Robust Estimation of Inverse Probability Weights for Marginal Structural Models*

Kosuke Imai† Marc Ratkovic‡

First Draft: February 22, 2013
This Draft: May 28, 2013

Abstract

Marginal structural models (MSMs) are becoming increasingly popular among applied researchers as a tool to make causal inference from longitudinal data. Unlike standard regression models, MSMs can adjust for time-dependent observed confounders while avoiding post-treatment bias. Despite their theoretical appeal, a main practical challenge of MSMs is the difficulty in estimating inverse probability weights. Previous studies have found that MSMs can be highly sensitive to model misspecification of treatment assignment model even when the number of time periods is moderate. The effect of misspecification can propagate across time periods because inverse probability weights used for MSMs are typically based on the product of propensity score estimated separately at each time period. To address this problem, we generalize the Covariate Balancing Propensity Score (CBPS) methodology of Imai and Ratkovic (2013), which estimates the inverse probability weights such that the resulting covariate balance is optimized. The proposed methodology incorporates all covariate balancing conditions associated with inverse probability weights. Since the number of these conditions grows exponentially as the number of time period increases, we orthogonalize them to make estimation feasible in this high-dimensional setting. Our small scale simulation study suggests that the CBPS significantly improves the empirical performance of MSMs by making the treatment assignment model robust to misspecification. Open-source software is available for implementing the proposed methods.

Key Words: causal inference, covariate balancing propensity score, inverse propensity score weighting, observational studies, sequential ignorability, time-dependent treatments

*The proposed methodology can be implemented via open-source software `CBPS` (Ratkovic *et al.*, 2012), which is freely available as an R package at the Comprehensive R Archive Network (CRAN http://cran.r-project.org/package=CBPS). We thank seminar participants at the University of Michigan (Economics Department) and the Atlantic Causal Inference Conference (Harvard University) for helpful suggestions.

†Associate Professor, Department of Politics, Princeton University, Princeton NJ 08544. Phone: 609–258–6601, Email: kimai@princeton.edu, URL: http://imai.princeton.edu

‡Lecturer, Department of Politics, Princeton University, Princeton NJ 08544. Phone: 608–658–9665, Email: ratkovic@princeton.edu, URL: http://www.princeton.edu/∼ratkovic

1 Introduction

Since its introduction by Robins (1999), marginal structural models (MSMs) have quickly gained popularity among applied researchers in biomedical and other fields as a tool for making causal inference from longitudinal data in observational studies. The paper that popularized MSMs in the field of epidemiology have more than 1,000 Google Citations as of May 2013 (Robins *et al.*, 2000) and the method has been introduced to other disciplines (e.g., Blackwell, 2013). As explained by Robins *et al.* (2000), when estimating the causal effects of time-varying treatments, standard regression models fail to appropriately adjust for time-dependent observed confounders that are affected by previous treatments. In contrast, MSMs allow one to estimate the causal effects of different treatment sequences while avoiding this post-treatment bias.

Despite their theoretical appeal, a main practical challenge of MSMs is the difficulty in estimating inverse probability weights. Using simulation and empirical studies, a number of previous studies have found that MSMs can be highly sensitive to model misspecification of treatment assignment model even when the number of time periods is moderate (e.g., Cole and Hernán, 2008; Howe *et al.*, 2011; Kang and Schafer, 2007; Lefebvre *et al.*, 2008; Mortimer *et al.*, 2005). The effect of misspecification can propagate across time periods because the inverse probability weights used for MSMs are typically based on the product of propensity score estimated separately at each time period.

To address this problem, we introduce the Covariate Balancing Propensity Score (CBPS) methodology as a robust estimation method for inverse probability weights of MSMs. The CBPS estimates the inverse probability weights such that the resulting covariate balance is optimized. The idea was first introduced by Imai and Ratkovic (2013) to improve the estimation of propensity score in the cross section settings. In this paper, we generalize the CBPS methodology to the longitudinal data. After briefly reviewing MSMs and their assumptions (Section 2), we describe the proposed methodology (Section 3). The proposed methodology incorporates all covariate balancing conditions associated with inverse probability weights. Since the number of these conditions grows exponentially as the number of time period increases, we orthogonalize them to make estimation feasible in this high-dimensional setting. We then present simulation studies, which suggest that the CBPS can dramatically improve the empirical performance of MSMs when the treatment assignment model is misspecified (Section 4). The final section gives concluding remarks and discusses future research agenda.

2 A Review of Marginal Structural Models

In this section, we briefly review the marginal structural models (MSMs) of Robins (1999). See Robins *et al.* (2000) for a more detailed introduction of MSMs. Suppose that we have a simple random sample of size n from a population. For each unit, repeated measurements are taken at each of J time periods. Specifically, at each time period $j = 1, 2, \ldots, J$, we observe the time-dependent treatment variable T_{ij} as well as the time-dependent confounders X_{ij} that are possibly affected by previous treatments. We assume that X_{ij} is realized before the treatment at time j and therefore is not affected by T_{ij}. We further assume that the treatment variable is binary where $T_{ij} = 1$ ($T_{ij} = 0$) implies unit i receives (does not receive) the treatment at time j. Next, for each unit, we denote the treatment and covariate history up to time j by $\overline{T}_{ij} = \{T_{i1}, T_{i2}, \ldots, T_{ij}\}$ and $\overline{X}_{ij} = \{X_{i1}, X_{i2}, \ldots, X_{ij}\}$, respectively. We also denote the set of possible treatment and covariate values at time j as $\overline{\mathcal{T}}_j$ and $\overline{\mathcal{X}}_j$. Finally, we observe the outcome variable Y_i for unit i at the end of the study, i.e., time J, after the treatment for the same time period, i.e., T_{iJ}, is administered.

The potential outcome framework of causal inference was originally developed by Neyman (1923) and Rubin (1973) in the cross-section setting, but Robins (1986) generalized it to the longitudinal analysis. Under this framework, we use $Y_i(\bar{t}_J)$ to represent the potential value of the eventual outcome variable for unit i measured at time J under the entire treatment history $\overline{T}_{iJ} = \bar{t}_J$ where $\bar{t}_J \in \overline{\mathcal{T}}_J$. Thus, the observed outcome is given by $Y_i = Y_i(\overline{T}_J)$. Similarly, $X_{ij}(\bar{t}_{j-1})$ denotes the potential values of covariates for unit i at each time period j under the treatment history up to time $j-1$, i.e., $\overline{T}_{i,j-1} = \bar{t}_{j-1}$. Therefore, the observed values of covariates can be written as $X_{ij} = X_{ij}(\overline{T}_{i,j-1})$ for unit i at time j. This setup relies upon the consistency assumption that the potential values of outcome and covariates for each unit are only functions of its own treatment history up to that point in time. The assumption excludes the possible interference between units (but not between time periods), implying that the potential values of outcome and covariates are not influenced by the treatment history of other units.

MSMs are based on the assumption of sequential ignorability, which states that the treatment assignment of unit i at time j is exogenous given the treatment and covariate history of the same unit up to that point in time. In other words, MSMs assume no unmeasured confounding at each time period. This sequential ignorability assumption can be formally written as,

$$Y_i(\bar{t}_J) \perp\!\!\!\perp T_{ij} \quad | \quad \overline{T}_{i,j-1} = \bar{t}_{j-1}, \overline{X}_{ij} = \bar{x}_j \tag{1}$$

at any time period j for a given treatment history $\bar{t}_J = \{\bar{t}_{j-1}, t_j, \ldots, t_J\} \in \overline{\mathcal{T}}_J$ and covariate history

$\bar{x}_j \in \overline{\mathcal{X}}_j$. We also assume that the conditional probability of treatment assignment is bounded away from zero and one at each time period. That is,

$$0 \ < \ \Pr(T_{ij} = 1 \mid \overline{T}_{i,j-1} = \bar{t}_{j-1}, \overline{X}_{ij} = \bar{x}_j) \ < \ 1 \tag{2}$$

at any time period j for a given treatment history $\bar{t}_{j-1} \in \overline{\mathcal{T}}_{j-1}$ and covariate history $\bar{x}_j \in \overline{\mathcal{X}}_j$.

Under these assumptions, Robins (1999) showed that the inverse-probability-of-treatment weighting (IPTW) can be used to consistently estimate the marginal mean of any potential outcome, i.e., $\mathbb{E}\{Y_i(\bar{t}_J)\}$ for any treatment sequence $\bar{t}_J \in \mathcal{T}_J$. For the reason that will become clear later, we first define the potential value of this weight for unit i under treatment history $\bar{t}_J$ as,

$$w_i(\bar{t}_J, \overline{X}_{iJ}(\bar{t}_{J-1})) \ = \ \frac{1}{P(\overline{T}_{iJ} = \bar{t}_J \mid \overline{X}_{iJ}(\bar{t}_{J-1}))} \ = \ \prod_{j=1}^{J} \frac{1}{P(T_{ij} = t_{ij} \mid \overline{T}_{i,j-1} = \bar{t}_{j-1}, \overline{X}_{ij}(\bar{t}_{j-1}))} \tag{3}$$

This weight is typically small and therefore the estimates become highly variable. Therefore, researchers commonly follow the suggestion given in the literature and use the stabilized weights of the form, $w_i^*(\bar{t}_J, \overline{X}_{iJ}(\bar{t}_{J-1})) = P(\overline{T}_{iJ} = \bar{t}_J)/P(\overline{T}_{iJ} = \bar{t}_J \mid \overline{X}_{iJ}(\bar{t}_{J-1}))$, when fitting the outcome model. We denote the observed values of these weights as $w_i = w_i(\overline{T}_{iJ}, \overline{X}_{iJ})$ and $w_i^* = w_i^*(\overline{T}_{iJ}, \overline{X}_{iJ})$.

In an observational study, these weights are unknown and must be estimated. Typically, a parametric model is used to estimate the conditional probability of treatment assignments given the set of covariates,

$$w_i^{-1} \ = \ \pi_\beta(\overline{T}_{iJ}, \overline{X}_{iJ}) \tag{4}$$

where β is a finite dimensional vector of unknown parameters. A common choice of parametric model is the logistic regression independently applied to each time period,

$$\pi_\beta(\overline{T}_{iJ}, \overline{X}_{iJ}) \ = \ \prod_{j=1}^{J} \text{expit}\{(2T_{ij} - 1)\beta_j^\top \overline{X}_{ij}^*\} \tag{5}$$

where $\overline{X}_{ij}^* = [\overline{T}_{i,j-1} \ \overline{X}_{ij}]$, $\text{expit}(z) = \{1 + \exp(-z)\}^{-1}$, and β_j is a vector of unknown coefficients. The numerator of the stabilized weight is typically estimated using the sample proportion for each treatment sequence.

Once the (stabilized) weights are estimated, the conditional expectation of outcome is modeled as a function of treatment history alone without covariates, i.e., $\mathbb{E}(Y_i \mid \overline{T}_{iJ})$. Robins (1999) has shown that this yields a consistent estimate of the mean potential outcome, i.e., $\mathbb{E}\{Y_i(\bar{t}_J)\}$ thereby allowing researchers to compute the average outcome under any sequence of treatment assignments over time.

3 The Proposed Methodology

In this section, we propose a robust estimation procedure for the inverse-probability-treatment weight w_i for MSMs. Specifically, we estimate the weight such that time-dependent covariates are balanced across appropriate sub-populations by generalizing the the covariate balancing propensity score (CBPS) of Imai and Ratkovic (2013) to the longitudinal data settings. The CBPS methodology estimates the propensity score such that the resulting covariate balance is optimized. Imai and Ratkovic (2013) present the simulation and empirical evidence that the CBPS significantly improves the robustness of propensity score matching and weighting methods to the model misspecification of treatment assignment model. We show how to extend the CBPS to the causal analysis of panel data with time-dependent treatments and confounders. In what follows, for the ease of exposition, we first present the proposed methodology in the case of two time periods. We then consider the general case of possibly more than two time periods.

3.1 The Two Time Period Case

To convey the intuition for the proposed methodology, we first present the CBPS for the case of two time periods. For each unit i, we observe the outcome variable Y_i measured at the end of study, the binary treatment variable T_{ij}, and a vector of confounders X_{ij} for each time period $j = 1, 2$. Suppose that we are interested in using MSMs to estimate the marginal mean of potential outcome measured at the end of the second period, $\mathbb{E}\{Y_i(\bar{t}_2)\}$, where $\bar{t}_2$ can take any of the four possible values, i.e., $\bar{t}_2 \in \mathcal{T}_2 = \{(0,0),(0,1),(1,0),(1,1)\}$.

Covariate Balancing Conditions. We derive the moment conditions based on the covariate balancing property of the weight for MSMs. To do this, we first express these moment conditions as functions of the (potential) weight defined in equation (3). Specifically, at the first time period, across all possible treatment histories, the weight should balance the mean of the baseline covariate, X_{i1}. Formally, for all $\bar{t}_2 = (t_1, t_2) \in \{(0,0),(0,1),(1,0),(1,1)\}$, we have

$$\mathbb{E}(X_{i1}) \quad = \quad \mathbb{E}\left[\mathbf{1}\{T_{i1} = t_1, T_{i2} = t_2\} w_i(\bar{t}_2, \overline{X}_{i2}(t_1)) X_{i1}\right]. \tag{6}$$

This gives the total of three moment conditions because the above equality holds across four different treatment histories and one such equality is redundant. While there exist numerous ways to represent these three moment conditions, we choose the following equalities that are orthogonal to each other

Time period	Treatment history: (t_1, t_2) (0,0)	(0,1)	(1,0)	(1,1)	Moment condition
time 1	$+$	$+$	$-$	$-$	$\mathbb{E}\left\{(-1)^{T_{i1}} w_i X_{i1}\right\} = 0$
	$+$	$-$	$+$	$-$	$\mathbb{E}\left\{(-1)^{T_{i2}} w_i X_{i1}\right\} = 0$
	$+$	$-$	$-$	$+$	$\mathbb{E}\left\{(-1)^{T_{i1}+T_{i2}} w_i X_{i1}\right\} = 0$
time 2	$+$	$-$	$+$	$-$	$\mathbb{E}\left\{(-1)^{T_{i2}} w_i X_{i2}\right\} = 0$
	$+$	$-$	$-$	$+$	$\mathbb{E}\left\{(-1)^{T_{i1}+T_{i2}} w_i X_{i2}\right\} = 0$

Table 1: Orthogonal Representation of Covariate Balancing Moment Conditions in the Two Time Period Case. The first and second time periods have three and two moment conditions, respectively. There are four distinct values of treatment history with t_j representing the value of the treatment variable at time j. The symbols, "+" and "−", in these four treatment history columns show whether the weighted average of covariates among units with a certain treatment history is added or subtracted when formulating the moment condition. Within each time period, row vectors of +'s and −'s for the treatment history combinations are orthogonal to one another. The last column represents the corresponding moment condition.

and can be written in a compact notation using the observed weight instead of its potential values,

$$\mathbb{E}\{(-1)^{T_{i1}} w_i X_{i1}\} = 0 \tag{7}$$

$$\mathbb{E}\{(-1)^{T_{i2}} w_i X_{i1}\} = 0 \tag{8}$$

$$\mathbb{E}\{(-1)^{T_{i1}+T_{i2}} w_i X_{i1}\} = 0 \tag{9}$$

This orthogonal representation of covariate balancing conditions is summarized in the first three rows of Table 1. In the table, if we treat $+$ and $-$ as $+1$ and -1 in the table, row vectors for each time period are orthogonal to each other.

The covariate balancing conditions at the second time period are similar to those at time 1, except that the covariates measured at time 2 are a function of the treatment at time 1, i.e., $X_{i2} = X_{i2}(T_{i1})$. Using the potential outcomes notation, for all $\bar{t}_2 = \{t_1, t_2\}$, we can write these covariate balancing conditions as follows,

$$\mathbb{E}\{X_{i2}(t_1)\} = \mathbb{E}\left[\mathbf{1}\{T_{i1} = t_1, T_{i2} = t_2\} w_i(\bar{t}_2, \overline{X}_{i2}(t_1)) X_{i2}(t_1)\right] \tag{10}$$

Because $X_{i2}(t_1)$ is observed only when $T_{i1} = t_1$, the above covariate balancing equation implies the following two sets of moment conditions, i.e., one for each value of t_1, which again can be compactly written as,

$$\mathbb{E}\left\{(-1)^{T_{i2}} w_i X_{i2}\right\} = 0 \tag{11}$$

$$\mathbb{E}\left\{(-1)^{T_{i1}+T_{i2}} w_i X_{i2}\right\} = 0 \tag{12}$$

The bottom two rows of Table 1 summarize this result. While at time 1 both T_{i1} and T_{i2} are varied to generate 3 moment conditions, only T_{i2} is varied at time 2.

The benefits of this orthogonalization is twofold. First, it reduces the statistical dependence between moment conditions, simplifying the estimation and improving the efficiency of the resulting weight. Second, as shown in Section 3.2, its compact notation allows one to extend the proposed methodology to the general case of more than two time periods.

Estimation. Since the number of moment conditions exceeds the number of parameters to be estimated, we use the generalized method of moments (GMM; Hansen, 1982) estimation to combine the covariate balancing conditions derived above. Our GMM estimator is given by,

$$\begin{aligned}\hat{\beta} &= \underset{\beta\in\Theta}{\operatorname{argmin}}\ \operatorname{vec}(\mathbf{G})^\top\{\mathbf{I}_3\otimes\mathbf{W}\}^{-1}\operatorname{vec}(\mathbf{G}) \qquad (13)\\ &= \underset{\beta\in\Theta}{\operatorname{argmin}}\ \operatorname{trace}(\mathbf{G}^\top\mathbf{W}^{-1}\mathbf{G}) \qquad (14)\end{aligned}$$

where the sample moment conditions are given by,

$$\mathbf{G} = \frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix}(-1)^{T_{i1}}w_iX_{i1} & (-1)^{T_{i2}}w_iX_{i1} & (-1)^{T_{i1}+T_{i2}}w_iX_{i1}\\ 0 & (-1)^{T_{i2}}w_iX_{i2} & (-1)^{T_{i1}+T_{i2}}w_iX_{i2}\end{bmatrix}, \qquad (15)$$

and $\mathbf{W}$ is given by,

$$\mathbf{W} = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\begin{array}{cc|c}w_i^2X_{i1}X_{i1}^\top & w_i^2X_{i1}X_{i2}^\top & \\ w_i^2X_{i2}X_{i1}^\top & w_i^2X_{i2}X_{i2}^\top & \end{array}\ X_{i1},X_{i2}\right] \qquad (16)$$

where the expectation is calculated analytically.

It is important to note that this formulation rests upon a simplifying assumption in the GMM weighting matrix structure. In particular, we assume that different covariate balancing conditions are uncorrelated with each other. The assumption implies that each column of $\mathbf{G}$ in equation (15) is uncorrelated with each other. While these assumptions do not generally hold, the orthogonalization of covariate balancing conditions reduces the possible correlation. The main motivation for this simplifying assumption is computational scalability. As shown in Section 3.2, this formulation prevents the dimension of $\mathbf{W}$ from increasing exponentially as the number of time periods increases. While it is known that the optimal weighting matrix equals the inverse of the covariance matrix of all moment conditions, this assumption does not affect the consistency of the GMM estimator.

3.2 The General Case

We extend the above formulation to the general case with more than two time periods, i.e., $j = 1, 2, \ldots, J$. We first generalize the covariate balancing conditions derived above and then describe the GMM estimation.

Covariate Balancing Conditions. We characterize the covariate balancing conditions in the general case with an arbitrary number of time periods $J \geq 2$. Recall that in the two time period case, the weight for MSMs balances the covariates at the first time period across all possible values of the entire treatment vector. At the second time period, however, the weight only balances covariates across the treatment values at that time period among the units who receive the same treatment value in the first time period. In general, the weight balances covariates at a given time period across all possible combinations of the current and future treatment conditions, but not across the past treatment combinations. Formally, for a given time period j and fixed past treatment sequence up to that point $\bar{t}_{j-1}$, we can write the covariate balancing conditions across all treatment sequences of the current and future time periods $\underline{t}_j = \{t_j, t_{j+1}, \ldots, t_J\}$ as,

$$\mathbb{E}\{X_{ij}(\bar{t}_{j-1})\} \quad = \quad \mathbb{E}[\mathbf{1}\{\overline{T}_{j-1} = \bar{t}_{j-1}, \underline{T}_{ij} = \underline{t}_j\} w_i(\bar{t}_J, \overline{X}_{iJ}(\bar{t}_{J-1})) X_{ij}(\bar{t}_{j-1})] \tag{17}$$

where $\underline{T}_{ij} = \{T_{ij}, T_{i,j+1}, \ldots, T_{iJ}\}$ represents a vector of observed current and future treatment conditions.

In the two time period case, the balance conditions are characterized in terms of the sums and differences of $w_i X_{ij}$ across all groups defined by a distinct value of the entire treatment sequence. We generalize that formulation here. Specifically, for each time period, we orthogonalize the covariate balancing conditions given in equation (17) by aliasing the past treatment effects on the covariates at time j. Since there exist a total of 2^{J-j+1} possible current and future treatments, equation (17) implies $2^{J-j+1} - 1$ orthogonal constraints given a particular history of treatment up to time $j-1$, i.e., $\bar{t}_{j-1}$. There are a total of 2^{j-1} possible treatment histories and hence all together we have $(2^J - 2^{j-1})$ covariate balancing conditions for each time period j.

To formalize this idea, we utilize the theoretical framework developed for analyzing and designing randomized experiments based on the 2^J full factorial design (see e.g., Box *et al.*, 2005). In Table 2, we present a running example of the case with $J = 3$ where the first three columns present the design matrix in Yates order with $+$'s and $-$'s indicating the presence and absence of the treatment at each time period, respectively. It is well known that the full 2^J factorial design can be represented by Hadamard matrix of order 2^J. Recall that Hadamard matrix of order n, denoted by $\mathbf{H}_n$, is an $n \times n$ matrix of $+1$'s and -1's whose rows are orthogonal to one another, implying that $\mathbf{H}_n^\top \mathbf{H}_n = n\mathbf{I}_n$.

To construct a Hadamard matrix that corresponds to the full 2^J factorial design, let $\mathbf{D}$ be the $2^J \times J$ "negative" design matrix of $+1$'s and -1's sorted in Yates order,

$$\mathbf{D} \quad = \quad [d_0, d_1, d_2, d_{12}, d_3, d_{13}, d_{23}, d_{123}, d_4, d_{14}, \ldots]^\top \tag{18}$$

Design matrix			Treatment history: (t_1, t_2, t_3)								Time periods		
			(0,0,0)	(1,0,0)	(0,1,0)	(1,1,0)	(0,0,1)	(1,0,1)	(0,1,1)	(1,1,1)			
T_{i1}	T_{i2}	T_{i3}	h_0	h_1	h_2	h_{12}	h_{13}	h_3	h_{23}	h_{123}	1	2	3
−	−	−	+	+	+	+	+	+	+	+	✗	✗	✗
+	−	−	+	−	+	−	+	−	+	−	✓	✗	✗
−	+	−	+	+	−	−	+	+	−	−	✓	✓	✗
+	+	−	+	−	−	+	+	−	−	+	✓	✓	✗
−	−	+	+	+	+	+	−	−	−	−	✓	✓	✓
+	−	+	+	−	+	−	−	+	−	+	✓	✓	✓
−	+	+	+	+	−	−	−	−	+	+	✓	✓	✓
+	+	+	+	−	−	+	−	+	+	−	✓	✓	✓

Table 2: Orthogonal Representation of Covariate Balancing Moment Conditions in the Three Time Period Case Using the 2^3 Factorial Experiment Framework. The first three columns show the design matrix of the factorial experiment in Yates order where the symbol "+" ("−") represents the presence (absence) of each treatment factor. The next eight columns show the Hadamard matrix of this factorial experiment based on this design matrix that corresponds to the eight distinct values of treatment history with t_j representing the value of the treatment variable at time j. The symbols, "+" and "−", in these eight treatment history columns also indicate the orthogonal representation of covariate balancing moment conditions. Finally, the symbol ✓(✗) in the last three columns indicates that the corresponding covariate balancing moment condition is (not) binding for each time period.

where d_0 is a J dimensional column vector of 1's and d_j is a column vector of length J where the elements of set j indicate the indexes of the elements of the vector with -1 and the other elements of the vector are 1's. For example, when $J = 3$, we have $d_{12} = (-1, -1, 1)^\top$. Thus, +'s and −'s in Table 2 correspond to -1's and $+1$'s in $\mathbf{D}$, respectively. Let c_j be the jth column of $\mathbf{D}$ so that $\mathbf{D} = [c_1, c_2, \ldots, c_J]$. Further, denote the common component of d_j and c_k by d_{jk}. For a subset t of $\mathbb{N}_J = \{1, \ldots, J\}$, let the Hadamard product, denoted by h_t, of columns c_k with $k \in t$ be a 2^J dimensional column vector with its jth element being $\prod_{k \in t} d_{jk}$. Then, the Hadamard matrix of order 2^J can be constructed by collecting in Yates order all the Hadamard products of the columns of $\mathbf{D}$. The result is given by the following $2^J \times 2^J$ matrix,

$$\mathbf{H}_{2^J} \quad = \quad [h_0, h_1, h_2, h_{12}, h_3, h_{13}, h_{23}, h_{123}, h_4, h_{14}, \ldots] \tag{19}$$

where h_0 is a column vector of +1's. This matrix in the case of $J = 3$ is given in the middle columns of Table 2.

Thus, the Hadamard matrix representation allows us to orthogonalize the covariate balancing moment conditions in a systematic way regardless of the number of time periods. Moreover, the successive multiplication procedure used for the construction of this Hadamard matrix directly justifies the notation used in equations (11) and (12). In fact, it has long been known that this Hadamard matrix representation can be used to compute the mod 2 discrete Fourier transform (Good, 1958).

For example, the second and sixth rows of Table 2 corresponds to the following covariate moment conditions,

$$\mathbb{E}\{(-1)^{T_{i1}} w_i X_{ij}\} = 0 \tag{20}$$

$$\mathbb{E}\{(-1)^{T_{i1}+T_{i3}} w_i X_{ij}\} = 0 \tag{21}$$

That is, one can use the design matrix to form the treatment variables that enter the exponent of -1 in the compact expression of the covariate balancing moment conditions. In sum, the 2^J factorial experiment framework allows us to directly generalize the orthogonal representation of the covariate moment conditions given in Section 3.1 to the general case with more than two time periods.

Therefore, this full 2^J factorial design framework clearly shows which covariate balancing moment conditions are binding at any given time period for the estimation of the weight for MSMs. As noted above, the stabilized weight balances covariates measured at time j across all possible current and future treatments but it does not balance across past treatments. Therefore, the covariate balancing moment conditions, which correspond to the effects of past treatments and their interactions, are not binding. These conditions can be easily identified by the design matrix. For example, in Table 2, we see that the second row, corresponding to the main effect of time 1 treatment, i.e., T_{i1}, is not binding for time 2 covariates X_{i2}. Similarly, for time 3 covariates, the moment conditions corresponding to the effects of T_{i1} and T_{i2} as well as their interaction are not binding. In general, for covariates measured at time j, the first 2^{j-1} rows of Hadamard matrix $\mathbf{H}_{2^J}$ can be ignored when constructing the covariate balancing moment conditions.

Estimation. As in the two time period case, we use the GMM to combine all the covariate balancing conditions. Our GMM estimator is given in equation (14) where the expression for the vector of covariate balancing conditions $\mathbf{G}$ and their sample covariance matrix $\mathbf{W}$ in the general case are constructed as follows. We begin by defining the following two matrices,

$$\widetilde{\mathbf{X}}_j = \begin{bmatrix} w_1 X_{1j}^\top \\ w_2 X_{2j}^\top \\ \vdots \\ w_n X_{nj}^\top \end{bmatrix} \quad \text{and} \quad \mathbf{R}_j = \begin{bmatrix} \mathbf{0}_{2^{j-1}\times 2^{j-1}} & \mathbf{0}_{2^{j-1}\times(2^J-2^{j-1})} \\ \mathbf{0}_{(2^J-2^{j-1})\times 2^{j-1}} & \mathbf{I}_{2^J-2^{j-1}} \end{bmatrix} \tag{22}$$

where $\widetilde{\mathbf{X}}_j$ represents the matrix of weighted time-dependent covariates and $\mathbf{R}_j$ is the "selection" matrix which identifies the binding covariate balancing conditions for each time period. Next, we construct the $n \times (2^J - 1)$ model matrix $\mathbf{M}$ based on the design matrix $\mathbf{D}$ arranged in Yates' order as follows,

$$\mathbf{M} \;=\; [m_0, m_1, m_2, m_{12}, m_3, m_{13}, m_{23}, m_{123}, m_4, m_{14}, \ldots] \tag{23}$$

where m_0 is a n dimensional column vector of ones, and the ith element of an $n \times 1$ vector m_t is $(-1)^{\sum_{k \in t} T_{ik}}$ with $t \in \{1, 2, 12, 3, 13, 23, 123, 4, 14, \ldots\}$. For example, the ith element of m_{23} equals $(-1)^{T_{i2}+T_{i3}}$. In fact, the ith row of M is given by the row of the Hadamard matrix in Table 2 that corresponds to the treatment sequence of the ith observation. Finally, the set of balancing conditions and their sample covariance matrix are given by

$$\mathbf{G} \;=\; \begin{bmatrix} \widetilde{\mathbf{X}}_1^\top \mathbf{M}\mathbf{R}_1 \\ \vdots \\ \widetilde{\mathbf{X}}_J^\top \mathbf{M}\mathbf{R}_J \end{bmatrix} \quad \text{and} \quad \mathbf{W} \;=\; \begin{bmatrix} \mathbb{E}(\widetilde{\mathbf{X}}_1 \widetilde{\mathbf{X}}_1^\top \mid \mathbf{X}) & \cdots & \mathbb{E}(\widetilde{\mathbf{X}}_1 \widetilde{\mathbf{X}}_J^\top \mid \mathbf{X}) \\ \vdots & \ddots & \vdots \\ \mathbb{E}(\widetilde{\mathbf{X}}_J \widetilde{\mathbf{X}}_1^\top \mid \mathbf{X}) & \cdots & \mathbb{E}(\widetilde{\mathbf{X}}_J \widetilde{\mathbf{X}}_J^\top \mid \mathbf{X}) \end{bmatrix}. \tag{24}$$

where the expectation is calculated analytically. Thus, the estimator derived in Section 3.1 can be directly generalized to the case with an arbitrary number of time periods.

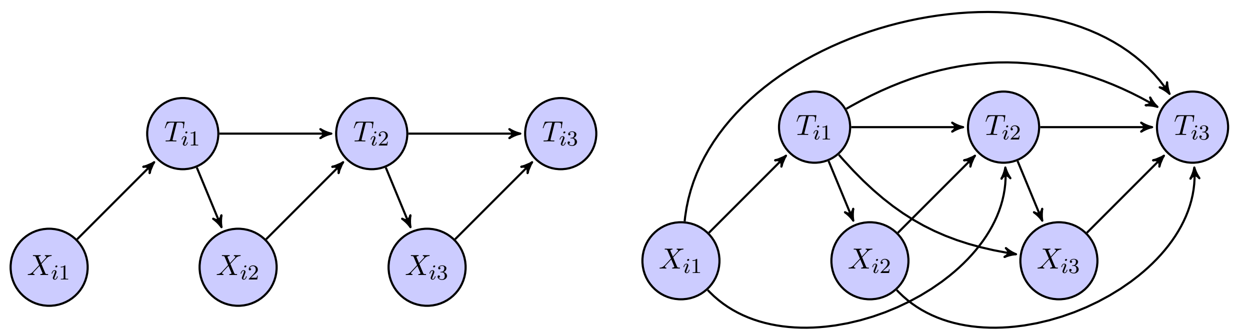
3.3 Combining with the Score Conditions

In the cross section settings, Imai and Ratkovic (2013) suggest that the CBPS may combine the score conditions with the covariate balancing conditions under the GMM framework. The idea is to exploit the dual characteristics of propensity score: if the propensity score is correctly estimated it should predict the treatment assignment *and* balance the covariates. In the current longitudinal settings, we can also take the same strategy by including the score condition from each time period as another set of moment conditions in the GMM objective function given in equation (14).

Since the covariance between the score and covariate balancing conditions is complex, in practice we may assume the independence among them. In our simulations, we account for the dependence within each score condition across time periods using their sample covariance. Specifically, we apply a logistic regression to model the treatment assignment at each time period. As shown in Imai and Ratkovic (2013), in this case, a score condition can be written as another covariate balancing condition with different weights. Thus, we include these score conditions as another set of columns of $\widetilde{\mathbf{X}}_j$ in equation (22).

3.4 Extension to Multiple Binary Treatments

The method described above naturally extends to the setting where there exist multiple binary treatments. Indeed, dynamic treatment regimes considered in this paper is essentially a special case of J multiple binary treatments. The only difference is that for dynamic treatment regimes some



(a) Simulation 1: Correct Lag Structure

(b) Simulation 2: Incorrect Lag Structure

Figure 1: Treatment Variable Data Generating Process in Simulation Studies. In the first set of simulations summarized by the diagram of panel (a), a relatively simple treatment assignment model is used and we only misspecify the functional form while maintaining the correct lag structure. In the second set of simulations summarized by the diagram of panel (b), a more complex data generating process is used and we examine the impact of incorrectly specifying the lag structure. The results of these simulations are given in Figure 2 and 3, respectively

of the covariate balancing conditions are not binding as indicated by zero elements of **G** matrix in equations (15) and (24). In contrast, for multiple binary treatments, all of these covariate balancing conditions are binding. However, aside from this difference, the estimation for the case of multiple binary treatments proceeds in an identical manner.

4 Simulation Studies

We conduct four sets of simulation studies in order to assess the empirical performance of the proposed CBPS estimation. First, we show that when the treatment assignment model is correctly specified, the proposed methodology does as well as the standard maximum likelihood estimation. Second, we also examine several scenarios where the treatment assignment model is misspecified in terms of either the lag structure or the functional form of the covariates (or both). We find that the CBPS significantly reduces the bias and mean squared error of the standard method in each of these model misspecification scenarios.

In all four simulation scenarios, we consider the case of three time periods, i.e., $J = 3$, and use four different sample sizes $n = 500, 1,000, 2,500$, and $10,000$. Across these four simulations, we vary both whether the lag structure and functional form for the treatment assignment model are properly modeled. Figure 1 summarizes the treatment variable data generating processes used in our simulations. In the first set of simulations summarized by the diagram of panel (a), a relatively simple treatment assignment model is used and we only misspecify the functional form while maintaining the correct lag structure. In practice, however, both the treatment variables and the covariates may be affected by the previous treatment. In the second set of simulations summarized by the diagram of panel (b), a more complex data generating process is used and we examine the impact of incorrectly specifying the lag structure. All simulations use the identical outcome variable model.

Specifically, in the first set of simulations, for time j, we use the covariates $X_{ij} = (Z_{ij1} \cdot U_{ij}, Z_{ij2} \cdot U_{ij}, |Z_{ij3} \cdot U_{ij}|, |Z_{ij4} \cdot U_{ij}|)^\top$ where each Z_{ijk} is an i.i.d. draw from the standard normal distribution, and U_{ij} is constructed as $U_{ij} = 2 + (2T_{i,j-1} - 1)/3$ for $j = 2, 3$ and $U_{ij} = 1$ for $j = 1$. The treatment assignment model is given by $\Pr(T_{ij} = 1) = \text{expit}\{-T_{i,j-1} + \gamma^\top X_{ij} + (-1/2)^j\}$ where $\gamma = (1, -0.5, 0.25, 0.1)^\top$ and $T_{i0} = 0$. Finally, the outcome model is defined as $Y_i = 250 - 10 \cdot \sum_{j=1}^{3} T_{ij} + \sum_{j=1}^{3} \delta^\top X_{ij} + v_i$ where $\delta = (27.4, 13.7, 13.7, 13.7)^\top$ and v_i is a normal disturbance with mean zero and standard deviation five. To consider the functional form misspecification, we use the following non-linear transformation of the covariates, $X_{ij}^* = (Z_{ij1}^3, 6 \cdot Z_{ij2}, \log(|Z_{ij3}|), 1/|Z_{ij4}|)^\top$, and estimate the treatment assignment model with these covariates.

In the second set of simulations, we consider the misspecification of lag structure where only the covariates from the current period and the treatment from the most immediately previous time period are used. Recall that in these simulations we generate the treatment at any given time period as a function of treatments and covariates from all previous time periods. As with the first two simulations, we also consider the misspecification of functional forms using nonlinear transformation. Specifically, the treatment assignment in the second set of simulations is given by $\Pr(T_{ij} = 1) = \text{expit}\{\sum_{j'=1}^{j} (T_{i,j-1} + \gamma^\top X_{ij'})/2^{j-j'} + (-1/2)^j\}$. The true treatment assignment model is a function of the entire covariate and treatment history for each observation, but each method is applied using the most immediate covariates and treatment. In order to generate our covariates for this set of simulations, we adjust U_{ij} such that $U_{ij} = \prod_{j'=1}^{j-1} \left\{2 + (2T_{ij'} - 1)/3\right\}$ for $j = 2, 3$ and $U_{ij} = 1$ for $j = 1$. The new set of covariates are then constructed as $X_{ij} = (Z_{ij1}U_{ij}, Z_{ij2}U_{ij}, |Z_{ij3}U_{ij}|, |Z_{ij4}U_{ij}|)^\top$ so that they are a function of all past treatments. The outcome model is the same as the one used for the first set of simulations except that the definition of X_{ij} is different. As before, we assess each methods' performance when using the correct covariates, X_{ij}, and the covariates after a mild nonlinear transformation, $X_{ij}^* = \{(Z_{ij1}U_{ij})^3, 6 \cdot Z_{ij2}U_{ij}, \log|Z_{ij3}U_{ij}|, 1/|Z_{ij4}U_{ij}|\}^\top$.

To evaluate the performance of our proposed CBPS methodology, we simulate 10,000 data sets using the aforementioned data generating processes. We then fit a logistic regression model (GLM) as the treatment assignment model independently for each time period using correct and incorrect model specifications as discussed above. We also fit the same exact logistic model using the proposed CBPS estimation but in two ways: first with covariate balancing conditions alone (CBPS1) and second with

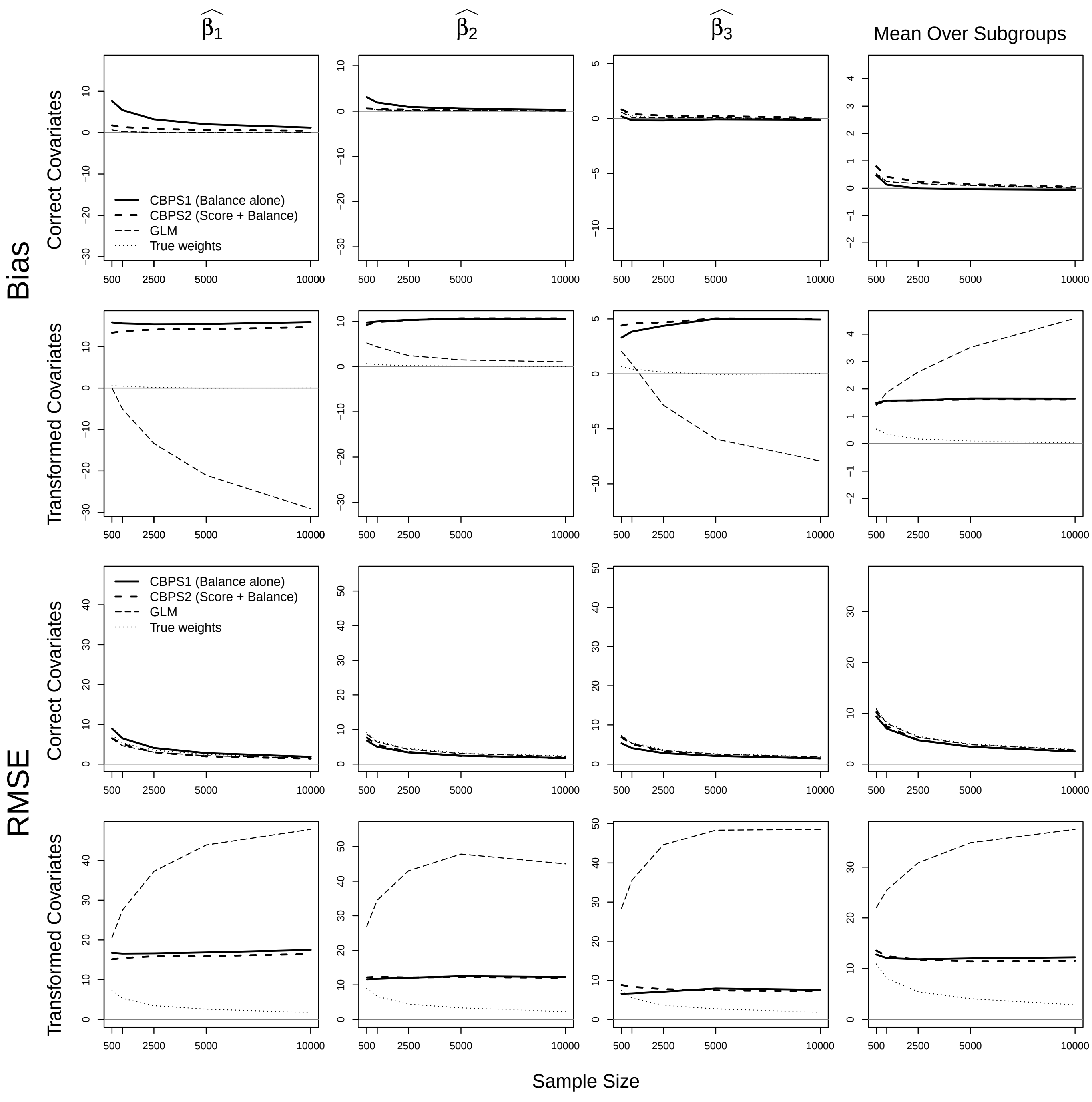


Figure 2: Impact of Treatment Assignment Model Misspecification based on Simulations with Correct Lag Structure. Two cases are considered where the treatment assignment model is either correctly specified or misspecified. In the latter scenario, the functional form misspecification is considered while the lag structure is correctly modeled. The first three columns show that the bias and root mean squared error (RMSE) for the estimated regression coefficients of the three treatment variables (one for each of the three time periods) from the marginal structural model. The final column presents the bias and RMSE for the estimated mean potential outcome, $\mathbb{E}(Y_i(t_1, t_2, t_3))$, averaged across eight unique treatment sequences. Overall, CBPS1 (thick solid lines; balance conditions) and CBPS2 (thick dashed lines; balance and score conditions alone) outperform the GLM (thin dashed lines) when the model is misspecified. The dotted lines represent the results for the estimates based on the true weights.

the both covariate balancing and score conditions (CBPS2). Finally, the marginal structural model (MSM) weights are constructed from each of the fitted models and then we regress the outcome variable on all three treatment variables using the MSM weights. The resulting regression coefficients are then compared with the numerical estimates of true regression coefficients obtained from a large number of simulations with the true treatment assignment probabilities.

Figure 2 presents the results from the first set of simulations where the misspecification of treatment assignment model is confined to the functional form and the correct lag structure is maintained. The first three columns show that the bias (upper two rows) and root mean squared error or RMSE (bottom two rows) for the estimated regression coefficients of the three treatment variables (one for each of the three time periods, i.e., $\hat{\beta}_1, \hat{\beta}_2$, and $\hat{\beta}_3$, respectively) from the MSM. That is, our MSM is a weighted linear regression where the outcome is regressed on three treatments using the MSM weights. The final column presents the bias and RMSE for the estimated mean potential outcome, $\mathbb{E}(Y_i(t_1, t_2, t_3))$, averaged across eight unique treatment sequences. These estimates are obtained by calculating the weighted average of the outcome using the subset of data for each treatment sequence.

When the treatment assignment model is correctly specified, all methods have small bias (the first row) and small RMSE (third row) for all quantities of interest. However, when the model is misspecified, the bias and RMSE are large and even grow in sample size for GLM (thin dashed lines). In contrast, CBPS1 (thick solid line; covariate balancing conditions alone) and CBPS2 (thick dashed line; covariate balancing and score conditions combined) have much smaller bias and RMSE across parameters. Unlike the GLM, both the bias and RMSE do not grow in sample size, thereby outperforming the standard estimation technique.

In the first row of Figure 3, the misspecified lag structure induces noticeable bias across all methods, with the CBPS methods showing modest gains in bias (first row) and RMSE (third row). When the lag structure is misspecified and the transformed covariates are included (second and fourth rows), the standard GLM estimation leads to large bias and RMSE that increase in the sample size when the model is misspecified. In contrast, the CBPS methods minimize the impact of model misspecification and stays within a reasonable range for bias and RMSE across all quantities of interest.

5 Concluding Remarks

In this paper, we have extended the CBPS methodology of Imai and Ratkovic (2013) to the estimation of inverse probability weights for marginal structural models, which are often used in the

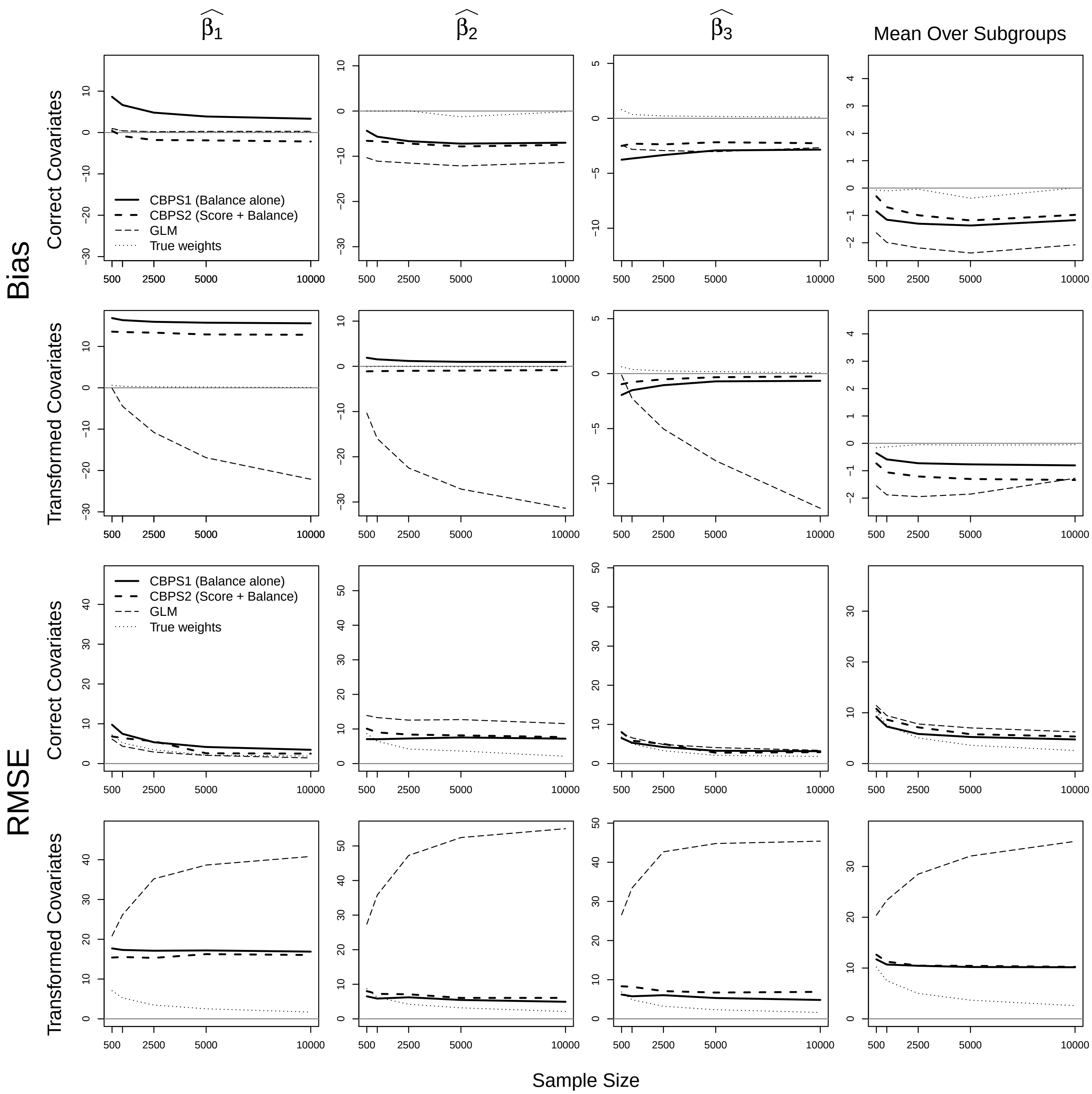


Figure 3: Impact of Treatment Assignment Model Misspecification based on Simulations with Incorrect Lag Structure. Two cases are considered where the treatment assignment model is either correctly specified or misspecified. In the latter scenario, both the functional form and lag structure are misspecified. The first three columns show that the bias and root mean squared error (RMSE) for the estimated regression coefficients of the three treatment variables (one for each of the three time periods) from the marginal structural model. The final column presents the bias and RMSE for the estimated mean potential outcome, $\mathbb{E}(Y_i(t_1, t_2, t_3))$, averaged across eight unique treatment sequences. Overall, CBPS1 (thick solid lines; covariate balancing conditions alone) and CBPS2 (thick dashed lines; covariate balancing and score conditions combined) outperform the GLM (thin dashed lines) when the model is misspecified. The dotted lines represent the results for the estimates based on the true weights.

analysis of longitudinal data. The proposed methodology estimates these weights by optimizing the resulting covariate balance. This is an important advantage because checking covariate balance, after fitting the treatment assignment models, is a difficult task even when the number of time periods is moderate. As a result, detecting model misspecification is much more challenging in longitudinal data settings than simple cross-section data settings. In addition, because the marginal structural model (MSM) weights are constructed by multiplying the inverse of the estimated propensity scores from each time period, MSMs can be highly sensitive to the misspecification of treatment assignment models. In contrast, the CBPS methodology provides a robust estimation method for inverse probability weights by ensuring optimal balance of covariates. Our simulation studies illustrate the effectiveness of the proposed method over the standard maximum likelihood estimation.

One important future research agenda is the question of how to select covariate balancing conditions when there exist many such conditions. In this paper, we have essentially assumed that the number of time periods is relatively small. The standard asymptotic properties of generalized method of moments are applicable so long as we fix the number of time periods and let the sample size tend to infinity. This is the setting we assume in this paper. However, suppose that we fix the sample size and let the number of time periods increase. Under this scenario, the number of possible treatment sequences increases rapidly, implying that the data will become sparse and some treatment sequences have extremely small number of observations. As a result, the number of covariate balancing conditions also increase at an exponential rate while their strength declines. Essentially, this is the problem of many and weak moment conditions. We plan to investigate how the proposed CBPS methodology performs in such a situation and develop effective strategies for addressing this issue.

References

Blackwell, M. (2013). A framework for dynamic causal inference in political science. *American Journal of Political Science* **57**, 2, 504–520.

Box, G. E., Hunter, J. S., and Hunter, W. G. (2005). *Statistics for Experimenters: Design, Innovation, and Discovery.* Wiley-Interscience, New York, 2nd edn.

Cole, S. R. and Hernán, M. A. (2008). Constructing inverse probability weights for marginal structural models. *American Journal of Epidemiology* **168**, 6, 656–664.

Good, I. J. (1958). The interaction algorithm and practical Fourier analysis. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* **20**, 2, 361–372.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* **50**, 4, 1029–1054.

Howe, C. J., Cole, S. R., Chmiel, J. S., and Muñoz, A. (2011). Limitation of inverse probability-of-censoring weights in estimating survival in the presence of strong selection bias. *American Journal of Epidemiology* **173**, 5, 569–577.

Imai, K. and Ratkovic, M. (2013). Covariate balancing propensity score. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* Forthcoming. preprint available at `http://imai.princeton.edu/research/CBPS.html`.

Kang, J. D. and Schafer, J. L. (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data (with discussions). *Statistical Science* **22**, 4, 523–539.

Lefebvre, G., Delaney, J. A. C., and Platt, R. W. (2008). Impact of mis-specification of the treatment model on estimates from a marginal structural model. *Statistics in Medicine* **27**, 18, 3629–3642.

Mortimer, K. M., Neugebauer, R., van der Laan, M., and Tager, I. B. (2005). An application of model-fitting procedures for marginal structural models. *American Journal of Epidemiology* **162**, 4, 382–388.

Neyman, J. (1923). On the application of probability theory to agricultural experiments: Essay on principles, section 9. (translated in 1990). *Statistical Science* **5**, 465–480.

Ratkovic, M., Imai, K., and Fong, C. (2012). Cbps: R package for covariate balancing propensity score. available at the Comprehensive R Archive Network (CRAN). `http://CRAN.R-project.org/package=CBPS`.

Robins, J. (1999). *Statistical Models in Epidemiology, the Environment and Clinical Trials (eds. M. E. Halloran and D. A. Berry)*, chap. Marginal Structural Models Versus Structural Nested Models as Tools for Causal Inference, 95–134. Springer, New York.

Robins, J. M. (1986). A new approach to causal inference in mortality studies with sustained exposure periods: Application to control of the healthy worker survivor effect. *Mathematical Modeling* **7**, 1393–1512.

Robins, J. M., Hernán, M. A., and Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology* **11**, 5, 550–560.

Rubin, D. B. (1973). Matching to remove bias in observational studies. *Biometrics* **29**, 159–183.