
					18.78	0.806
$\hat{\phi}_{\text{IPW}}(\hat{\eta}_{\text{CBPS}})$	12.67	19.51	12.53	14.84	26.51	1
					9.707	0.714
$\hat{\phi}_{\text{IPW}}(\hat{\eta}_{\text{SEAVE}})$	0.020	2.663	1.834	2.665	26.37	1
					2.576	0.941
$\hat{\phi}_{\text{SB-IPW}}(\hat{\eta}_{\text{MLE}})$	2.021	11.77	7.151	11.60	8.716	0.832
					7.519	0.758
$\hat{\phi}_{\text{SB-IPW}}(\hat{\eta}_{\text{CBPS}})$	0.540	7.835	4.956	7.820	8.792	0.939
					5.238	0.819
$\hat{\phi}_{\text{SB-IPW}}(\hat{\eta}_{\text{SEAVE}})$	0.020	2.663	1.834	2.665	8.878	1
					2.576	0.941
$n = 6000$						
$\hat{\phi}_{\text{IPW}}(\hat{\eta}_{\text{MLE}})$	0.324	21.35	10.17	21.36	20.24	0.962
					14.84	0.844
$\hat{\phi}_{\text{IPW}}(\hat{\eta}_{\text{CBPS}})$	7.600	13.23	8.434	10.84	19.29	0.999
					7.506	0.761
$\hat{\phi}_{\text{IPW}}(\hat{\eta}_{\text{SEAVE}})$	−0.029	1.879	1.306	1.879	18.85	1
					1.836	0.949
$\hat{\phi}_{\text{SB-IPW}}(\hat{\eta}_{\text{MLE}})$	0.986	9.886	5.723	9.841	7.342	0.826
					6.566	0.772
$\hat{\phi}_{\text{SB-IPW}}(\hat{\eta}_{\text{CBPS}})$	0.237	6.316	3.903	6.314	7.126	0.955
					4.302	0.861
$\hat{\phi}_{\text{SB-IPW}}(\hat{\eta}_{\text{SEAVE}})$	−0.029	1.879	1.306	1.879	6.856	1
					1.836	0.949

Note: The top/bottom standard errors (and the corresponding confidence intervals) are obtained using the uncorrected/corrected standard deviations defined in (14) and (15), respectively (see Section 5.1). The corrected standard errors are considered to account for the uncertainty in the estimated propensity scores.

[a]No convergence was attained for SEAVE in one run out of 1000.

where $\rho_1, \dots, \rho_{T+1}$ are unknown causal parameters. Using the proposed estimating equations (6), the propensity scores $\hat{\pi}_0^{r_0}(1 - \hat{\pi}_0)^{1-r_0}, \dots, \hat{\pi}_T^{r_T}(1 - \hat{\pi}_T)^{1-r_T}$ are obtained for each feasible treatment regime $\bar{r}$ separately. Next, we define the inverse probability weights as

$$w_i(\bar{r}, \overline{X}_{T,i}) = \prod_{t=0}^{T} \frac{1}{\hat{\pi}_{t,i}^{r_t}(1 - \hat{\pi}_{t,i})^{1-r_t}}$$

for $i = 1, \dots, n$. Furthermore, the MSM coefficients $\rho = (\rho_0, \dots, \rho_{T+1})$ are estimated through a weighted least squares regression

$$\hat{\rho} = \arg\min_{\rho} \sum_{i=1}^{n} W_i (Y_i - \rho_0 - \sum_{t=0}^{T} \rho_{t+1} R_{t,i})^2,$$

where $W_i = w_i(\overline{R}_{T,i}, \overline{X}_{T,i})$. We demonstrate the performance of the proposed inverse weights for estimating the MSM coefficients and compare with that of MLE and CBPS on a particular numerical study. The weights W_i are estimated separably for each possible regime. We use Scenario 1 described in Subsection 5.2.1 and we estimate MSM coefficients at sample sizes $n = 500, 1000$, and 2000. It can be shown that the MSM is correctly specified under the data-generating mechanism and has the following form:

$$E(Y^{(r_0,r_1)}) = 1 + r_0 + 0.5r_1.$$

Table 5 summarizes the results of our study. We observe that our proposed methodology provides more stable estimators of ρ than MLE and CBPS methods. The results demonstrate that RMSE and MAE of the MSM coefficients are much lower if the corresponding weights are obtained through the SEAVE method compared with those for MLE and CBPS.

Remark 3. Although this article does not focus on estimating the parameters of the MSM, this simple example shows that the stability of the inverse probability weights positively affects the estimated MSM coefficients. This simulation study simply demonstrates the stability of the SEAVE approach which is consistent with the results obtained for the simulation Scenario 1 (see Section 5.2.1).

7 | APPLICATION

In this section, we illustrate the performance of the proposed methodology through an empirical application. We use a data set previously analyzed by Blackwell (2013) who focused on estimating the impact of negative advertisements (when a politician or party focuses on criticizing the opponents rather than highlighting their own qualities) on election outcomes. The data set includes information on 5 weeks leading to the elections from 114 U.S. election races from the period 2000–2006. In our study, we only focus on two time periods: the first week and the last week of the 5-week period. In each week $t \in \{0, 1\}$, the candidate i may run a negative campaign ($R_{t,i} = 1$) or remain positive ($R_{t,i} = 0$). The outcome (Y) is the "Democratic" percentage of the votes. We estimate the expected percentage of the votes under two regimes: running a negative campaign in the first week and in the last week (i.e., $\overline{r} = (1, 1)$), and remaining positive in the first week and in the last week (i.e., $\overline{r} = (0, 0)$). The following covariates are considered: campaign length, share of the polls, type of office, baseline share, and baseline proportion of undecided voters (observed at the first week of the considered five week period). We estimate counterfactual mean ϕ_0 using MLE, CBPS, and SEAVE methods. Table 6 summarizes the results. We observe that under the regime $\overline{r} = (1, 1)$ the estimators show roughly the same results. On the other hand, we observe that under the regime $\overline{r} = (0, 0)$ the proposed SEAVE method shows relatively lower corrected standard error. The results show that under the regime $\overline{r} = (0, 0)$ the inverse probability weights estimated using the proposed SEAVE approach are slightly more stable than those estimated using MLE or CBPS methods.

Additionally, we repeat the same calculations by considering three time periods: the first, the third, and the last week of the 5-week period leading to the elections. We estimate the expected percentage of the votes under two regimes: running a negative campaign in three periods (i.e., $\overline{r} = (1, 1, 1)$), and remaining positive in three periods (i.e., $\overline{r} = (0, 0, 0)$). The considered covariates are the same as above. Table 7 provides the results. We observe that under the regime $\overline{r} = (1, 1, 1)$ the estimators based on MLE and SEAVE show roughly the same results. On the other hand, we

TABLE 5 Estimation of the MSM coefficients. Simulation results based on 1000 replications. Scenario 1.

	Estimator	Bias	RMSE	MAE
$n = 500$				
	MLE	−0.166	0.518	0.337
ρ_0	CBPS	−0.061	0.412	0.264
	SEAVE	−0.034	0.251	0.170
	MLE	0.055	0.460	0.248
ρ_1	CBPS	−0.001	0.351	0.209
	SEAVE	0.020	0.227	0.155
	MLE	0.167	0.436	0.306
ρ_2	CBPS	0.010	0.298	0.183
	SEAVE	0.028	0.197	0.117
$n = 1000$				
	MLE	−0.079	0.478	0.267
ρ_0	CBPS	−0.030	0.351	0.208
	SEAVE	−0.011	0.177	0.113
	MLE	0.049	0.428	0.230
ρ_1	CBPS	0.001	0.296	0.183
	SEAVE	0.009	0.159	0.106
	MLE	0.066	0.375	0.225
ρ_2	CBPS	−0.004	0.256	0.151
	SEAVE	0.003	0.132	0.085
$n = 2000$				
	MLE	−0.065	0.350	0.211
ρ_0	CBPS	−0.022	0.282	0.173
	SEAVE	0.002	0.123	0.082
	MLE	0.026	0.298	0.165
ρ_1	CBPS	−0.004	0.227	0.140
	SEAVE	−0.001	0.112	0.072
	MLE	0.059	0.273	0.179
ρ_2	CBPS	−0.0007	0.209	0.126
	SEAVE	−0.002	0.096	0.061

Abbreviations: bias, Monte Carlo bias; CBPS, covariate balancing propensity score; MAE, median of absolute errors; MCSD, Monte Carlo standard deviation; MLE, maximum likelihood propensity score estimator; MSM, marginal structural model; RMSE, root mean square error; SEAVE, Stable Estimand AdaptiVE IPW estimation.

TABLE 6 Estimation of the counterfactual mean $E(Y^{\bar{r}})$ for two time periods

Estimator	IPW	SE	Corrected SE	SB-IPW	SE	Corrected SE
$\bar{r} = (1,1)$						
MLE	48.56	7.078	0.886	49.11	0.812	0.680
CBPS	48.86	7.048	0.797	49.05	0.833	0.707
SEAVE	49.07	7.245	0.698	49.07	0.823	0.698
$\bar{r} = (0,0)$						
MLE	50.24	8.535	1.733	49.86	1.998	1.163
CBPS	57.86	10.71	7.623	49.31	2.530	1.640
SEAVE	49.49	8.322	0.897	49.49	1.928	0.897

Abbreviations: CBPS, covariate balancing propensity score; IPW, inverse probability weighting; MLE, maximum likelihood propensity score estimator; SB-IPW, sample-bounded IPW; SEAVE, Stable Estimand AdaptiVE IPW estimation.

TABLE 7 Estimation of the counterfactual mean $E(Y^{\bar{r}})$ for three time periods

Estimator	IPW	SE	Corrected SE	SB-IPW	SE	Corrected SE
$\bar{r} = (1,1,1)$						
MLE	48.67	7.645	1.030	49.21	0.930	0.764
CBPS[a]	51.57	8.428	6.669	48.13	1.602	1.962
SEAVE	49.25	7.922	0.736	49.25	0.910	0.736
$\bar{r} = (0,0,0)$						
MLE	53.59	17.09	9.695	49.79	2.530	1.899
CBPS[b]	88.21	43.37	34.58	46.87	2.158	1.707
SEAVE	49.97	15.28	0.946	49.97	2.566	0.946

Abbreviations: CBPS, covariate balancing propensity score; IPW, inverse probability weighting; MLE, maximum likelihood propensity score estimator; SB-IPW, sample-bounded IPW; SEAVE, Stable Estimand AdaptiVE IPW estimation.

[a,b]CBPS method provides propensity score estimates close to one, which leads to extreme target estimates and standard errors for the IPW estimator.

observe that under the regime $\bar{r} = (0,0,0)$ the proposed SEAVE method shows relatively lower corrected standard error. In our study, we do not consider all five time periods since this leads to convergence issues for the considered approaches.

Remark 4. The results show that with three time periods the CBPS method provides propensity score estimates close to one. This leads to extreme target estimates and standard errors.

8 | DISCUSSION

In this article, we have developed a new approach for estimating the nuisance parameters indexing the propensity score model in IPW estimators for longitudinal studies. The proposed

method is based on a system of estimating equations and guarantees stability of the inverse weights at each time point. Moreover, it provides balancing conditions on the covariates. Unlike other existing estimation methods, our proposed approach adapts to the estimand, that is, depending on the target of interest, the corresponding nuisance parameter is obtained by solving different estimating equations. The IPW estimator that employs the proposed inverse probability weights equals the efficient and robust AIPW estimator when the outcome model is linear. Through an extensive numerical analysis we have demonstrated that our proposed method provides more stable IPW estimation than MLE and CBPS approaches. In our study, we considered the estimation of counterfactual means. In addition, we have demonstrated the performance of the considered methods on a simple example of fitting MSMs.

Despite desirable features of our proposed methodology, an important limitation is that solving (6) requires a separate model in each time period, and that (6) moreover may become difficult to solve at the larger time points where the observed treatment regime becomes sparse. The proposed approach may therefore not be flexible in problems where the number of time periods is large or there are many confounders in the study. In future work, we will remedy this by applying generalized method of moments (Hansen et al., 1996) to solve the estimating equations in (6) simultaneously. This will significantly increase the usability of the proposed method in settings such as the estimation of MSMs.

ORCID

Vahe Avagyan https://orcid.org/0000-0001-5838-4850
Stijn Vansteelandt https://orcid.org/0000-0002-4207-8733

REFERENCES

Athey S., Imbens G. W., Wager S. (2018). Approximate residual balancing: debiased inference of average treatment effects in high dimensions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *80*, (4), 597–623.

Blackwell, M. (2013). A framework for dynamic causal inference in political science. *American Journal of Political Science*, *57*, 504–520.

Chan, K. C. G., Yam, S. C. P., & Zhang, Z. (2016). Globally efficient non-parametric inference of average treatment effects by empirical balancing calibration weighting. *The Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, *78*, 673–700.

Deville, J.-C., & Särndal, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, *87*, 376–382.

Fan, J., Imai, K., Liu, H., Ning, Y., & Yang, X. (2016). *Improving covariate balancing propensity score: A doubly robust and efficient approach. Technical report*. Princeton University.

Fong, C., Hazlett, C., & Imai, K. (2018). Covariate balancing propensity score for a continuous treatment: Application to the efficacy of political advertisements. *The Annals of Applied Statistics*, *12*, 156–177.

Fu, W. J. (1998). Penalized regressions: The bridge versus the lasso. *Journal of Computational and Graphical Statistics*, *7*, 397–416.

Hainmueller, J. (2012). Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis*, *20*, 25–46.

Han, P. (2016). Intrinsic efficiency and multiple robustness in longitudinal studies with drop-out. *Biometrika*, *103*, 683–700.

Hansen, L. P., Heaton, J., & Yaron, A. (1996). Finite-sample properties of some alternative gmm estimators. *Journal of Business and Economic Statistics*, *14*, 262–280.

Hernán, M. A., & Robins, J. M. (2006). Estimating causal effects from epidemiological data. *Journal of Epidemiology and Community Health*, *60*, 578–586.

Hirano, K., & Imbens, G. W. (2001). Estimation of causal effects using propensity score weighting: An application to data on right heart catheterization. *Health Services and Outcomes Research Methodology*, *2*, 259–278.

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

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

Imai, K., & Ratkovic, M. (2014). Covariate balancing propensity score. *The Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, *76*, 243–263.

Imai, K., & Ratkovic, M. (2015). Robust estimation of inverse probability weights for marginal structural models. *Journal of the American Statistical Association*, *110*, 1013–1023.

Kallus, N., & Santacatterina, M. (2018). Optimal balancing of time-dependent confounders for marginal structural models. *arXiv preprint arXiv:1806.01083*.

Kang, J. D., & Schafer, J. L. (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science*, *22*, 523–539.

Kott, P. S., & Liao, D. (2012). Providing double protection for unit nonresponse with a nonlinear calibration-weighting routine. *Survey Research Methods*, *6*, 105–111.

Lunceford, J. K., & Davidian, M. (2004). Stratification and weighting via the propensity score in estimation of causal treatment effects: A comparative study. *Statistics in Medicine*, *23*, 2937–2960.

Ridgeway, G., & McCaffrey, D. F. (2007). Comment: Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science*, *22*, 540–543.

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

Robins, J., Sued, M., Lei-Gomez, Q., & Rotnitzky, A. (2007). Comment: Performance of double-robust estimators when "inverse probability" weights are highly variable. *Statistical Science*, *22*, 544–559.

Robins, J. M. (1999). Association, causation, and marginal structural models. *Synthese*, *121*, 151–179.

Robins, J. M. (2000). *Marginal structural models versus structural nested models as tools for causal inference*. Springer.

Robins, J. M., Hernan, M. A., & Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology*, *7*, 550–560.

Robins, J. M., & Rotnitzky, A. (1995). Semiparametric efficiency in multivariate regression models with missing data. *Journal of the American Statistical Association*, *90*, 122–129.

Robins, J. M., Rotnitzky, A., & Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, *89*, 846–866.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*, 41–55.

Scharfstein, D. O., Rotnitzky, A., & Robins, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association*, *94*, (448), 1096–1120.

Seaman, S. R., & White, I. R. (2013). Review of inverse probability weighting for dealing with missing data. *Statistical Methods in Medical Research*, *22*, 278–295.

Tsiatis, A. (2007). *Semiparametric theory and missing data*. Springer Science & Business Media.

van der Laan, M. J., & Robins, J. M. (2003). *Unified methods for censored longitudinal data and causality*. Springer.

Wang, Y., & Zubizarreta, J. R. (2020). Minimal dispersion approximately balancing weights: Asymptotic properties and practical considerations. *Biometrika*, *107*, 93–105.

Wu, C. (2003). Optimal calibration estimators in survey sampling. *Biometrika*, *90*, 937–951.

Yiu, S., & Su, L. (2018). Covariate association eliminating weights: A unified weighting framework for causal effect estimation. *Biometrika*, *105*, 709–722.

Zubizarreta, J. R. (2015). Stable weights that balance covariates for estimation with incomplete outcome data. *Journal of the American Statistical Association*, *110*, 910–922.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.