Estimating dynamic causal effects with unobserved confounders: a latent class version of the inverse probability weighted estimator

Bartolucci, Francesco and Grilli, Leonardo and Pieroni, Luca

University of Perugia

8 October 2012

Estimating dynamic causal effects with unobserved confounders: a latent class version of the inverse probability weighted estimator

Francesco Bartolucci*, Leonardo Grilli† and Luca Pieroni‡

August 7, 2012

Abstract

We consider estimation of the causal effect of a sequential binary treatment (typically corresponding to a policy or a subsidy in the economic context) on a final outcome, when the treatment assignment at a given occasion depends on the sequence of previous assignments as well as on time-varying confounders. In this case, a popular modeling strategy is represented by Marginal Structural Models; within this approach, the causal effect of the treatment is estimated by the Inverse Probability Weighting (IPW) estimator, which is consistent provided that all the confounders are observed (sequential ignorability). To alleviate this serious limitation, we propose a new estimator, called Latent Class Inverse Probability Weighting (LC-IPW), which is based on two steps: first, a finite mixture model is fitted in order to compute latent-class-specific weights; then, these weights are used to fit the Marginal Structural Model of interest. A simulation study shows that the LC-IPW estimator outperforms the IPW estimator for all the considered configurations, even in cases of no unobserved confounding. The proposed approach is applied to the estimation of the causal effect of wage subsidies on employment, using a dataset of Finnish firms observed for eight years. The LC-IPW estimate confirms the existence of a positive effect, but its magnitude is nearly halved with respect to the IPW estimate, pointing out the substantial role of unobserved confounding in this setting.

Keywords: Causal inference, Longitudinal design, Mixture model, Potential outcomes, Sequential treatment.

JEL: C32, C33, C54, H25.

*Department of Economics, Finance, and Statistics, University of Perugia (IT), bart@stat.unipg.it
†Department of Statistics, University of Florence (IT), grilli@ds.unifi.it
‡Department of Economics, Finance, and Statistics, University of Perugia (IT), lpieroni@unipg.it

1 Introduction

In many fields, including Economics, longitudinal studies are often designed to assess the causal effect of a dynamic treatment on an outcome measured at the end of the period. In the economic context, this treatment corresponds to a policy or, as in the application motivating this paper, to a subsidy for firms satisfying a certain requirement. We consider the common case of a binary treatment assigned repeatedly over time, for example an incentive to firms. The analysis is complicated by the fact that, usually, the treatment assignment at a given occasion (time point or interval) depends on the sequence of previous assignments, as well as on time-varying confounders, namely covariates varying in time and affecting both the treatment assignment and the outcome. In such cases, conditioning on the observed sequence of time-varying confounders yields biased results (Robins et al., 2000). A popular solution is represented by Marginal Structural Models (MSMs); see Robins (1999), Robins et al. (2000), Gill and Robins (2001), and Cole and Hernan (2008).

A MSM is a regression model for the relationship between the outcome and the sequence of treatment assignments: the time-varying confounders are not included in the model, but their distorting effect is neutralized by weighting each observation with the inverse of the probability of the sequence of observed treatments. Such a probability is the product of the occasion-specific probabilities of being assigned to treatment conditional on the history, namely previous treatment assignments and previous values of the confounders. The resulting *Inverse Probability-of-Treatment Weighted* or simply *Inverse Probability Weighted* (IPW) estimator of the causal effect is consistent under quite general assumptions. Note that the method involves the specification of two models: an *outcome model* and a *treatment model* for estimating the weights. Both models are subject to misspecification, with different consequences on the results (Lefebvre et al., 2008; Lefebvre and Gustafson, 2010). The IPW method can be implemented using standard software for regression or specific routines, such as the `R` package `ipw` (van der Wal and Geskus, 2011).

The issue of casual effects of dynamic treatments has been initially considered in Epidemiology, where the main methods are MSM coupled with IPW (Robins et al., 2000) and sequential propensity score adjustment (Achy-Brou et al., 2010). Recently, Lechner

(2009) and Lechner and Miquel (2010) considered the estimation of dynamic causal effects in Economics. Their approach focuses on non-parametric identification and estimation of dynamic causal effects defined either on the whole population or on specific subsets, such as the treated subjects.

The IPW estimator for a MSM yields valid results provided that all the confounders are observed (*sequential ignorability* or *dynamic conditional independence*). To alleviate this serious limitation, we propose an extension of the IPW estimator to account for unobserved pre-treatment confounders. Specifically, we assume that the unobserved confounders are summarized by a discrete latent variable, thus we estimate the probabilities of treatment using a finite mixture (or latent class) model. The new estimator, called Latent Class IPW (LC-IPW), is based on two steps. First, we fit a finite mixture model to treatment indicators and covariates in order to classify the subjects into a small number of groups that we name *latent classes*; subjects in the same latent class are assumed to have the same behavior in terms of the effect of the unobserved confounders. Second, for each subject the weight is computed according to the assigned latent class, and the IPW estimator is applied. Standard errors for the resulting estimates, and corresponding confidence intervals, are obtained by non-parametric bootstrap. We implemented the proposed estimator by combining existing software for latent class modeling and inverse probability weighting.

To assess the properties of the proposed LC-IPW estimator, we rely on a simulation study with a continuous outcome, a sequential binary treatment and some observed and unobserved covariates (potential confounders). The LC-IPW estimator outperforms both the IPW and the OLS regression estimators for all the considered combinations of sample size, number of occasions, distribution of the unobserved confounder and direction of confounding. It is worth to note that LC-IPW may be more accurate than IPW even when there is no unobserved confounding, which is consistent with results on over-adjustment in inverse probability weighting (Rotnitzky et al., 2010).

The proposed approach is illustrated by estimating the effect of firms' subsidies on employment using a dataset of Finnish firms covering the period 1995-2002. This is an instance of sequential binary treatment with time-varying covariates. In this application, the standard IPW method yields a positive estimate of the causal effect of subsidies. The proposed LC-IPW method confirms the existence of a positive effect, but its magnitude is nearly halved, pointing out the substantial role of unobserved confounding in this setting.

The article is organized as follows. In Section 2, we review the standard IPW method, whereas in Section 3 we introduce the LC-IPW method to account for unobserved confounding. In Section 4, we show the results of a simulation study of the performance of the LC-IPW estimator compared with conventional estimators. In Section 5 we illustrate the new method through an application about subsidies on employment. Section 6 offers some concluding remarks.

2 Standard inverse probability weighting

Consider a random sample of n units observed at T consecutive occasions (time points or intervals). In general, we denote each subject by i, $i = 1, \dots, n$, and each occasion by t, $t = 1, \dots, T$, although, to simplify the notation, we will usually omit the subject index i. Moreover, we adopt the following notation: Y is the outcome of interest (measured after the last occasion), S_t is the binary indicator of treatment at occasion t, $\boldsymbol{V}$ is a column vector of pre-treatment covariates (measured before the first occasion), and $\boldsymbol{X}_t$ is a column vector of time-varying covariates (possibly including the outcome variable measured during the observation period). We use the subscript $1:t$ to denote a column vector obtained by stacking vectors of variables measured from occasion 1 until occasion t, namely $\boldsymbol{S}_{1:t} = (S_1, \dots, S_t)'$ and $\boldsymbol{X}_{1:t} = (\boldsymbol{X}'_1, \dots, \boldsymbol{X}'_t)'$. Lowercase letters denote realizations of these variables.

The covariates $\boldsymbol{V}$ and $\boldsymbol{X}_t$ are confounders, namely they simultaneously affect both S_t and Y. Following the potential outcome approach, the outcome Y has a potential version for each treatment sequence, denoted with $Y^{(\boldsymbol{s}_{1:T})}$. The vector $\boldsymbol{Y}^{(all)}$ contains all these potential outcomes (2^T in case of a sequential binary treatment with T occasions). For any subject, the observed outcome is the potential outcome corresponding to the observed treatment sequence. This basic assumption is known as the *observation rule* (Lechner and Miquel, 2010) or the *consistency assumption* (Cole and Hernan, 2008).

The *history* of a variable is the set of variables determined before it, thus potentially affecting it. We solve the simultaneity issue by postulating, as in Lechner and Miquel (2010), that S_t is determined before $\boldsymbol{X}_t$. Therefore, the history of S_t includes $\boldsymbol{V}$, $\boldsymbol{X}_{1:t-1}$ and $\boldsymbol{S}_{1:t-1}$, whereas the history of $\boldsymbol{X}_t$ includes $\boldsymbol{V}$, $\boldsymbol{X}_{1:t-1}$ and $\boldsymbol{S}_{1:t}$ (for $t = 1$ both $\boldsymbol{X}_{1:t-1}$ and $\boldsymbol{S}_{1:t-1}$ vanish). At the last occasion $\boldsymbol{X}_t$ is not a confounder since it does not affect any treatment indicator, thus $\boldsymbol{X}_T$ can be ignored (note that in some applications the outcome Y is a just one of the variables included in $\boldsymbol{X}_T$).

Figure 1 shows the causal Directed Acyclic Graph (DAG) for two occasions ($T = 2$), where the ordering of the variables from left to right reflects the time ordering (the graph does not include $\boldsymbol{X}_t$ at $t = 2$ because it is not a confounder).

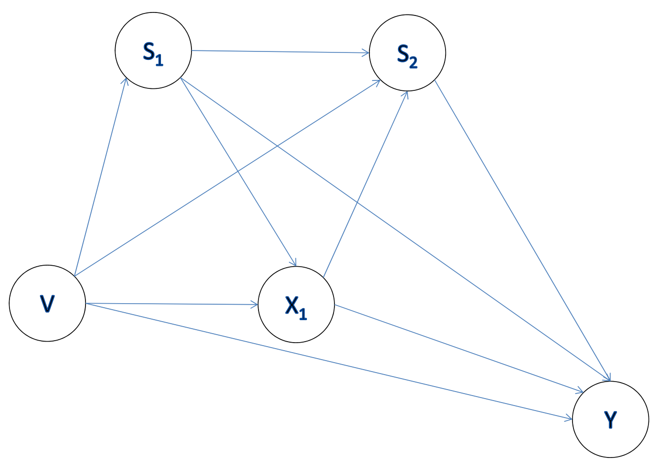

Figure 1: *Causal DAG for two occasions ($T = 2$) without unobserved confounders.*

The DAG in Figure 1 makes it clear the double nature of $\boldsymbol{X}_1$: it is a post-treatment variable relative to S_1 (as it is determined after S_1), whereas it is a confounder relative to S_2 (as it affects both S_2 and Y). This explains the difficulties in getting unbiased estimates of treatment effects: in fact, one should condition on $\boldsymbol{X}_1$ because it is a confounder; at the same time, one should not condition on $\boldsymbol{X}_1$ because it is a post-treatment variable (Rubin, 2005). The IPW estimation method of Robins et al. (2000) solves this controversy as it adjusts for time-varying confounders without conditioning on them.

In the spirit of Robins et al. (2000), we specify a MSM, namely a model for the relationship between the potential outcome $Y^{(\boldsymbol{s}_{1:T})}$ and the treatment indicators $\boldsymbol{s}_{1:T}$. In the continuous case, we may specify a linear model:

$$E(Y^{(\boldsymbol{s}_{1:T})}) = \beta_0 + \boldsymbol{g}(\boldsymbol{s}_{1:T})'\boldsymbol{\beta}_1, \tag{1}$$

where $\boldsymbol{g}(\boldsymbol{s}_{1:T})$ is a function summarizing the treatment sequence, such as the scalar $s_+ = \sum_{t=1}^{T} s_t$ or the vector $(s_+, I(s_T = 1))'$. Under the identification assumptions discussed below, the parameters of interest $\boldsymbol{\beta}_1$ have a causal interpretation, allowing us to make inference on average treatment effects on the whole population (note that, in the context of sequential treatments, the causal effects are sometimes called *dynamic*). For example, the average causal effect of *always treated* versus *never treated* is

$$E(Y^{(1,1,\dots,1)}) - E(Y^{(0,0,\dots,0)}) = [\boldsymbol{g}(1,1,\dots,1) - \boldsymbol{g}(0,0,\dots,0)]'\boldsymbol{\beta}_1, \tag{2}$$

which simplifies to $T\beta_1$ if $\boldsymbol{g}(\boldsymbol{s}_{1:T}) = s_+$.

The first assumption for the identification of causal effects is the *Stable Unit Treatment Value Assumption* (SUTVA) of Rubin (1974), which implies that the potential outcomes of a given subject only depend on the treatment sequence of that subject, thus excluding any interference with other subjects (Achy-Brou et al., 2010).

The second assumption, indifferently called *random assignment* (Achy-Brou et al., 2010), *common support requirement* (Lechner and Miquel, 2010), or *positivity* (Cole and Hernan, 2008), states that the conditional probability of being assigned to treatment is neither zero nor one:

$$0 < Pr(S_t = 1 \mid \boldsymbol{S}_{1:t-1}, \boldsymbol{X}_{1:t-1}, \boldsymbol{V}) < 1, \quad t = 1, \dots, T. \tag{3}$$

In general, a statement conditional on a set of variables is intended to hold for any admissible combination of values of the variables after the conditioning bar.

The third assumption rules out unobserved confounders. This is usually the more difficult assumption to justify in an observational study. In a static setting, namely with a single occasion of administration of the treatment, it states that the treatment indicator is independent of the potential outcomes given the observed covariates. This is called *ignorability* in the literature on causal inference (*selection on observables* in Economics or *exchangeability* in Epidemiology); see Rubin (2005). In a dynamic setting with a sequential treatment, the ignorability assumption is formulated in a recursive way. The *Sequential Ignorability Assumption* (SIA) is written as

$$S_t \perp \boldsymbol{Y}^{(all)} \mid \boldsymbol{S}_{1:t-1}, \boldsymbol{X}_{1:t-1}, \boldsymbol{V}, \quad t = 1, \ldots, T. \tag{4}$$

In Lechner and Miquel (2010) the SIA, supplemented with the common support requirement, is called *Weak Dynamic Conditional Independence Assumption* (whereas the strong version rules out observed confounders). The SIA states that, conditionally on the history up to occasion $t-1$, the treatment assignment at occasion t is independent of the potential outcomes: therefore, for each sub-population defined by the history up to the previous occasion, the treatment is assigned as if it was randomized.

In the following, SUTVA and *random assignment* will be taken as true and we will focus on SIA and its modification.

Under SIA, the causal parameters of a MSM can be consistently estimated by the IPW method: subject i is weighted by the inverse of the probability of its observed treatment sequence, namely $\prod_{t=1}^{T} Pr(S_{it} = s_{it} \mid \boldsymbol{s}_{i,1:t-1}, \boldsymbol{x}_{i,1:t-1}, \boldsymbol{v}_i)$. The occasion-specific probabilities are usually estimated through a pooled logistic regression, namely a standard logistic regression applied to the subject-occasion dataset (any subject contributes with one record for each occasion where it is observed).

The efficiency of the IPW estimator may be unsatisfactory when the weights have a high variability. In particular, subjects with tiny probabilities have large weights which increase the variance. A common method to improve the efficiency is to use *stabilized weights* (Robins et al., 2000):

$$w_i = \frac{\prod_{t=1}^{T} Pr(S_{it} = s_{it} \mid \boldsymbol{s}_{i,1:t-1})}{\prod_{t=1}^{T} Pr(S_{it} = s_{it} \mid \boldsymbol{s}_{i,1:t-1}, \boldsymbol{x}_{i,1:t-1}, \boldsymbol{v}_i)}. \tag{5}$$

The probabilities at the numerator are estimated with the same approach used for the denominator, except that the set of regressors is restricted to the treatment indicators.

3 The proposed approach: latent class inverse probability weighting

We extend the IPW method with the aim of introducing a consistent estimator of causal effects in the presence of a pre-treatment unobserved confounder. To represent such a confounder, we extend the DAG in Figure 1 by adding a further node U pointing at all existing nodes. The extended DAG is shown in Figure 2.

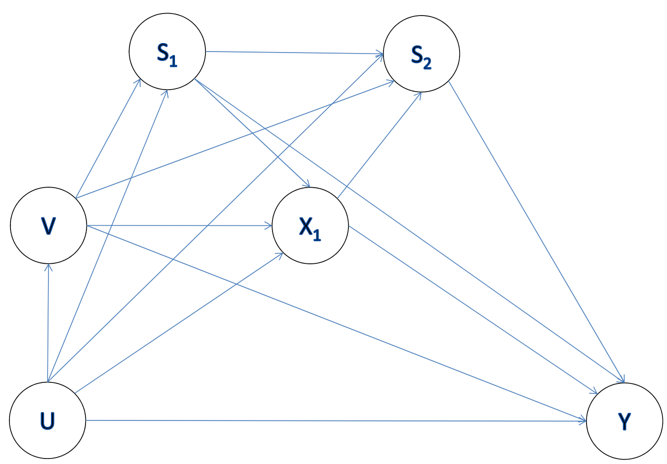

Figure 2: *Causal DAG for two occasions ($T = 2$) with a pre-treatment unobserved confounder U.*

We assume that the confounder is a categorical latent variable U with categories $c = 1, \ldots, k$ corresponding to the different latent classes or mixture components. Therefore, the population is divided into a finite number of latent classes having different parameters for the distribution of the observed variables. It is worth noting that the number of latent classes k and their probabilities $\pi_c = Pr(U = c)$ are parameters to be estimated, thus the approach is flexible enough to satisfactorily approximate also continuous unobserved confounders.

We relax the ignorability assumption (SIA) defined in (4) by requiring that the independence holds within the latent classes induced by the unobserved confounder U. We call it the *Latent Class Sequential Ignorability Assumption* (LC-SIA):

$$S_t \perp \boldsymbol{Y}^{(all)} \mid \boldsymbol{S}_{1:t-1}, \boldsymbol{X}_{1:t-1}, \boldsymbol{V}, U, \quad t = 1, \ldots, T. \tag{6}$$

Clearly, LC-SIA is weaker than SIA because the independence statement is conditional on U, thus in general it does not hold marginally (namely, if U is ignored). Therefore, under LC-SIA the standard IPW estimator may be biased, but it is possible to devise a suitable modification to correct it. A route, noted by Robins et al. (2000) but never implemented, consists in computing the weights using probabilities conditioned on U:

$$Pr(S_{it} = s_{it} \mid \boldsymbol{s}_{i,1:t-1}, \boldsymbol{x}_{i,1:t-1}, \boldsymbol{v}_i, U_i = c_i). \tag{7}$$

Since the latent class c_i of subject i is unknown, it has to predicted on the basis of the available data. For this end, we fit an auxiliary latent class model for the treatment indicators and the observed covariates. Therefore, we propose a two-step estimation procedure:

1. fit an auxiliary latent class model to assign subjects to latent classes;

2. fit a MSM, such as the one in equation (1), using weights computed with the latent-class-specific probabilities (7).

3.1 First step: auxiliary finite mixture model

In order to assign subjects to latent classes, we fit a finite mixture model for the treatment indicators and the observed covariates. For this end, the joint distribution of the observed variables is written as follows:

$$f(\boldsymbol{v}, \boldsymbol{x}_{1:T}, \boldsymbol{s}_{1:T}) = \sum_{c=1}^{k} f(\boldsymbol{v}, \boldsymbol{x}_{1:T}, \boldsymbol{s}_{1:T} \mid c)\pi_c. \tag{8}$$

Conditionally on the latent class, the joint distribution of the observed variables is recursively factorized as follows:

$$f(\boldsymbol{v}, \boldsymbol{x}_{1:T}, \boldsymbol{s}_{1:T} \mid c) = f(\boldsymbol{v}|c) \prod_{t=1}^{T} f(s_t \mid \boldsymbol{s}_{1:t-1}, \boldsymbol{x}_{1:t-1}, \boldsymbol{v}, c) f(\boldsymbol{x}_t \mid \boldsymbol{s}_{1:t}, \boldsymbol{x}_{1:t-1}, \boldsymbol{v}, c). \tag{9}$$

For every probability or density function in (9), we have to choose a suitable model. For the component $f(s_t \mid \boldsymbol{s}_{1:t-1}, \boldsymbol{x}_{1:t-1}, \boldsymbol{v}, c)$ we use logistic regression:

$$f(s_t \mid \boldsymbol{s}_{1:t-1}, \boldsymbol{x}_{1:t-1}, \boldsymbol{v}, c) = p_{t|c}^{s_t}(1 - p_{t|c})^{(1-s_t)} \tag{10}$$

with

$$p_{t|c} = Pr(S_t = 1 \mid \boldsymbol{s}_{1:t-1}, \boldsymbol{x}_{1:t-1}, \boldsymbol{v}, c) = \text{expit}(\eta_{t|c}^{(1)} + \boldsymbol{s}'_{1:t-1}\boldsymbol{\eta}_{t|c}^{(S)} + \boldsymbol{x}'_{1:t-1}\boldsymbol{\eta}_{t|c}^{(X)} + \boldsymbol{v}'\boldsymbol{\eta}_{t|c}^{(V)}), \tag{11}$$

where $\text{expit}(x) = \exp(x)/(1 + \exp(x))$. Note there are specific parameters for every combination of occasion t and latent class (or mixture component) c.

The other components of the distribution (9), namely $f(\boldsymbol{v}|c)$ and $f(\boldsymbol{x}_t \mid \boldsymbol{s}_{1:t}, \boldsymbol{x}_{1:t-1}, \boldsymbol{v}, c)$, should be modeled according to the nature of the variables: for example, for continuous variables the simplest choice is a multivariate normal regression model.

Under random sampling, the log-likelihood of the auxiliary latent class model is the sum over subjects of the logarithm of the joint distribution (8):

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log[f(\boldsymbol{v}_i, \boldsymbol{x}_{i,1:T}, \boldsymbol{s}_{i,1:T})], \tag{12}$$

where $\boldsymbol{\theta}$ is the vector of all model parameters. These parameters can be estimated by maximizing $\ell(\boldsymbol{\theta})$ by means of an EM algorithm implemented along the same lines as in . For reasons of computational efficiency, we have written a `MATLAB` code, but the auxiliary model could be fitted by existing software for latent class or mixture models, for example the `R` package `flexmix` (Grün and Leisch, 2008) and by the STATA package `fmm` by Kit Baum[1].

[1]The series of `MATLAB" functions are available from the authors upon request.`

Once the parameters have been estimated, every subject is assigned to a latent class on the basis of the estimated posterior probabilities $\hat{q}_{ic}$, $c = 1, \dots, k$, where

$$\hat{q}_{ic} = \frac{\hat{f}(\boldsymbol{v}_i, \boldsymbol{x}_{i,1:T}, \boldsymbol{s}_{i,1:T} \mid c)\hat{\pi}_c}{\sum_{c=1}^{k} \hat{f}(\boldsymbol{v}_i, \boldsymbol{x}_{i,1:T}, \boldsymbol{s}_{i,1:T} \mid c)\hat{\pi}_c}. \tag{13}$$

Subject i is assigned to the latent class $\hat{c}_i$ with the highest estimated posterior probability.

Even if the aim is to estimate the probabilities of the treatment indicators, the auxiliary latent class model is a model for both the treatment indicators $\boldsymbol{S}_{1:T}$ and the covariates $\boldsymbol{V}$ and $\boldsymbol{X}_{1:T}$. Indeed, modeling the covariates is not necessary, but it yields better predictions of the latent classes, thus increasing the precision of the weighted estimator of the causal effect.

The choice of the number of latent classes k is a critical issue. The most popular method is based on the Bayesian Information Criterion (BIC) of Schwarz (1978); see McLachlan and Peel (2000), Chapter 6, among others. This criterion is based on the minimization of the index

$$\text{BIC}_k = -2\ell(\hat{\boldsymbol{\theta}}_k) + \log(n)\#\text{par}_k,$$

where $\ell(\hat{\boldsymbol{\theta}}_k)$ is the maximum log-likelihood of the model with k components and $\#\text{par}_k$ denotes the number of free parameters. This index accounts for the goodness-of-fit of the model (measured by the log-likelihood) and the model complexity (in terms of number of free parameters). An alternative criterion to select the number of latent classes is the Normalized Entropy Criterion (NEC) of Celeux and Soromenho (1996); see also Biernacki et al. (1999). The NEC explicitly accounts for the quality of the classification, in addition to the goodness-of-fit, since it is based on the minimization of the index

$$\text{NEC}_k = \frac{-\sum_{i=1}^{n}\sum_{c=1}^{k} \hat{q}_{ic}\log(\hat{q}_{ic})}{\ell(\hat{\boldsymbol{\theta}}_k) - \ell(\hat{\boldsymbol{\theta}}_1)},$$

with $\text{NEC}_1 \equiv 1$.

In practice, we recommend to select the number of latent classes by considering both BIC and NEC. In fact, even if BIC should be regarded as the main criterion, we noted that it sometimes leads to a large number of classes, with some classes containing few sample units: this situation is potentially harmful since the parameter estimates for small classes may be unreliable with negative effects on the stability of the weights. Therefore, we suggest to also consider NEC, which may tend to select a smaller number of latent classes. In practice, NEC is used as a counterpart to check whether BIC leads to choose an excessive number of latent classes; when, as in the application illustrated in Section 5, there is a significant difference between the model selected by BIC and that selected by NEC, we prefer the second one. On the contrary, in the simulation study of Section 4 we relied solely on BIC since, as we experimented, the two criteria yielded similar results.

3.2 Second step: weighted estimation of the causal model

The second step of the proposed LC-IPW method entails fitting the MSM (1) with a modified IPW procedure where the weight of each subject is computed conditionally on the assigned latent class. Specifically, the stabilized weights (5) become

$$w_{i,\hat{c}_i} = \frac{\prod_{t=1}^{T} Pr(S_{it} = s_{it} \mid \boldsymbol{s}_{i,1:t-1}, U_i = \hat{c}_i)}{\prod_{t=1}^{T} Pr(S_{it} = s_{it} \mid \boldsymbol{s}_{i,1:t-1}, \boldsymbol{x}_{i,1:t-1}, \boldsymbol{v}_i, U_i = \hat{c}_i)}. \tag{14}$$

The probabilities at the denominator are estimated by using the logistic models in equations (10) and (11) after assigning the latent class and replacing parameters with maximum likelihood estimates. Such estimates are available from the model fitted in the first step. The probabilities at the numerator are computed in a similar way, but without conditioning on the covariates and the latent classes. The standard errors and corresponding confidence intervals can be computed via non-parametric bootstrap (Efron and Tibshirani, 1994): the method consists in resampling with replacement and repeating the whole estimation process, including the estimation of the finite mixture auxiliary model. As will be clear in the application, these confidence interval may be used to test that the treatment has no causal effect on the target variable.

4 Simulation study

In order to study the performance of the proposed LC-IPW estimator and to make comparisons with conventional estimators, we carried out a simulation study with an unobserved pre-treatment covariate U having different roles (confounder or not) and different distributions (discrete or continuous). The number of occasions and the sample size include values comparable to those of the application, about subsidies received by a group of firms, which will be illustrated in Section 5. In the following, we first describe the simulation design and then we discuss the results.

4.1 Design

The simulation design is based on a model for $T = 4$ or $T = 8$ occasions with a continuous outcome Y, a sequential binary treatment S_t, a pre-treatment continuous covariate V, a time-varying continuous covariate X_t, and an unobserved pre-treatment covariate U. The covariates always affect the outcome and later values of the time-varying covariates; however, we introduced parameters regulating the effect of the covariates on the treatment indicators, so that the covariates may be *confounders* (if they also affect the treatment indicators) or not (if they do not affect the treatment indicators, a case sometimes called *pure predictors of outcome*).

The data generating model is as follows:

$$
\begin{aligned}
\text{logit}Pr(S_{it}=1) &= \begin{cases} -1 + u_i\phi_1(4/T) + v_i\phi_2(4/T), & t=1,\\ -1 + u_i\phi_1(4/T) + x_{i,t-1}\phi_2(4/T) - s_{i,t-1}, & t=2,\ldots,T, \end{cases}\\
X_{it} &= \begin{cases} -0.25 + u_i/2 + v_i + s_{it} + \epsilon_{it}, & t=1,\\ -0.25 + u_i/2 + x_{i,t-1} + s_{it} + \epsilon_{it}, & t=2,\ldots,T-1, \end{cases}\\
Y_i &= u_i/2 + x_{i,T-1} + s_{iT} - 0.25 + \epsilon_{iT},
\end{aligned}
$$

where ϵ_{it} are iid $N(0, 0.25)$ and V_i are iid $N(0, 1)$. The level of confounding is regulated by the parameters $\phi_1 = -0.5, 0, 0.5$ (unobserved confounding via U), and $\phi_2 = -0.5, 0, 0.5$ (observed confounding via V and X_t).

For the unobserved pre-treatment covariate U we used several alternative distributions:

(i) LC2: discrete Uniform on $\{-1, 1\}$; (ii) LC3-type1: discrete Uniform on $\{-\sqrt{1.5}, 0, \sqrt{1.5}\}$; ($iii$) LC3-type2: discrete Uniform on $\{-2, 0, 2\}$; (iv) Normal: standard Normal; (v) Uniform: continuous Uniform on the interval $(-\sqrt{3}, \sqrt{3})$. Note that the previous distributions have all mean 0 and variance 1, except LC3-type2 which has variance equal to 8/3.

Regardless of the distribution of U and the values of ϕ_1 and ϕ_2, the MSM for the outcome is

$$E(Y^{(\boldsymbol{s}_{1:T})}) = \beta_0 + s_+\beta_1,$$

where $\beta_1 = 1$ for both $T = 4$ and $T = 8$ (the intercept is $\beta_0 = -1$ for $T = 4$ and $\beta_0 = -2$ for $T = 8$, but it is not considered later on since it is not of direct interest).

Concerning the sample size, we considered $n = 1000$ and $n = 4000$. Overall, for the number of configurations considered for distribution every of U is 36 ($2 \times 2 \times 3 \times 3$ values of n, T, ϕ_1, ϕ_2). For each of these configurations we generated 1000 random samples and fitted the MSM with four estimation methods: (i) OLS (least squares unweighted regression), (ii) IPW (with stabilized weights), (iii) LC-IPW (with stabilized weights and a number of latent classes chosen by BIC and NEC in order to prevent an excessive number of classes), (iv) true weights (similar to LC-IPW except that true weights are used instead of estimated weights).

The results for the last estimator, which is feasible only in a simulation setting, allows us to assess the impact of estimating the weights on the performance of the LC-IPW estimator.

4.2 Results

Tables 1 to 5 summarize the simulation results for the causal parameter $\beta_1 = 1$ using the four estimators under consideration. Each table refers to a distribution of the unobserved potential confounder U and reports the median bias and MAE (Median Absolute Error) for the 36 configurations defined by n, T, ϕ_1, and ϕ_2. The choice of median-based summary measures is suggested by the asymmetry of the sampling distributions that we observed to be strong in certain cases.

In the following, we first comment the results for U having a discrete distribution with two mass points (LC2) and then we discuss the other discrete cases (LC3-type1, LC3-type2) and continuous cases (Normal, Uniform). In order to easily assess the results, for each considered distribution we also computed the number of cases in which the proposed LC-IPW estimator outperforms the IPW estimator, in terms of median bias or MAE, and the average of the difference between the absolute median bias, or MAE, of the first estimator with respect to the second. These results are reported in Table 6.

Table 1: *Simulation results (Median Bias and MAE) for U discrete with two mass points (LC2) with $T = 4, 8$ and $n = 1000, 4000$*

			Median Bias				MAE			
			$T=4$		$T=8$		$T=4$		$T=8$	
ϕ_1	ϕ_2	Method	1000	4000	1000	4000	1000	4000	1000	4000
-0.5	-0.5	OLS	-1.785	-1.786	-2.362	-2.355	1.785	1.786	2.362	2.355
		IPW	-0.591	-0.531	-0.500	-0.372	0.596	0.532	0.544	0.398
		LC-IPW	-0.159	-0.087	-0.199	-0.086	0.233	0.137	0.296	0.167
		true weights	-0.140	-0.062	-0.202	-0.073	0.242	0.142	0.318	0.173
-0.5	0.0	OLS	-0.960	-0.958	-1.043	-1.040	0.960	0.958	1.043	1.040
		IPW	-0.549	-0.540	-0.407	-0.409	0.549	0.540	0.407	0.409
		LC-IPW	-0.016	-0.010	-0.002	0.000	0.045	0.021	0.034	0.019
		true weights	-0.005	0.004	0.000	0.002	0.079	0.043	0.097	0.048
-0.5	0.5	OLS	0.302	0.304	0.932	0.928	0.302	0.304	0.932	0.928
		IPW	-0.525	-0.530	-0.497	-0.505	0.525	0.530	0.497	0.505
		LC-IPW	0.007	-0.001	0.009	0.004	0.056	0.030	0.085	0.046
		true weights	0.017	0.001	0.017	0.000	0.088	0.041	0.121	0.057
0.0	-0.5	OLS	-1.399	-1.399	-2.090	-2.091	1.399	1.399	2.090	2.091
		IPW	-0.051	-0.026	-0.088	-0.054	0.125	0.075	0.187	0.110
		LC-IPW	-0.061	-0.026	-0.099	-0.055	0.124	0.070	0.194	0.108
		true weights	-0.046	-0.022	-0.081	-0.042	0.148	0.082	0.221	0.127
0.0	0.0	OLS	0.004	0.000	0.000	0.003	0.070	0.037	0.090	0.045
		IPW	-0.001	0.001	0.000	0.004	0.045	0.021	0.051	0.027
		LC-IPW	-0.001	0.000	0.000	0.001	0.029	0.013	0.026	0.013
		true weights	0.006	0.000	0.001	0.003	0.070	0.037	0.090	0.046
0.0	0.5	OLS	1.047	1.049	1.525	1.524	1.047	1.049	1.525	1.524
		IPW	0.033	0.020	0.047	0.020	0.097	0.061	0.153	0.074
		LC-IPW	0.036	0.019	0.047	0.017	0.098	0.054	0.138	0.070
		true weights	0.033	0.014	0.044	0.021	0.109	0.065	0.162	0.079
0.5	-0.5	OLS	-0.511	-0.508	-1.418	-1.420	0.511	0.508	1.418	1.420
		IPW	0.442	0.454	0.315	0.332	0.442	0.454	0.315	0.332
		LC-IPW	-0.013	-0.002	-0.048	-0.010	0.071	0.039	0.114	0.069
		true weights	-0.024	-0.003	-0.026	-0.006	0.109	0.059	0.162	0.089
0.5	0.0	OLS	0.964	0.963	1.059	1.047	0.964	0.963	1.059	1.047
		IPW	0.571	0.571	0.508	0.497	0.571	0.571	0.508	0.497
		LC-IPW	0.016	0.012	0.008	0.002	0.053	0.027	0.050	0.024
		true weights	0.013	0.004	0.018	0.005	0.086	0.038	0.101	0.049
0.5	0.5	OLS	1.394	1.394	1.784	1.785	1.394	1.394	1.784	1.785
		IPW	0.637	0.601	0.692	0.632	0.638	0.601	0.694	0.632
		LC-IPW	0.134	0.075	0.110	0.045	0.188	0.114	0.206	0.116
		true weights	0.121	0.053	0.107	0.046	0.188	0.112	0.228	0.130

First of all consider the situation of no confounding at all ($\phi_1 = \phi_2 = 0$), a situation in which all the estimators are expected to be unbiased. Under the LC2 model for the distribution of U, in fact, the median bias is always very close to 0. Moreover, as may be

Table 2: *Simulation results (Median Bias and MAE) for U discrete with three mass points (LC3-type1) with $T = 4, 8$ and $n = 1000, 4000$*

			Median Bias				MAE			
			$T=4$		$T=8$		$T=4$		$T=8$	
ϕ_1	ϕ_2	Method	1000	4000	1000	4000	1000	4000	1000	4000
-0.5	-0.5	OLS	-1.775	-1.778	-2.349	-2.346	1.775	1.778	2.349	2.346
		IPW	-0.569	-0.524	-0.490	-0.399	0.576	0.524	0.526	0.418
		LC-IPW	-0.216	-0.138	-0.208	-0.128	0.271	0.174	0.349	0.209
		true weights	-0.156	-0.088	-0.223	-0.114	0.252	0.156	0.351	0.214
-0.5	0.0	OLS	-0.965	-0.959	-1.042	-1.043	0.965	0.959	1.042	1.043
		IPW	-0.540	-0.535	-0.407	-0.406	0.540	0.535	0.407	0.406
		LC-IPW	-0.056	-0.005	-0.002	0.005	0.064	0.024	0.046	0.023
		true weights	-0.013	-0.002	-0.007	-0.002	0.087	0.043	0.104	0.049
-0.5	0.5	OLS	0.302	0.305	0.934	0.926	0.302	0.305	0.934	0.926
		IPW	-0.543	-0.546	-0.506	-0.516	0.543	0.546	0.506	0.516
		LC-IPW	-0.052	0.020	0.022	0.015	0.069	0.037	0.088	0.046
		true weights	0.010	0.003	0.010	0.002	0.084	0.048	0.115	0.064
0.0	-0.5	OLS	-1.400	-1.397	-2.086	-2.087	1.400	1.397	2.086	2.087
		IPW	-0.062	-0.030	-0.132	-0.051	0.134	0.067	0.221	0.120
		LC-IPW	-0.026	-0.036	-0.142	-0.051	0.123	0.070	0.229	0.119
		true weights	-0.058	-0.029	-0.118	-0.057	0.157	0.080	0.250	0.124
0.0	0.0	OLS	0.009	-0.001	-0.007	-0.001	0.070	0.035	0.090	0.042
		IPW	0.005	0.000	0.003	-0.001	0.043	0.021	0.055	0.026
		LC-IPW	0.033	0.001	0.001	0.000	0.041	0.016	0.032	0.014
		true weights	0.009	-0.001	-0.008	0.000	0.071	0.034	0.091	0.042
0.0	0.5	OLS	1.043	1.048	1.522	1.518	1.043	1.048	1.522	1.518
		IPW	0.029	0.013	0.054	0.019	0.104	0.059	0.142	0.084
		LC-IPW	0.066	0.018	0.068	0.018	0.108	0.054	0.141	0.076
		true weights	0.036	0.017	0.061	0.017	0.112	0.067	0.152	0.093
0.5	-0.5	OLS	-0.512	-0.512	-1.421	-1.421	0.512	0.512	1.421	1.421
		IPW	0.462	0.466	0.325	0.326	0.462	0.466	0.325	0.326
		LC-IPW	0.090	-0.023	-0.057	-0.033	0.099	0.047	0.132	0.075
		true weights	-0.023	-0.012	-0.039	-0.023	0.112	0.059	0.166	0.085
0.5	0.0	OLS	0.953	0.958	1.048	1.046	0.953	0.958	1.048	1.046
		IPW	0.564	0.562	0.497	0.493	0.564	0.562	0.497	0.493
		LC-IPW	0.137	0.003	0.001	-0.003	0.137	0.029	0.054	0.024
		true weights	0.010	0.002	0.008	0.004	0.085	0.045	0.099	0.049
0.5	0.5	OLS	1.387	1.387	1.776	1.775	1.387	1.387	1.776	1.775
		IPW	0.611	0.588	0.675	0.627	0.613	0.588	0.678	0.628
		LC-IPW	0.233	0.096	0.138	0.067	0.258	0.127	0.237	0.135
		true weights	0.119	0.067	0.131	0.060	0.195	0.118	0.250	0.135

expected, the MAE decreases with n, while it increases with the number of occasions T, with the notable exception of LC-IPW, whose MAE seems to be less sensitive to T. It also worth noting that in this case the MAE of the LC-IPW estimator tends to be smaller

Table 3: *Simulation results (Median Bias and MAE) for U discrete with three mass points (LC3-type2) with $T = 4, 8$ and $n = 1000, 4000$*

			Median Bias				MAE			
			$T = 4$		$T = 8$		$T = 4$		$T = 8$	
ϕ_1	ϕ_2	Method	1000	4000	1000	4000	1000	4000	1000	4000
-0.5	-0.5	OLS	-2.662	-2.664	-3.570	-3.569	2.662	2.664	3.570	3.569
		IPW	-1.207	-1.103	-1.399	-1.048	1.210	1.106	1.442	1.113
		LC-IPW	-0.692	-0.453	-1.025	-0.630	0.749	0.510	1.123	0.773
		true weights	-0.575	-0.399	-1.010	-0.637	0.664	0.478	1.152	0.787
-0.5	0.0	OLS	-2.083	-2.077	-2.468	-2.465	2.083	2.077	2.468	2.465
		IPW	-1.071	-1.060	-0.834	-0.825	1.071	1.060	0.834	0.825
		LC-IPW	-0.096	-0.053	-0.021	-0.007	0.160	0.083	0.123	0.056
		true weights	-0.063	-0.030	-0.017	-0.012	0.198	0.102	0.201	0.098
-0.5	0.5	OLS	-0.088	-0.083	1.761	1.746	0.101	0.083	1.761	1.746
		IPW	-1.097	-1.094	-1.093	-1.112	1.097	1.094	1.093	1.112
		LC-IPW	-0.009	-0.009	0.052	0.020	0.091	0.048	0.191	0.104
		true weights	0.006	0.003	0.039	0.022	0.126	0.074	0.222	0.123
0.0	-0.5	OLS	-2.336	-2.338	-3.647	-3.651	2.336	2.338	3.647	3.651
		IPW	-0.152	-0.074	-0.384	-0.220	0.248	0.147	0.597	0.370
		LC-IPW	-0.140	-0.083	-0.397	-0.229	0.235	0.143	0.602	0.375
		true weights	-0.140	-0.074	-0.394	-0.226	0.268	0.166	0.594	0.380
0.0	0.0	OLS	0.017	-0.002	-0.014	-0.002	0.101	0.050	0.136	0.066
		IPW	0.006	-0.001	0.004	0.000	0.062	0.030	0.081	0.038
		LC-IPW	-0.003	0.000	0.000	-0.001	0.029	0.014	0.034	0.014
		true weights	0.018	-0.001	-0.018	-0.001	0.103	0.050	0.137	0.066
0.0	0.5	OLS	1.759	1.762	2.758	2.754	1.759	1.762	2.758	2.754
		IPW	0.102	0.049	0.214	0.142	0.205	0.121	0.413	0.258
		LC-IPW	0.117	0.051	0.245	0.142	0.197	0.109	0.402	0.252
		true weights	0.112	0.051	0.218	0.132	0.218	0.130	0.428	0.255
0.5	-0.5	OLS	-0.194	-0.191	-2.689	-2.698	0.196	0.191	2.689	2.698
		IPW	1.039	1.040	0.648	0.683	1.039	1.040	0.648	0.683
		LC-IPW	0.004	0.006	-0.129	-0.061	0.108	0.060	0.288	0.163
		true weights	-0.009	-0.007	-0.106	-0.056	0.155	0.085	0.314	0.186
0.5	0.0	OLS	2.070	2.077	2.470	2.467	2.070	2.077	2.470	2.467
		IPW	1.307	1.294	1.187	1.182	1.307	1.294	1.187	1.182
		LC-IPW	0.094	0.052	0.015	0.008	0.155	0.089	0.126	0.061
		true weights	0.058	0.020	0.030	0.007	0.185	0.103	0.202	0.098
0.5	0.5	OLS	2.143	2.144	2.797	2.799	2.143	2.144	2.797	2.799
		IPW	1.337	1.295	1.738	1.590	1.338	1.295	1.742	1.592
		LC-IPW	0.543	0.348	0.744	0.442	0.593	0.405	0.891	0.579
		true weights	0.487	0.298	0.727	0.458	0.561	0.357	0.871	0.567

than that of the IPW estimator and of the estimator based on the true weights. Even if less frequently, this phenomenon is also observed about the median bias.

We then consider the situations of observed confounding only ($\phi_1 = 0, \phi_2 \neq 0$), again

Table 4: *Simulation results (Median Bias and MAE) for U continuous (Normal) with $T = 4, 8$ and $n = 1000, 4000$*

			Median Bias				MAE			
			$T=4$		$T=8$		$T=4$		$T=8$	
ϕ_1	ϕ_2	Method	1000	4000	1000	4000	1000	4000	1000	4000
-0.5	-0.5	OLS	-1.753	-1.753	-2.325	-2.323	1.753	1.753	2.325	2.323
		IPW	-0.524	-0.472	-0.478	-0.375	0.529	0.474	0.515	0.394
		LC-IPW	-0.322	-0.231	-0.369	-0.215	0.350	0.247	0.413	0.272
		true weights	-0.197	-0.115	-0.315	-0.171	0.276	0.180	0.402	0.247
-0.5	0.0	OLS	-0.945	-0.943	-1.040	-1.039	0.945	0.943	1.040	1.039
		IPW	-0.518	-0.510	-0.402	-0.399	0.518	0.510	0.402	0.399
		LC-IPW	-0.094	-0.078	0.001	-0.010	0.102	0.080	0.073	0.034
		true weights	-0.013	-0.005	-0.003	-0.005	0.096	0.052	0.109	0.056
-0.5	0.5	OLS	0.303	0.310	0.927	0.920	0.303	0.310	0.927	0.920
		IPW	-0.588	-0.595	-0.518	-0.542	0.588	0.595	0.518	0.542
		LC-IPW	-0.043	-0.040	0.075	0.027	0.073	0.045	0.133	0.059
		true weights	0.001	0.004	0.027	0.004	0.087	0.046	0.133	0.065
0.0	-0.5	OLS	-1.389	-1.386	-2.066	-2.062	1.389	1.386	2.066	2.062
		IPW	-0.063	-0.032	-0.152	-0.086	0.137	0.075	0.219	0.143
		LC-IPW	-0.041	-0.023	-0.129	-0.086	0.126	0.075	0.215	0.144
		true weights	-0.068	-0.031	-0.142	-0.078	0.154	0.087	0.247	0.153
0.0	0.0	OLS	0.009	0.002	-0.002	0.003	0.071	0.036	0.085	0.043
		IPW	0.000	0.002	0.002	0.001	0.043	0.021	0.051	0.026
		LC-IPW	0.052	0.023	0.082	0.038	0.058	0.027	0.083	0.038
		true weights	0.009	0.001	-0.003	0.002	0.071	0.036	0.085	0.043
0.0	0.5	OLS	1.037	1.040	1.498	1.502	1.037	1.040	1.498	1.502
		IPW	0.047	0.018	0.071	0.045	0.108	0.065	0.180	0.103
		LC-IPW	0.110	0.047	0.186	0.095	0.141	0.073	0.225	0.124
		true weights	0.041	0.015	0.064	0.043	0.112	0.071	0.186	0.111
0.5	-0.5	OLS	-0.516	-0.510	-1.407	-1.409	0.516	0.510	1.407	1.409
		IPW	0.498	0.505	0.308	0.334	0.498	0.505	0.308	0.334
		LC-IPW	0.155	0.090	0.036	0.018	0.157	0.091	0.128	0.075
		true weights	-0.017	0.000	-0.046	-0.019	0.109	0.058	0.173	0.094
0.5	0.0	OLS	0.951	0.948	1.044	1.038	0.951	0.948	1.044	1.038
		IPW	0.540	0.538	0.489	0.483	0.540	0.538	0.489	0.483
		LC-IPW	0.189	0.127	0.168	0.077	0.190	0.127	0.169	0.078
		true weights	0.029	0.013	0.029	0.003	0.104	0.049	0.114	0.058
0.5	0.5	OLS	1.368	1.369	1.753	1.752	1.368	1.369	1.753	1.752
		IPW	0.563	0.531	0.618	0.578	0.563	0.532	0.625	0.579
		LC-IPW	0.307	0.201	0.334	0.194	0.315	0.208	0.365	0.228
		true weights	0.164	0.078	0.222	0.121	0.221	0.132	0.298	0.190

under the LC2 model. Obviously, the OLS estimator is biased (even strongly in certain cases), whereas all the weighted estimators have a reduced median bias that, in absolute value, decreases with n and increases with T. A similar behavior is observed for the

Table 5: *Simulation results (Median Bias and MAE) for U continuous (Uniform) with $T = 4, 8$ and $n = 1000, 4000$*

			Median Bias				MAE			
			$T = 4$		$T = 8$		$T = 4$		$T = 8$	
ϕ_1	ϕ_2	Method	1000	4000	1000	4000	1000	4000	1000	4000
-0.5	-0.5	OLS	-1.768	-1.772	-2.340	-2.338	1.768	1.772	2.340	2.338
		IPW	-0.548	-0.511	-0.509	-0.367	0.554	0.513	0.543	0.394
		LC-IPW	-0.222	-0.146	-0.275	-0.127	0.271	0.177	0.359	0.220
		true weights	-0.152	-0.094	-0.273	-0.116	0.245	0.166	0.365	0.233
-0.5	0.0	OLS	-0.950	-0.954	-1.040	-1.043	0.950	0.954	1.040	1.043
		IPW	-0.529	-0.530	-0.406	-0.405	0.529	0.530	0.406	0.405
		LC-IPW	-0.057	-0.016	0.026	0.019	0.070	0.028	0.054	0.029
		true weights	-0.011	-0.001	-0.003	-0.003	0.086	0.047	0.102	0.052
-0.5	0.5	OLS	0.306	0.306	0.924	0.926	0.306	0.306	0.924	0.926
		IPW	-0.551	-0.554	-0.509	-0.520	0.551	0.554	0.509	0.520
		LC-IPW	-0.051	0.004	0.052	0.036	0.079	0.034	0.107	0.055
		true weights	0.015	0.003	0.014	0.009	0.090	0.047	0.131	0.061
0.0	-0.5	OLS	-1.392	-1.396	-2.089	-2.083	1.392	1.396	2.089	2.083
		IPW	-0.063	-0.028	-0.135	-0.063	0.126	0.070	0.229	0.132
		LC-IPW	-0.022	-0.018	-0.092	-0.055	0.121	0.066	0.210	0.123
		true weights	-0.059	-0.025	-0.123	-0.064	0.144	0.079	0.243	0.142
0.0	0.0	OLS	0.001	0.003	0.009	0.002	0.066	0.034	0.090	0.044
		IPW	0.004	0.000	0.007	0.002	0.044	0.022	0.054	0.026
		LC-IPW	0.047	0.015	0.076	0.025	0.050	0.020	0.076	0.026
		true weights	0.001	0.004	0.008	0.001	0.066	0.034	0.089	0.044
0.0	0.5	OLS	1.045	1.046	1.512	1.513	1.045	1.046	1.512	1.513
		IPW	0.048	0.019	0.052	0.026	0.102	0.061	0.154	0.084
		LC-IPW	0.090	0.038	0.147	0.055	0.119	0.062	0.189	0.094
		true weights	0.048	0.020	0.057	0.024	0.111	0.065	0.170	0.088
0.5	-0.5	OLS	-0.506	-0.510	-1.410	-1.418	0.506	0.510	1.410	1.418
		IPW	0.469	0.473	0.322	0.330	0.469	0.473	0.322	0.330
		LC-IPW	0.125	0.018	0.031	-0.008	0.128	0.040	0.119	0.072
		true weights	-0.015	-0.006	-0.029	-0.020	0.103	0.053	0.161	0.090
0.5	0.0	OLS	0.957	0.954	1.043	1.039	0.957	0.954	1.043	1.039
		IPW	0.561	0.554	0.493	0.490	0.561	0.554	0.493	0.490
		LC-IPW	0.159	0.045	0.119	0.032	0.159	0.047	0.120	0.036
		true weights	0.015	-0.001	0.008	-0.003	0.097	0.044	0.102	0.050
0.5	0.5	OLS	1.381	1.382	1.770	1.768	1.381	1.382	1.770	1.768
		IPW	0.591	0.573	0.629	0.609	0.593	0.574	0.635	0.610
		LC-IPW	0.240	0.143	0.239	0.114	0.265	0.162	0.301	0.162
		true weights	0.113	0.078	0.119	0.078	0.198	0.128	0.253	0.157

MAE. It is again worth observing that, even if the latent class adjustment is not needed, the LC-IPW estimator is, in most cases, more efficient than both the standard weighted estimator and the infeasible estimator with true weights.

Table 6: *Summary of the comparison between the LC-IPW estimator and IPW estimator in terms of median bias and MAE; for every distribution the table shows the number of times the LC-IPW outperforms the IPW estimator in terms of median bias and MAE and the corresponding average differences (in absolute value for the median bias)*

distribution	Better LC-IPW		Average difference	
	median bias	MAE	median bias	MAE
LC2	31/36	34/36	-0.3094	-0.2816
LC3-type1	29/36	33/36	-0.2881	-0.2651
LC3-type2	29/36	34/36	-0.6030	-0.5565
Normal	28/36	26/36	-0.2253	-0.2169
Uniform	28/36	29/36	-0.2638	-0.2489

Finally, in the presence of unobserved confounding ($\phi_1 \neq 0$), both the OLS and the IPW estimators tend to be severely biased, whereas the LC-IPW estimator usually has a small bias (always lower than the bias of the competing estimators); even when LC-IPW has not a negligible bias (for instance when $\phi_1 = \phi_2 = -0.5$ and $\phi_1 = \phi_2 = 0.5$), this bias decreases with the sample size n (while the bias of IPW is stable). Therefore, for some model specifications the good asymptotic properties of LC-IPW may show up only in very large samples. In any case, LC-IPW represents a striking improvement over IPW in terms of bias and accuracy for any configuration and sample size. This conclusion is confirmed by the result in Table 6 which shows that, under the distribution LC2 for U, the proposed estimator outperforms the IPW estimator in 31 cases (in terms of median bias) and in 34 cases (in terms of MAE) out of 36.

The previous findings are confirmed for alternative distributions of the unobserved covariate U, both discrete (Tables 2 and 3) and continuous (Tables 4 and 5). In particular, also with the help of the summary statistics in Table 6, we conclude that under the LC3 distribution the results are very similar than under the LC2 distribution. This is rather plausible since under both models the latent variable U has, by assumption, the same variance. On the other hand, the advantage of the LC-IPW estimator over the IPW estimator considerably increases under the LC3-type2 distribution, since in this case the distribution of U has a higher variance, and then the effect of the unobserved confounding is more significant. Regarding the continuous distributions, we observe that the proposed estimator has a reduced advantage over the IPW estimator with respect to the LC2 and

LC3 cases, in which the distribution of U has the same variance. However, it is worth noting that, even in these cases, the proposed method allows us to strongly reduce the bias in estimating the causal effect due to the unobserved confounding. For these continuous cases, a detailed analysis of the simulation results (not reported here) reveals that the required number of latent classes is limited, not larger than six. Specifically, for any configuration of n, T, ϕ_1 and ϕ_2, the number of latent classes k is selected by jointly use BIC and NEC, separately for each of the 1000 samples, thus generating a Monte Carlo distribution of k. In our simulations for U Normal and U Uniform, the modal value of k ranges from 2 to 5; higher values of k are selected for U Normal (as compared to U Uniform) and for larger sample sizes.

The efficiency of the LC-IPW estimator in the case of no unobserved confounding may appear as surprising, but it is in line with the simulation results of Lefebvre et al. (2008), who compared the performance of the IPW estimator under different specifications of the model for the treatment indicators. Their baseline specification includes only confounders, namely predictors of both treatment and outcome. The key result is that, in terms of mean squared error and then efficiency, adding *pure predictors of treatment* is deleterious, whereas adding *pure predictors of outcome* is beneficial. To see the connection with our simulations, note that $\phi_1 = 0$ corresponds to the latent variable U affecting Y and $\boldsymbol{X}_t$ but not S_t, thus it is a *pure predictor of outcome*: IPW ignores U, whereas LC-IPW accounts for U through the latent classes. The efficiency of the LC-IPW estimator in this situation is also consistent with theoretical results on over-adjustment in inverse probability weighting (Rotnitzky et al., 2010).

Another apparently odd result is the higher efficiency of LC-IPW with respect to the infeasible estimator with true weights. This is due to the flexibility of the estimated weights which account for the patterns in the realized sample, a phenomenon already noted in several settings; see, among others, Rosenbaum (1987).

5 Application: effect of wage subsidies on employment

We illustrate the LC-IPW method outlined in Section 3 through an application to a dataset extracted from the registers compiled by the Finnish Tax Authority from 1995 to 2002. These registers cover the whole population of firms that pay taxes in Finland and also contain information about the received subsidies.

In our application, the research question concerns the effect of wage subsidies on employment, a labor market policy mainly proposed to achieve the aim of long term unemployment reductions. This is an instance of a sequential binary treatment with time-varying covariates, where conventional estimators are likely to be biased due to unobserved confounders, namely latent traits of the firms affecting both the subsidy receipt and observed variables such employment and profit. Therefore, we estimate the effect of the subsidies bythe LC-IPW approach and compare it with the results of OLS regression and the conventional IPW.

5.1 Description of the dataset

Our dataset refers to a sample of $n = 1640$ Finnish firms (manufactures and services) between 20 and 200 employees in the period 1995-2002. Although several types of subsidies are available for firms (e.g. investment and R&D subsidies), the most common type is represented by wage subsidies, which were receivedat least once by about 65% of the firms in the sample. This result depends on the wide eligibility to wage subsidies for Finnish firms that only excludes firms that are non-profitable or threatened of bankruptcy. A firm is eligible to receive a wage subsidy if it just demonstrates that the job is new. For a discussion on the wage subsidy scheme in Finland, see Kangasharju (2007). Note that the subsidy is partial because it complements private wages, in line with the idea that subsidized jobs should tend to exploit public wage incentives to fill the gap between the wage that the firm is willing to pay and the unionized wage level.

The dataset includes for every firm the following variables measured at every year: *employment* (number of employees), *wage* (*total* and *per employee*), *fixed capital*, *sales* and *profit*.

In this application the number of occasions is $T = 8$, the treatment variable S_t is an indicator taking the value 1 if the firm receives a wage subsidy in year t, the outcome is the employment at the end of the period Y, whereas the potential confounders $\boldsymbol{X}_t$ are all the variables measured at the end of year t (possibly including lagged values).

Table 7 shows the observed distribution of the 3121 subsidies by year: the percentage of firms receiving a subsidy falls from 35% in 1995 to about 14% in the last two years. Table 8 reports the observed distribution of the number of subsidies during the eight years: 65% of the firms were subsidized at least once, but less than 14% were subsidized more than 4 times.

Table 7: *Subsidized firms by year*

Year	Number	Percentage
1995	582	35.49
1996	448	27.32
1997	491	29.94
1998	450	27.44
1999	383	23.35
2000	293	17.87
2001	242	14.76
2002	232	14.15
Total	3121	100.00

Table 9 compares subsidized firms (receiving the subsidy at least once in the period) and non-subsidized firms by listing the mean and standard deviation of the variables

Table 8: *Observed distribution of subsidies*

Number of subsidies	% firms	% cumulative
0	34.94	34.94
1	18.54	53.48
2	15.18	68.66
3	10.24	78.90
4	7.44	86.34
5	6.16	92.50
6	4.09	96.59
7	1.71	98.29
8	1.71	100

measured at the last occasion (year 2002): subsidized firms are larger in terms of number of employees and fixed capital, but they have lower profitability and lower wage per employee.

Table 9: *Descriptive statistics about firms in 2002 (monetary variables in thousands of Euros)*

Variable	Subsidized (1067)		Non-subsidized (573)	
	Mean	St.dev.	Mean	St.dev.
Employment	73.57	74.72	60.78	55.74
Wage (total)	1900	2775	1702	1872
Wage (per employee)	24	16	28	14
Fixed capital	3113	15400	2605	9788
Sales	11100	37800	10800	26800
Profit	470	2365	625	3691

5.2 Estimation of the casual effect

The subsidies are expected to affect all the variables considered here. Our focus is on the effect on employment, which represents the aim of the public policy. Therefore, we specify a Marginal Structural Model (MSM) for the effect of the subsidies on the number of employees at the end of the period. We fit two versions of the model:

- M1: $E(Y^{(\boldsymbol{s})}) = \beta_0 + s_+\beta_1$
- M2: $E[\log(Y^{(\boldsymbol{s})})] = \beta_0^* + s_+\beta_1^*$

where $s_+ = \sum_{t=1}^{T} s_t$ is the number of years receiving a subsidy. The two versions of the model differ for the scale of the response (raw vs logarithmic): the parameter β_1 in $M1$ is the causal effect of subsidies on employment expressed as an absolute variation, whereas β_1^* in $M2$ is the same effect expressed as an approximate relative variation.

As a benchmark, Table 10 reports the results for both M1 and M2 using OLS and IPW (standard inverse probability weighting), where the 95% confidence intervals are obtained via non-parametric bootstrap. As expected, the estimates of the causal effect of subsidies on employment are positive and statistically significant. The OLS estimates are $\hat{\beta}_1 = 6.6$ (every year receiving a subsidy entails an average increase of 6.6 employees) and $\hat{\beta}_1^* = 0.075$ (the average relative variation is about 7.5%).

Table 10: *Results for the Marginal Structural Models M1 and M2 using the OLS and IPW methods*

		OLS			IPW		
Model	Parameter	Estimate	95% Conf.Int.		Estimate	95% Conf.Int.	
M1	β_0	62.385	57.853	68.187	67.958	63.306	72.365
	β_1	6.596	3.754	8.986	3.932	2.207	6.052
M2	β_0^*	3.822	3.776	3.876	3.889	3.837	3.945
	β_1^*	0.075	0.054	0.091	0.048	0.034	0.064

The confounding generated by the observed variables can be controlled by the standard IPW method. To compute the weights, the treatment indicators S_t are modeled by a logistic regression with a dummy variable for each year and the following covariates measured at years $t-1$ and $t-2$: treatment indicator, log(employment), log(wage per employee), log(fixed capital), log(sales), and sign(profit) $\sqrt[n]{|\text{profit}|}$. The last transformation is a sort of log-transformation which, however, maintains the sign (positive or negative) of the original variable. The method could be implemented using only the covariates at year $t-1$; however, we added a further time lag to properly account for time dependencies (note that the model for S_1 has no covariates, whereas the model for S_2 only has covariates measured at $t = 1$).

The results of the IPW method are reported in the right-most part of Table 10. The estimate of the causal effect is positive and statistically significant but, compared to OLS, the magnitude is much lower: the estimated effect of each year receiving a subsidy on the employment at end of the period is 3.9 employees (from model $M1$) or about 5% (from model $M2$). A positive causal effect is in line with the hypothesis that, exploiting wage subsidies, firms decrease the marginal cost of additional labor, allowing to employ workers who would not otherwise have been employed. The IPW estimate is about one half of the estimate obtained by Kangasharju (2007) using a slightly larger dataset. Therefore, we argue that the reduction in the magnitude of the IPW estimate is due the capacity of this method to adjust for observed time-varying confounders.

The simulation study of Section 4 showed that unobserved pre-treatment confounders may severely bias the IPW estimator. In this application it is likely that certain characteristics of firms, such as type of organization and management, strongly affect both the employment and the receipt of subsidies. These characteristics can be viewed as unobserved pre-treatment confounders since they are unmeasured and normally change very slowly over time. This kind of confounding can be removed by the LC-IPW method.

The implementation of the LC-IPW method is based on the following auxiliary model: (i) the treatment indicators S_t are modeled by a separate logistic model for every latent class (or mixture component) with the same regressors used for the IPW method; (ii) these regressors are modeled by a separate multivariate regression for every latent class, which includes the lagged versions of these variables and the treatment indicators for the present and the previous occasion (with the exception of the first two occasions). These models have been fitted with a number of latent classes k from 1 to 5. The results of this preliminary analysis are reported in Table 11 in terms of log-likelihood and fit indexes BIC and NEC defined in Section 3: BIC favors a solution with at least $k = 5$ classes, whereas NEC favors a solution with $k = 4$ classes.

Table 11: *Selection of the number of latent classes for the auxiliary model*

k	log-lik.	BIC_k	NEC_k
1	-74380	150100	1.000
2	-70672	143920	0.015
3	-68416	140645	0.010
4	-66589	138227	0.008
5	-65275	136837	0.016

Table 12 reports the estimates of the parameters of the Marginal Structural Models M1 and M2 for k from 1 to 5, so as to perform a sensitivity analysis with respect to the number of latent classes. We also report confidence intervals obtained via non-parametric bootstrap, which may be used to test the hypothesis that the causal effect is equal to 0. Note that the LC-IPW estimates for the single-class solution $k = 1$ are indeed IPW estimates and thus they are equal to those in Table 10. As the number of latent classes increases, the LC-IPW methods increases the adjustment for unobserved confounding and the estimates of the causal effects β_1 and β_1^* diminish monotonically up to $k = 4$, which is the solution suggested by the NEC index. The estimates increase for $k = 5$, but this could be a consequence of the instability of the estimated weights. Therefore, we choose the solution with $k = 4$ classes, yielding $\hat{\beta}_1 = 2.2$ (every year receiving a subsidy entails an average increase of 2.2 employees) and $\hat{\beta}_1^* = 0.035$ (the average relative variation is about 3.5%). The bootstrap confidence intervals confirm that the causal effects are statistically significant.

Table 12: *Results for the Marginal Structural Models M1 and M2 using the LC-IPW method*

Model	k	Parameter	Estimate	95% Conf. interval	
M1	1	β_0	67.958	63.458	72.967
		β_1	3.932	2.217	6.140
	2	β_0	69.079	65.033	74.328
		β_1	3.024	1.361	5.294
	3	β_0	69.450	64.174	74.230
		β_1	2.551	0.846	5.139
	4	β_0	70.280	65.032	75.385
		β_1	2.156	0.257	4.499
	5	β_0	68.725	64.037	77.441
		β_1	2.959	0.238	6.166
M2	1	β_0^*	3.889	3.844	3.940
		β_1^*	0.048	0.033	0.062
	2	β_0^*	3.890	3.841	3.942
		β_1^*	0.044	0.028	0.060
	3	β_0^*	3.893	3.844	3.939
		β_1^*	0.038	0.020	0.057
	4	β_0^*	3.900	3.846	3.955
		β_1^*	0.035	0.015	0.063
	5	β_0^*	3.884	3.831	3.973
		β_1^*	0.042	0.012	0.072

On the basis of the LC-IPW estimates we still conclude that subsidized jobs stimulated employment in Finland firms. Unlike Hujer et al. (2002) for Germany and Dahlberg and Forslund (2005) for Sweden, that using similar data found no increase in employment, we estimate a small positive effect, a result qualitatively in accordance with specific investigations in Finland (Hämäläinen and Ollikainen, 2004). However, even if the proposed LC-IPW method confirms the existence of a positive effect, its magnitude is substantially reduced with respect to OLS (no control for confounding) and IPW (controlling for observed confounders only). For example, β_1 is estimated to be 6.6 using OLS, 3.9 using IPW (a 41% reduction) and 2.2 using LC-IPW (an additional 44% reduction). Those results point out the substantial role of both observed and unobserved confounding in this setting.

Different arguments may help to explain these results. The prevalent interpretation is that, as the subsidy is partial, their use reflects different unobserved abilities of the firm on employment and managerial and financial policies. That is, the actions of more efficient firms in managerial terms are postulated to be positively correlated with the exploiting of wage subsidies as a long run strategy of firm. Thus, while individually firms contribute to the total employment performance heterogeneously, this result appears to be affected by the ability to substitute in an inter-temporal perspective private employment with that subsidized from the public sector. This implies that accounting for unobserved pre-treatment confounders through the proposed LC-IPW method should reduce the potential upward bias in the estimates, a prediction consistent with the findings of Table 12.

6 Conclusions

In this paper, we consider an extension of the Inverse Probability Weighting (IPW) estimator for Marginal Structural Models (Robins et al., 2000), which may be used for causal inference in the presence of certain forms of unobserved confounding with sequential binary treatments. The proposed extension, called Latent Class Inverse Probability Weighting (LC-IPW), is based on two steps: first, a finite mixture model is fitted in order to compute latent-class-specific weights; then, these weights are used to fit the Marginal Structural Model of interest. The properties of the LC-IPW estimator are studied by simulation under different scenarios, whereas its empirical implementation is demonstrated by an application based on a dataset of Finnish firms observed for eight years, where it is of interest to estimate the causal effect of a form of wage subsidy on employment.

The main advantage of the proposed LC-IPW method over the standard IPW method is that it properly corrects for unobserved pre-treatment confounders. This conclusion is rather obvious for the case in which: (*i*) this type of confounding may be represented by a discrete latent variable having a reduced number of levels, corresponding to *latent classes* in the population of interest, and (*ii*) provided that the auxiliary finite-mixture model to compute the weights, and singling out these latent classes, is correctly specified.

However, as shown by the simulation, the reduction of the bias due to the adoption of the proposed approach may also be consistent with unobserved confounding due to latent variables having a continuous distribution.

Even if it may be rather surprising, another relevant advantage of the LC-IPW method is that it may reduce the bias of the standard IPW method even in the absence of unobserved confounding. This is clearly shown by our simulation results and is coherent with some results available in the literature (e.g., Lefebvre et al., 2008). This leads to the important conclusion that the use of the proposed method is advisable even if we are not sure of the presence of unobserved confounding, provided that the sample size is large enough and we do not use an excessive number of latent classes.

Finally, as an advantage we also have to consider that the proposed method may be readily available software for finite mixture models. We mention, in particular, the `R` package `flexmix` (Grün and Leisch, 2008) and the Stata package `fmm` by Kit Baum, which may be use to fit the auxiliary model. We recall that fitting this model is the most challenging part of the proposed approach. In any case, we make available our Matlab implementation to the reader upon request.

Obviously, the advantages of the proposed LC-IPW method over the standard IPW method are at the cost of a greater complexity in formulating the auxiliary finite-mixture model for the weights and, in particular, in choosing the number of the mixture components (latent classes), denoted by k. However, suitable criteria, such as the Bayesian Information Criterion (Schwarz, 1978) or the Normalized Entropy Criterion (Celeux and Soromenho, 1996), may effectively drive this choice. Moreover, as shown in the application about wage subsidy, trying different values of k may be useful from a perspective of sensitivity analysis. It is also important to recall that this complexity may imply that the advantages of the proposed method over the IPW method are consistent only with large sample sizes. This happens for certain configurations of the generating model adopted in the simulation study. However, as clarified above, there are no strong reasons to advise against the use of the proposed estimator even when there is no unobserved confounding.

As a further development, we consider of interest methods to compute standard errors for the estimates of the causal parameters, which are more direct to use with respect to the bootstrap method. Obviously, these methods must take also into account the uncertainty on the auxiliary model parameters. As an extension, it may also be of interest the adoption of the proposed approach even in connection with the longitudinal propensity score method (Achy-Brou et al., 2010). Moreover, we think that it is also possible to account for time-varying unobserved confounders by using a latent Markov model (Bartolucci et al., 2010) to compute the weights. Finally, we consider of interest to adopt, even within the proposed approach, methods to increase the stability of the weights and the efficiency of the resulting estimator of causal effects, such as truncation (Cole and Hernan, 2008). We expect that this technique may lead to advantages in the presence of a small sample or an excessive number of latent classes, when the weights may be unstable due to the reduced number of sample units assigned to some of these classes.

References

Achy-Brou, A. C., Frangakis, C. E., and Griswold, M. (2010). Estimating treatment effects of longitudinal designs using regression models on propensity scores. *Biometrics*, 66.

Bartolucci, F., Farcomeni, A., and Pennoni, F. (2010). An overview of latent Markov models for longitudinal categorical data. *Technical report available at http://arxiv.org/abs/1003.2804*.

Biernacki, C., Celeux, G., and Govaert, G. (1999). An improvement of the NEC criterion for assessing the number of clusters in a mixture model. *Non-Linear Analysis*, 20:267–272.

Celeux, G. and Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification*, 13(2):195–212.

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

Dahlberg, M. and Forslund, A. (2005). Direct displacement effects of labour market programmes. *Scandinavian Journal of Economics*, 107(3):475–494.

Efron, B. and Tibshirani, R. (1994). *An Introduction to the Bootstrap*. Chapman & Hall/CRC, London.

Gill, R. and Robins, J. (2001). Causal inference for complex longitudinal data: the continuous case. *Annals of Statistics*, 29:1785–1811.

Grün, B. and Leisch, F. (2008). Flexmix version 2: Finite mixtures with concomitant variables and varying and constant parameters. *Journal of Statistical Software*, 28:1–35.

Hämäläinen, K. and Ollikainen, V. (2004). Differential effects of active labour market programmes in the early stages of young people's unemployment. Research Reports 115, Government Institute for Economic Research Finland (VATT).

Hujer, R., Caliendo, M., and Radić, D. (2002). Estimating the effects of wage subsidies on the labour demand in west-germany using the IAB establishment panel. Technical report.

Kangasharju, A. (2007). Do wage subsidies increase employment in subsidized firms? *Economica*, 74(293):51–67.

Lechner, M. (2009). Sequential causal models for the evaluation of labor market programs. *Journal of Business & Economic Statistics*, 27:71–83.

Lechner, M. and Miquel, R. (2010). Identification of the effects of dynamic treatments by sequential conditional independence assumptions. *Empirical Economics*, 39:111–137.

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*, 27:3629–3642.

Lefebvre, G. and Gustafson, P. (2010). Impact of outcome model misspecification on regression and doubly-robust inverse probability weighting to estimate causal effect. *The International Journal of Biostatistics*, 6(2):1–24.

McLachlan, G. J. and Peel, D. (2000). *Finite Mixture Models*. Wiley, New York.

Robins, J. (1999). Marginal structural models versus structural nested models as tools for causal inference. In Halloran E., B. D., editor, *Epidemiology: The Environment and Clinical Trials*, pages 95–134. Springer, New York.

Robins, J. M., Hernan, M. A., and Brumback, B. (2000). Marginal structural models and causal inference. *Epidemiology*, 11:550–560.

Rosenbaum, P. R. (1987). Model-based direct adjustment. *Journal of the American Statistical Association*, 82:387–394.

Rotnitzky, A., Li, L., and Li, X. (2010). A note on overadjustment in inverse probability weighted estimation. *Biometrika*, 97:997–1001.

Rubin, D. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66.

Rubin, D. B. (2005). Causal inference using potential outcomes: design, modeling, decisions. *Journal of the American Statistical Association*, 100:322–331.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6:461–464.

van der Wal, W. and Geskus, R. (2011). ipw: An r package for inverse probability weighting. *Journal of Statistical Software*, 43:1–23.