TEST
https://doi.org/10.1007/s11749-023-00877-8

ORIGINAL RESEARCH PAPER

A causal hidden Markov model for assessing effects of multiple direct mail campaigns

Fulvia Pennoni[1] · **Leonard J. Paas**[2] · **Francesco Bartolucci**[3]

Received: 21 February 2023 / Accepted: 24 July 2023

Abstract

We propose assessing the causal effects of a dynamic treatment in a longitudinal observational study, given observed confounders under suitable assumptions. The causal hidden Markov model is based on potential versions of discrete latent variables, and it accounts for the estimated propensity to be assigned to each treatment level over time using inverse probability weighting. Estimation of the model parameters is carried out through a weighted maximum log-likelihood approach. Standard errors for the parameter estimates are provided by nonparametric bootstrap. The proposal is validated through a simulation study aimed at comparing different model specifications. As an illustrative example, we consider a marketing campaign conducted by a large European bank over time on its customers. Findings provide straightforward managerial implications.

Keywords Causal inference · Direct marketing · Expectation–Maximization algorithm · Generalized propensity score · Longitudinal observational data

Mathematics Subject Classification 6208 · 62H30 · 90B60

1 Introduction

Hidden Markov (HM) models (Wiggins 1973; Bartolucci et al. 2013; Zucchini et al. 2016) are widely employed to analyze longitudinal data as they represent a logical extension of the latent class model (Lazarsfeld 1950; Goodman 1974), which is used

✉ Fulvia Pennoni
fulvia.pennoni@unimib.it

[1] Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Via degli Arciboldi, 8, 20126 Milan, Italy

[2] Department of Marketing, Business School, The University of Auckland, Sir Owen G Glenn Building - Bldg 260, 12 Grafton RD, Auckland Central, Auckland 1010, New Zealand

[3] Department of Economics, University of Perugia, Via A. Pascoli, 20, 06123 Perugia, Italy

Published online: 07 September 2023

to discover unobserved clusters of individuals with cross-sectional data. For a recent review of some real-life applications of HM models, see Mor et al. (2021), Visser and Speekenbrink (2022), among others. The basic assumption of HM models without covariates is that, for every individual, the response variables referred to the same time occasion depend solely on a discrete latent variable having a finite number of support points corresponding to latent states. The sequence of these latent variables is assumed to follow a Markov chain of first order, so that it is possible to perform model-based dynamic clustering. Individual covariates may be included in the model in different ways. The formulation in which the covariates affect the distribution of the latent variables is of particular interest as it allows studying how the probability of belonging to the different clusters is affected by individual characteristics and other factors (Bartolucci et al. 2014).

HM models have traditionally been applied in marketing for customer segmentation and to infer switching dynamics of customers between segments over time (Ehrenberg 1965; Poulsen 1982; Paas et al. 2007; Wedel and Kamakura 2012; Chang and Zhang 2016; Vermunt and Paas 2017). Table 7 provided in "Appendix A.1" lists some of the main proposals that appeared in the marketing literature without claiming to be exhaustive since the context is varied. When estimating HM models with covariates related to firms' customer-level marketing activities, see Lemmens et al. (2012), Park et al. (2018) and Kappe et al. (2018) among others, the researcher is confronted with a possible differential selection into marketing stimuli and intensities. For example, in the empirical illustration on which this article is focused, different customers may receive varying numbers of direct mailings promoting financial products of a bank according to their individual characteristics and portfolio.

Many methods have been developed to make endogeneity adjustments deriving from a differential selection. In marketing, the use of an instrumental variable (Angrist et al. 1996) is the most common approach employed for handling selection bias in estimating a linear or nonlinear model. It requires finding relevant covariates independent of unobserved determinants of the outcome (Conley et al. 2012; Park and Gupta 2012). Kumar et al. (2011) proposed a trivariate Tobit HM model that accounts for endogeneity by incorporating factors influencing marketing activity levels as covariates considering the target equation approach proposed by Manchanda et al. (2004); see also Moon et al. (2007), Zhang et al. (2014) and Petrin and Train (2010). As Schweidel et al. (2011) and Papies et al. (2017) suggested, researchers should first include observed confounding variables as covariates in the model. The approach is particularly relevant when the variables causing the bias are observed, but information about the firm implementation of marketing efforts is not available. This situation can commonly occur if, for example, product managers apply different criteria for selecting prospects for marketing campaigns. However, it is unclear whether adding covariates may effectively correct for measured confounding factors, and moreover, it is required that the model is correctly specified.

We cast our proposal in the class of the methodologies based on inverse probability-of-treatment weighting (IPTW) proposed by Robins (1997) and widely employed with the latent class model (Lanza et al. 2013; McCaffrey et al. 2013) as a simple and effective method able to alleviate the possible bias resulting from endogeneity; see also Skrondal and Rabe-Hesketh (2014). A first proposal can be found in Bartolucci et al. (2016) where, differently from the standard potential outcome framework (Rubin 2005), under specific assumptions, potential versions of latent variables are conceived. The approach allows estimating average treatment effects (ATEs) on the general population or on the treated, maximizing the log-likelihood based on longitudinal observational data through the Expectation–Maximization (EM) algorithm (Dempster et al. 1977).

Starting from the proposal of Tullio and Bartolucci (2022), we further develop the causal HM approach proposing a model that corrects for selection effects and accommodating for time unobserved heterogeneity when the treatment is repeatedly assigned over time with varying amounts. The causal HM model on which the approach is based is still estimated by an EM algorithm accounting for individual weights that are preliminary estimated by a suitably formulated multinomial logit model for the probability of receiving different types of treatment given certain confounders. Standard errors for the parameter estimates are obtained by a nonparametric bootstrap method so as to account for the uncertainty on the estimation of the weights.

A simulation study validates the proposal where we consider different scenarios generating the latent potential outcomes and responses, and we compare the estimated parameters with those obtained under a randomized experiment. The reported empirical application assesses changes in customers' financial product portfolios at a bank due to direct mail campaigns proposing a novel approach for endogeneity correction when managers made time varying decisions on the marketing stimuli using observed customer relationship management (CRM) data to select prospects most likely to respond positively to the campaign.

As for the organization of the paper, in Sect. 2 we provide a detailed description of the proposed causal HM approach. In Sect. 3, we illustrate the simulation design and show the simulation results. In Sect. 4, we illustrate the marketing data, and we apply the proposal presenting the empirical results and we report a comparison with alternative HM model formulations. Section 5 concludes with a summary and brief discussion. Additional details on the application and simulation results are provided in "Appendix A".

2 Methodology

In the following, we introduce the conceptual framework of the proposal, and the causal HM model. Then, we illustrate the inferential approach for estimating the model parameters based on the preliminary estimation of the individual weights.

2.1 Conceptual rationale of the causal hidden Markov model

The proposed approach is consistent with the potential outcomes framework, see Holland (1986) and Rubin (2005) among others, where the causal effect of a certain treatment is typically of interest. Apart from the case of perfectly randomized experiments (Neyman 1923), the effect of a treatment across the population, such as the ATE on an outcome, cannot be directly estimated because the individual's response can only be observed for the treatment s/he effectively received. When data are collected from observational studies, the IPTW method (Robins et al. 2000) provides estimates of the parameters of the marginal model according to the probability of receiving each possible treatment, which is preliminary estimated on a set of candidate confounders. The method can also be employed to correct for endogeneity when unobserved endogenous variables correlate with observed confounders (Rosenbaum 2020).

The IPTW estimator within the HM framework was proposed in Bartolucci et al. (2016) to estimate the ATEs of different academic degrees on the type of contract, skills, and gross income at the beginning of the working careers of graduates with longitudinal observational data. For a related formulation, applied to assess the role of remittances for alleviating poverty, see Tullio and Bartolucci (2022). In the empirical illustration reported in Sect. 4, we aim to estimate the ATEs of a direct marketing campaign consisting of different intensities of interventions on customers' financial product portfolio at a bank. We consider dynamic counterfactuals, and we face the problem that the marketing stimuli are administered at different times and managers increase or decrease advertising activities each year. Therefore, time-varying individual weights are estimated on the basis of the available confounders referred to the period before each campaign is conducted by means of a multinomial logit model. The set of confounders may include any covariate collected up to the instant before the customer receives the advertising.

The current proposal differs from the other methods currently employed in the marketing literature; see Table 7 in "Appendix A.1" for a summary of some selected studies. To the best of our knowledge, when the HM model is employed, only the transition probabilities are directly affected by the marketing activities along with the other covariates without adjusting for observed confounders. In the proposed HM model, we also assume the initial segment memberships to be influenced by the marketing instruments. Moreover, the fact that the marketing campaign is perpetuating differently over time, causing an endogeneity with varying effects, is not specifically accounted for in the previous literature.

2.2 Model assumptions

With specific reference to the application in marketing illustrated in Sect. 4, let Z_{it} be the variable indicating the marketing stimuli for customer i at each time occasion t, with $i = 1, \dots, n$ and $t = 1, \dots, T$, where n is the number of customers and T denotes the number of observation times. This is an ordinal variable with l levels ranging from 0, in the absence of direct mails, to $l - 1$, in the case of the highest number of mails received from the customer. The financial product portfolio of each customer is defined according to a set of r binary variables taking value 1 if the customer owns a certain product and 0 otherwise, and collected in the vector denoted as $\boldsymbol{Y}_{it}$ for customer i at time occasion t. Time-varying observed confounders collected in the vector denoted as $\boldsymbol{X}_{it}$ refer to characteristics of the customers such as age and number of transactions conducted with the bank.

As in Lanza et al. (2013), Bartolucci et al. (2016) and Tullio and Bartolucci (2022), even within the current proposal potential outcomes are never observed. These are defined as latent customer- and time-specific variables denoted as $H_{it}^{(z_1,\ldots,z_t)}$, with $i = 1,\ldots,n$, $t = 1,\ldots,T$, $z_1,\ldots,z_t = 0,\ldots,l-1$, having a discrete distribution with support points $h = 1,\ldots,k$ and unspecified distribution. Therefore, on each time occasion, every customer has many latent potential states associated with each treatment level. For instance, $H_{i2}^{(0,2)}$ and $H_{i2}^{(0,3)}$ denote the potential product of prospect i at the second time occasion when s/he is exposed to marketing stimuli with intensity 2 or 3, respectively, with the absence of stimuli at the previous occasion. Let $\boldsymbol{H}_i^{(z_1,\ldots,z_T)} = (H_{i1}^{(z_1)},\ldots,H_{iT}^{(z_1,\ldots,z_T)})'$ be the vector collecting the sequence of latent potential outcomes. The longitudinal data structure is taken into account assuming that the distribution of the latent potential outcomes follows a Markov chain of first order.

Initial and transition probabilities of the latent Markov chain are suitably parameterized in order to express the ATEs; see also Bartolucci et al. (2016). In particular, the initial probabilities are function of the treatment as follows:

$$\log\frac{p(H_{i1}^{(z_1)}=h)}{p(H_{i1}^{(z_1)}=1)} = \alpha_h + \boldsymbol{d}(z_1)'\boldsymbol{\beta}_h, \qquad z_1 = 0,\ldots,l-1,\ \ h = 2,\ldots,k, \tag{1}$$

where α_h is the intercept specific for each latent state, $\boldsymbol{\beta}_h = (\beta_{h2},\ldots,\beta_{hl})'$ is a column vector of $l-1$ regression parameters, and $\boldsymbol{d}(z_1)$ is a column vector of $l-1$ zeros with the (z_1-1)th element equal to 1 if $z_1 > 0$. Since each element β_{hz_1} of $\boldsymbol{\beta}_h$ for $z_1 > 1$ is a shift parameter from the first category to the hth it can be interpreted as the ATE. With reference to the empirical illustration, it represents the average effect of the z_1th direct mail condition with respect to absence of treatment at the first time occasion.

Transition probabilities are modeled through a multinomial logit parameterization based on the following expression:

$$\log\frac{p(H_{it}^{(z_1,\ldots,z_t)}=h \mid H_{i,t-1}^{(z_1,\ldots,z_{t-1})}=\bar{h})}{p(H_{it}^{(z_1,\ldots,z_t)}=1 \mid H_{i,t-1}^{(z_1,\ldots,z_{t-1})}=\bar{h})} = \gamma_{\bar{h}h} + \boldsymbol{d}(z_t)'\boldsymbol{\delta}_h, \tag{2}$$

with $\bar{h} = 1,\ldots,k$, $h = 2,\ldots,k$, $t = 2,\ldots,T$, $z_1,\ldots,z_t = 0,\ldots,l-1$, where $\gamma_{\bar{h}h}$ is an intercept specific for each transition and $\boldsymbol{\delta}_h = (\delta_{h2},\ldots,\delta_{hl})'$ are vectors of regression coefficients. Each parameter δ_{hz} may be interpreted in terms of average causal effect of the treatment referred to the transition probability from state 1 to state h.

According to this parametrization, which is proposed and justified in Bartolucci et al. (2017), the effect is referred to the destination state. It is more parsimonious than parametrizations in which the effect depends on the starting and destination states, with advantages, in particular, when many latent states are estimated.

Finally, the latent potential outcomes are related to the response variables Y_{ijt} collected in the vectors $\boldsymbol{Y}_{it}$ by the conditional probabilities

$$\phi_{jy|h} = p\left(Y_{ijt} = y \mid H_{it}^{(z_1,\ldots,z_t)} = h\right), \quad h = 1,\ldots,k,\ j = 1,\ldots,r,\ y = 0,1, \tag{3}$$

for all t. We assume that the latent potential outcomes are the only variables that influence the observable responses and the effect of the pre-treatment covariates passes through these latent variables. This amounts to assuming that the response variables in every vector $\boldsymbol{Y}_{it}$ are conditionally independent given $H_{it}^{(z_1,\ldots,z_t)}$, and it implies that

$$p(\boldsymbol{Y}_{it} = \boldsymbol{y} \mid H_{it}^{(z_1,\ldots,z_t)} = h) = \prod_{j=1}^{r} \phi_{jy_j|h}, \tag{4}$$

where $\boldsymbol{y}$ is a possible realization of $\boldsymbol{Y}_{it}$ with elements $y_1,\ldots,y_r$.

The following assumptions are required to identify the ATEs; see also Tullio and Bartolucci (2022). The first is the Stable Unit Treatment Value Assumption (SUTVA), according to which the individual treatments are completely represented without interactions between the members of the population. The second assumption is that of exogeneity, requiring that confounders collected into the vector $\boldsymbol{X}_{it}$ are not influenced by the treatment. The third is the positivity assumption, that is, $0 < P(Z_{it} = z \mid \boldsymbol{X}_{it} = \boldsymbol{x}) < 1$ for $i = 1,\ldots,n$, $t = 1,\ldots,T$, $z = 0,\ldots,l-1$, and all $\boldsymbol{x}$ denoting the realized values of the observed covariates. This is also related to sequential ignorability given confounders that correspond to the longitudinal version of the assumption of absence of unmeasured confounders (unconfoundness) (Rosenbaum and Rubin 1983; Rubin 1990). Referred to the empirical application presented in Sect. 4, it implies that every customer has a positive probability of receiving any marketing stimulus at each time occasion in which it is administered, and this probability is mainly related to the observed covariates. Formally, this assumption may be expressed as

$$Z_{i1},\ldots,Z_{it} \perp\!\!\!\perp H_{it}^{(z_1,\ldots,z_t)} \mid \boldsymbol{X}_{it}, \tag{5}$$

for $i = 1,\ldots,n$, $t = 1,\ldots,T$, and $z_1,\ldots,z_t = 0,\ldots,l-1$. Therefore, the covariates collected into vector $\boldsymbol{X}_{it}$ are assumed to be sufficiently informative and they include all variables directly influencing both the treatment and the responses.

2.3 Parameter estimation

Let $\boldsymbol{x}_{it}$ be the observed vector of covariates for customer i, $i = 1,\ldots,n$, at time t, $t = 1,\ldots,T$, $\boldsymbol{y}_{it}$ be that of the responses, and z_{it} be the received treatment. Developing the approach proposed in Bartolucci et al. (2016), we adopt an estimation procedure which relies on two steps aimed at obtaining individual weights, based on the probability of the received treatment, and in maximizing a weighted log-likelihood.

2.3.1 Estimation of individual weights

Under the assumptions formulated above, and considering (5) in particular, the probability of treatment assignment is estimated dynamically for each time period. We note that the observed responses prior to treatment can be added among the covariates used to estimate the propensity score at a certain time period. With respect to the empirical illustration reported in Sect. 4, we refer to the probability of each customer to receive her/his own number of mails at repeated time t, $t = 1, \dots, T$. This probability is estimated through the following multinomial logit model:

$$\log \frac{p(z_{it} = z \mid \boldsymbol{x}_{it})}{p(z_{it} = 0 \mid \boldsymbol{x}_{it})} = \eta_z + \boldsymbol{x}'_{it}\boldsymbol{\lambda}_z, \qquad z = 1, \dots, l-1,\ t = 1, \dots, T, \tag{6}$$

where η_z and $\boldsymbol{\lambda}_z$ are the intercept and regression parameters, respectively. The estimated time-varying weights are determined as the inverse of the estimated probability for each treatment type as

$$\hat{w}_{it} = n \frac{1/\hat{p}(z_{it} \mid \boldsymbol{x}_{it})}{\sum_{i=1}^{n} 1/\hat{p}(z_{it} \mid \boldsymbol{x}_{it})}, \qquad i = 1, \dots, n,\ t = 1, \dots, T.$$

Then, the overall weight for each customer is given by

$$\hat{w}_i = \prod_{t=1}^{T} \hat{w}_{it}, \qquad i = 1, \dots, n. \tag{7}$$

It is worth noting that, similarly to the logic used in survey sampling and missing data problems, the above defined weights create pseudo-populations of customers corresponding to treatment intensities. These weights have the role to mimic a marketing campaign conducted under a completely randomized experiment. Huge weights for a few individuals can lead to instability and less precision of the ATEs. Limiting the weights to a maximum with a trimming method is suggested in such a situation (Robins and Rotnitzky 1995; Stuart 2010).

2.4 Maximization of the weighted log-likelihood

On the basis of the observed data, the weighted model log-likelihood is

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log p(\boldsymbol{y}_{i1}, \dots, \boldsymbol{y}_{iT} \mid z_{i1}, \dots, z_{iT}), \tag{8}$$

where $\boldsymbol{\theta}$ is the vector of all model parameters arranged in a suitable way. The *manifest probability* $p(\boldsymbol{y}_{i1}, \dots, \boldsymbol{y}_{iT} \mid z_{i1}, \dots, z_{iT})$ is computed by recursions developed in the literature (Baum et al. 1970; Welch 2003) and applied to a reduced form of the HM model for longitudinal data that directly derives from the causal model illustrated in Sect. 2.2. The reduced form is based on the same conditional response probabilities expressed as in (3) and (4) with $H_{it}^{(z_1,\dots,z_t)}$ substituted by H_{it}, on the basis of the consistency rule, and initial and transition probabilities for the latent process that directly derive from (1) and (2). In more detail, this model is based on individual sequences of latent variables $\boldsymbol{H}_i = (H_{i1}, \dots, H_{iT})'$ that are related to the sequences of latent potential outcomes $\boldsymbol{H}_i^{(z_1,\dots,z_T)}$ and with initial probabilities formulated as

$$\log \frac{p(H_{i1} = h \mid z_{it})}{p(H_{i1} = h \mid z_{it})} = \alpha_h + \boldsymbol{d}(z_{it})'\boldsymbol{\beta}_h, \qquad h = 2, \dots, k, \tag{9}$$

and transition probabilities formulated as

$$\log \frac{p(H_{it} = h \mid H_{i,t-1} = \bar{h}, z_{it})}{p(H_{it} = 1 \mid H_{i,t-1} = \bar{h}, z_{it})} = \gamma_{\bar{h}h} + \boldsymbol{d}(z_{it})'\boldsymbol{\delta}_h, \quad \bar{h} = 1, \dots, k,\ h = 2, \dots, k. \tag{10}$$

Since the previous latent configuration is not known for each customer, the EM algorithm (Baum et al. 1970; Dempster et al. 1977) maximizes the observed data log-likelihood by alternating the following two steps until convergence:

- *E-step*: compute the expected value of the complete data log-likelihood $\ell^*(\boldsymbol{\theta})$, corresponding to $\ell(\boldsymbol{\theta})$, given the observed data and the current value of the parameters;
- *M-step*: update $\boldsymbol{\theta}$ by maximizing the expected value of $\ell^*(\boldsymbol{\theta})$ obtained at the E-step.

Details about the implementation of this algorithm may be found in Bartolucci et al. (2014) among others.

Standard errors for the estimated parameters, and the ATEs in particular, may be obtained by applying a nonparametric bootstrap (Davison and Hinkley 1997) based on resampling units with replacement from the original sample, along with the observed pre-treatment covariates, treatment, and responses, a suitable number of times.

Finally, selection of the number of latent states, k, corresponding to the customers's segments may be based on information criteria such as the Bayesian information criterion (BIC) (Schwarz 1978). Other indices are also proposed in the literature; for a comparative study, see Bacci et al. (2014).

Overall, we notice that the causal HM approach is relatively easy to apply since estimation can be carried out by packages such `LMest` (Bartolucci et al. 2017) in `R`.

3 Simulation study

We performed an extensive simulation study, which complements that in Tullio and Bartolucci (2022), considering two different models to generate the latent potential outcomes, response variables, and treatment assignment mechanisms. In this way, we can assess the finite sample properties of the proposed estimator. Other theoretical asymptotic properties, such as consistency, concerning a similar estimator can be found in Bartolucci et al. (2016), and they are also valid for the current proposal. We considered two different sample sizes, n = 5000, 10,000, two different numbers of time occasions, $T = 4, 8$, and $r = 5$ response variables for all i and t. The treatment variable Z_{it} may assume $l = 3$ levels from 0 to 2. The simulation also includes two covariates: the first is a continuous variable and it is generated from a standard Gaussian distribution, whereas the second is a discrete variable assuming values -1 or 1 with probability 0.5. The values of the two covariates generated for individual i at time occasion t are collected in the vectors $\boldsymbol{x}_{it}$, where $i = 1, \dots, n$ and $t = 1, \dots, T$.

The first model to generate the potential latent variables and the response variables, denoted by M_1, is an HM model with $k = 3$ latent states assuming for all i the following parametrization for the initial probabilities:

$$\log \frac{p\big(H_{i1}^{(z_1)} = h \mid \boldsymbol{x}_{i1}\big)}{p\big(H_{i1}^{(z_1)} = 1 \mid \boldsymbol{x}_{i1}\big)} = \alpha_h^* + \boldsymbol{d}(z_1)'\boldsymbol{\beta}_h^* + \boldsymbol{x}_{i1}'\boldsymbol{\tau}_h^*, \quad h = 2, 3,$$

where $\alpha_2^* = -0.3$, $\alpha_3^* = -2$, $\boldsymbol{\beta}_1^* = (0.6, 2)'$, $\boldsymbol{\beta}_2^* = (2, 4)'$, $\boldsymbol{\tau}_2^* = (0.5, 0.5)'$, and $\boldsymbol{\tau}_3^* = (1, 1)'$; these values are chosen so that for $z_1 = 1$ the marginal probability (with respect to the covariates) of each category of latent state h is approximately equal to 1/3, whereas when $z_1 = 0$, the probability of the first state increases and that of the last decreases, and vice versa when $z_1 = 2$. At the same time, for any z_1 and as the covariate values increase, the conditional probability of the last state increases and that of the other states decreases. Regarding the transition probabilities, model M_1 assumes for all i and for $\bar{h} = 1, 2, 3$ and $t = 2, \dots, T$ that

$$\log \frac{p\big(H_{it}^{(z_1,\dots,z_t)} = h \mid H_{i,t-1}^{(z_1,\dots,z_{t-1})} = \bar{h}, \boldsymbol{x}_{it}\big)}{p\big(H_{it}^{(z_1,\dots,z_t)} = 1 \mid H_{i,t-1}^{(z_1,\dots,z_{t-1})} = \bar{h}, \boldsymbol{x}_{it}\big)} = \gamma_{\bar{h}h}^* + \boldsymbol{d}(z_t)'\boldsymbol{\delta}_h^* + \boldsymbol{x}_{it}'\boldsymbol{\psi}_h^*, \quad h = 2, 3,$$

where $\boldsymbol{\delta}_h^* = \boldsymbol{\beta}_h^*$ and $\boldsymbol{\psi}_h^* = \boldsymbol{\tau}_h^*$ for all h with the same arguments as above, with

$$\gamma_{\bar{h}h}^* = \begin{cases} -2.3, & \bar{h} = 1, \ h = 2, \\ -4.0, & \bar{h} = 1, \ h = 3, \\ 1.7, & \bar{h} = 2, \ h = 2, \\ -2.0, & \bar{h} = 2, \ h = 3, \\ -0.3, & \bar{h} = 3, \ h = 2, \\ 0.0, & \bar{h} = 3, \ h = 3, \end{cases}$$

to ensure a certain level of persistence and that the latent states are visited approximately the same number of times overall. Given a certain sequence of treatments $z_1, \dots, z_T$ for individual i and once the potential latent variables are drawn, the response variables Y_{ijt} are generated for all $j = 1, \dots, r$ and $t = 1, \dots, T$ from a Bernoulli distribution with success probabilities

$$\phi_{jy|h} = \begin{cases} 0.2, & h = 1, \\ 0.5, & h = 2, \\ 0.8, & h = 3. \end{cases}$$

The second data generating model used within the simulation, denoted by M_2, is a version of the previous one based on a continuous latent variable denoted by $C_{it}^{(z_1,\dots,z_t)}$ and on an autoregressive structure of first order. For the first time occasion, the model assumes that

$$C_{i1}^{(z_1)} = \varrho + \boldsymbol{x}_{i1}'\boldsymbol{\varsigma} + z_1\upsilon + \varepsilon_{i1},$$

where $\varrho = -1$, $\boldsymbol{\varsigma} = (0.5, 0.5)'$, and $\upsilon = 1$, while the error terms ε_{i1} are independent with standard Gaussian distribution. For the other time occasions, it assumes that

$$C_{it}^{(z_1,\dots,z_t)} = C_{i,t-1}^{(z_1,\dots,z_{t-1})}\rho + \boldsymbol{x}_{it}'\boldsymbol{\varsigma} + z_t\upsilon + \eta_{it}, \quad t = 2,\dots,T,$$

where $\rho = 0.5$ and the error terms η_{it} are independent with distribution $N(0, 0.5)$. Finally, every response variable Y_{ijt} is generated from a Bernoulli variable with probability of success equal to $\exp(C_{it})/[1 + \exp(C_{it})]$, where C_{it} is the selected value according to the assigned treatment at each time occasion.

Regarding the treatment assignment mechanism, we considered, first of all, a randomized scheme, denoted by R_1, in which each variable Z_{it} assumes values from 0 to $l - 1$ with the same probability $1/l$ each. We considered an alternative mechanism in which the assignment depends on the covariates to generate a form of endogeneity when the covariates are ignored. Under this scheme, denoted by R_2, every Z_{it} depends on the realizations of two time-varying exogenous covariates (X_{i1t}, X_{i2t}), one from a standard Gaussian distribution and the other from a Bernoulli distribution. In particular, the treatment Z_{it} is assigned assuming the multinomial logit model defined in (6) with $\eta_1 = 0.6$, $\eta_2 = 0$, $\boldsymbol{\lambda}_1 = (1, 1)'$, and $\boldsymbol{\lambda}_2 = (2, 2)'$, so that the three treatment levels have probability equal to around 1/3 marginally with respect to the covariates.

For any combination of the data generating model, M_1 or M_2, and of treatment assignment mechanism, R_1 and R_2, we simulated B = 1000 samples, and for each sample, we estimated three different models: (i) the naive model based on parametrization (9) and (10) without weights (unweighted); (ii) the corresponding model based on extending the previous equations as follows

$$\log\frac{p(H_{i1} = h \mid z_{it}, \boldsymbol{x}_{it})}{p(H_{i1} = h \mid z_{it}, \boldsymbol{x}_{it})} = \alpha_h + \boldsymbol{d}(z_{it})'\boldsymbol{\beta}_h + \boldsymbol{x}_{i1}'\boldsymbol{\tau}_h, \quad h = 2,\dots,k, \tag{11}$$

$$\log\frac{p(H_{it} = h \mid H_{i,t-1} = \bar{h}, z_{it}, \boldsymbol{x}_{it})}{p(H_{it} = 1 \mid H_{i,t-1} = \bar{h}, z_{it}, \boldsymbol{x}_{it})} = \gamma_{\bar{h}h} + \boldsymbol{d}(z_{it})'\boldsymbol{\delta}_h + \boldsymbol{x}_{it}'\boldsymbol{\psi}_h, \tag{12}$$

the second for $\bar{h} = 1,\dots,k$, $h = 2,\dots,k$, without weights (covariate); (iii) the proposed approach based on weights computed as in (7) and parametrization (9) and (10) described in Sect. 2.4 (weighted).

Tables 8–9 shown in "Appendix A.2" report the results of the study performed using B = 1000 bootstrap replications generated as illustrated above under the HM models: (i) unweighted; (ii) covariate; and (iii) weighted. The results are evaluated in terms of the average values, bias (in absolute value), and standard deviation.

From Table 8, reporting the results under model M_1, we observe that all the estimated average values of the ATEs on the initial and transition probabilities under the proposed model are improved over the unweighted HM models, considering the results under randomization as a benchmark when computing the bias. Moreover, the HM model with covariates still outperforms that with the only treatment. The bias taken in absolute value is smaller under the proposed approach. With the proposed estimator, the bias slightly increases when T is larger, and it decreases when n increases. We also notice that with our proposal reasonable standard deviations result, which are only slightly larger than those of the unweighted estimator and the estimator with covariates, and they decrease when the sample size increases at a rate close to $\sqrt{n}$.

From Table 9, reporting the results when the data are generated under model M_2 with continuous latent variables, we notice that the causal HM estimator outperforms alternatives. The unweighted HM model shows a bias up to two times higher than that of our proposal for some parameters. When $n = 10{,}000$, the bias is relatively low (below 0.9 with $T = 8$) while that of the other estimators is still high (it reaches the value of 3 for some parameters). The proposed estimators show low standard deviations in line with those of the other estimators, particularly when $n = 10{,}000$ and $T = 8$.

We conclude that with an endogenous treatment, the unweighted estimator is biased, and the estimator which only consider the covariates without weights cannot account properly for the observed confounding (Rosenbaum 1987; Joffe et al. 2004; Rosenbaum 2020). Therefore, the proposal is confirmed as superior in terms of bias and standard deviation with respect to competitive methods also when the potential latent variables and the response variables are generated assuming the HM model or its version with continuous variables.

4 Analysis of the marketing campaigns

A large anonymous European bank conducted a marketing campaign over three years to increase the ownership of the following products among customers: loans, credit cards, checking accounts, investment products, mortgages, savings accounts, and a paid phone service enabling customers to gain insights into their account balances. Data are collected on a representative random sample of 49,967 customers aged 18 years and older from December 31, 2000, to December 31, 2001. Note that direct mail was the dominant channel for making product offers to customers at the time. Nowadays, this is still an important communication channel for many firms.

The managers conducted the campaign by using transactional data stored in a data warehouse for prospect selection purposes. This choice was sometimes based on simple common sense heuristics, that is, customers between 30 and 50 years of age with an income of at least 3000 Euro a month were chosen to receive a mail promoting a savings account. At other times, logistic regression models or regression trees were used for prospect selection, akin to the modelling approaches described in Knott et al. (2002).

Some clients left the bank before 31 December 2001 (6.2%), and more left on or before 31 December 2002 (10.0%) and at or before 31 December 2003 (13.3%). For these churned clients, the number of mailings is missing in their final year at bank, and

Table 1 Observed proportions and number of customers according to the direct mail intensity by year

Mail intensity	2001	2002	2003	Number of customers
None	0.318	0.237	0.186	15,885
1–2	0.311	0.323	0.248	15,515
3–5	0.221	0.211	0.245	11,096
≥ 6	0.150	0.229	0.321	7471

it is not clear how many mailings these clients would have received if they had stayed at the bank the entire year. Therefore, we impute the number of mailings they received in the year prior to mailing corrected for the differences in the average number of mailings across the two years of interest.

4.1 Estimation of weights

The following categories are considered to define the treatment intensity: one, one or two, three to five, and more than five mails. The number of treated customers and proportions according to the yearly mails received are shown in Table 1. The model assumes that the treatment is sequentially ignorable given the observed confounders according to the assumptions listed in Sect. 2.1.

Descriptive statistics for the pre-treatment covariates collected in 2000 are listed in Table 10 in "Appendix A.3". These are the observed confounders used to estimate the propensity to be assigned to a certain treatment with respect to the marketing campaign conducted in 2001, whereas covariates collected in 2001 and 2002 are used to measure this propensity for the campaigns conducted in 2002 and 2003, respectively. For the first campaign, lagged responses, listed at the bottom of Table 10 in "Appendix A.3", are considered as well as confounders, whereas for years 2001 and 2002, they are excluded in the propensity score model illustrated in Sect. 2.3.

As can be seen from Table 10 customers who own a credit card are overrepresented among customers receiving more intensive treatment, while they are underrepresented among untreated customers. Therefore, the estimated weights have values deviating from one to those customers. Table 2 summarizes estimates of the model as in equation (6).

4.2 Results

Model selection is performed estimating the causal HM models having homogeneous transition probabilities with the number of latent states h ranging from 1 to 8. BIC values shown in Table 3 suggest a model with seven latent states. Table 4 reports the estimated response conditional probabilities defined in Eq. (3). The estimated customer's segments are ordered according to the average number of products owned. These segments are labeled at the bottom of the table with respect to the estimated conditional probabilities of product ownership. At the beginning of the period, considering the estimated averaged (over individuals) initial probabilities, the highest

Table 2 Estimates of the multinomial logit model parameters $\boldsymbol{\lambda}_z$, $z = 1, \ldots, l$, as in Eq. (6) for the direct mail intensity in 2001 (none is baseline category) with confounders (the reference category for money transferred is that not listed)

Effect		Mail intensity in 2001		
		1–2	3–5	≥ 6
Intercept		1.959	0.125	−3.412
Transactions		0.000	0.001	0.001
Profits		0.000	0.000	0.000
Age		−0.060	−0.006	0.079
Age^2		−0.061	−0.005	0.078
Loans		0.005	0.567	1.064
Credit cards		0.137	0.449	0.731
Investment products		0.306	1.057	1.962
Mortgages		0.686	1.316	2.030
Savings		0.286	0.494	0.736
Online phone service		−0.149	0.060	0.259
Money transferred	≤ 965	0.339	0.405	0.729
	966–1188	0.251	0.689	1.205
	1189–1883	−0.088	0.350	1.044
	≥ 1883	−0.375	−0.398	−0.327

All coefficients are significant at 1%

Table 3 Maximum log-likelihood, number of parameters, and BIC index under the causal HM models estimated with a number of latent states ranging from 1 to 8

h	$\ell(\hat{\boldsymbol{\theta}})$	#par	BIC
1	−12854308.63	6	25708682.32
2	−12518867.54	23	25038067.95
3	−11416417.95	41	22833429.27
4	−10558207.42	61	21117297.67
5	−10122811.53	83	20246824.30
6	−9930851.72	107	19863252.02
7	−9382009.71	133	18765944.30
8	−9452620.47	161	18907571.05

proportion of customers (40%) is allocated to segment 2 defined as "checking account only", while none of the customers (0%) are allocated to the churned segment (segment 1). The other five segments are characterized by ownership probabilities of one or close to one for the checking account, and nonzero for one or more other products. Segment 3 represents 5% of customers, and it is labeled as the "savers' segment" since we find that 98.2% of these customers own the savings account next to the checking account and a smaller proportion owns investment products (10.6%). Segment 4 labeled as "investors" includes 23% of the customers, and it is characterized for owning the savings account (24.0%), the investment products (49.8%), and the mortgage

Table 4 Estimated conditional probabilities $\hat{\phi}_{jy|h}$, $h = 1, \ldots, k$, of the financial products $j = 1, \ldots, 7$, under the causal HM model with $k = 7$ hidden states and labels of each customer segment

Products	Latent state (h)						
	1	2	3	4	5	6	7
Loans ($j = 1$)	0.000	0.003	0.006	0.017	0.022	0.951	0.046
Credit cards ($j = 2$)	0.001	0.006	0.004	0.010	0.014	0.273	0.995
Check accounts ($j = 3$)	0.000	1.000	1.000	0.999	1.000	1.000	1.000
Investment products ($j = 4$)	0.000	0.000	0.106	0.498	0.144	0.001	0.159
Mortgages ($j = 5$)	0.003	0.000	0.001	0.453	0.027	0.016	0.043
Savings accounts ($j = 6$)	0.000	0.012	0.982	0.240	0.469	0.255	0.538
Online phone service ($j = 7$)	0.000	0.002	0.003	0.019	0.993	0.263	0.434
Segments	*none*	*check*	*savers*	*investors*	*phone*	*loan*	*actives*

Table 5 Estimated average initial probabilities under the causal HM model with $k = 7$ latent states according increasing intensities of the marketing stimuli

Mail intensity	Latent state (h)						
	1	2	3	4	5	6	7
None	0.000	0.478	0.205	0.028	0.102	0.038	0.149
1–2	0.000	0.440	0.224	0.044	0.101	0.042	0.145
3–5	0.000	0.384	0.251	0.054	0.110	0.055	0.149
≥ 6	0.000	0.180	0.298	0.106	0.131	0.072	0.213

(45.3%). Segment 5 represents 11% of the customers, and it has been labelled as that of "phone service customers" since 99.3% of the customers own the phone service next to the checking account, and they also have relatively high probabilities for owning investment products (14.4%) and savings account (46.9%). We find that 95.1% of the customers in segment 6 own the loan next to the checking account, and they also have relatively high probabilities of owing all other products, except the investment products or the mortgage; hence, segment 6 is labelled "loan customers" and includes 16% of the customers. The percentage of customers allocated in segment 7 is 5%, and they have labelled "actives", as in this cluster customers have the highest average probably to own each of the products.

Table 11 in "Appendix A.3" displays the ATEs on the initial segment membership probabilities as in Eq. (9), where segment 2 ("checking account only") is considered as a reference category. These estimated parameters may be clearly interpreted through the estimated average initial probabilities reported in Table 5. We notice that, in 2001, receiving at least one direct mail reduces the probability of being allocated to the churn segment (segment 1). Receiving more direct mailings enhances the probabilities in segments labelled as "investors" and "loan customers", respectively.

Table 12 in "Appendix A.3" lists the estimated ATEs on the transition probabilities as in Eq. (10). These parameters may be interpreted through the estimated average

Table 6 Estimated average transition probabilities under the causal HM model with $k = 7$ latent states according to increasing marketing stimuli

Mail intensity	Latent state ($\bar{h}$)	Latent state (h)						
		1	2	3	4	5	6	7
None	1	1.000	0.000	0.000	0.000	0.000	0.000	0.000
	2	0.133	0.823	0.024	0.000	0.012	0.003	0.004
	3	0.040	0.040	0.903	0.000	0.012	0.001	0.005
	4	0.099	0.000	0.000	0.895	0.003	0.000	0.002
	5	0.079	0.019	0.002	0.000	0.890	0.000	0.010
	6	0.039	0.080	0.019	0.000	0.005	0.854	0.003
	7	0.052	0.020	0.012	0.001	0.004	0.000	0.911
1–2	1	1.000	0.000	0.000	0.000	0.000	0.000	0.000
	2	0.041	0.906	0.022	0.000	0.021	0.004	0.005
	3	0.013	0.046	0.912	0.000	0.022	0.001	0.006
	4	0.010	0.000	0.000	0.987	0.002	0.000	0.001
	5	0.015	0.013	0.001	0.000	0.964	0.000	0.007
	6	0.010	0.072	0.015	0.000	0.007	0.895	0.002
	7	0.016	0.021	0.011	0.003	0.007	0.000	0.942
3–5	1	1.000	0.000	0.000	0.000	0.000	0.000	0.000
	2	0.061	0.863	0.025	0.000	0.036	0.007	0.009
	3	0.017	0.040	0.899	0.000	0.033	0.001	0.010
	4	0.016	0.000	0.000	0.979	0.003	0.000	0.002
	5	0.014	0.007	0.001	0.000	0.970	0.000	0.008
	6	0.009	0.044	0.010	0.000	0.007	0.926	0.003
	7	0.012	0.010	0.006	0.001	0.006	0.000	0.964
≥ 6	1	1.000	0.000	0.000	0.000	0.000	0.000	0.000
	2	0.072	0.808	0.030	0.000	0.056	0.020	0.015
	3	0.017	0.031	0.894	0.000	0.042	0.003	0.014
	4	0.023	0.000	0.000	0.967	0.006	0.000	0.004
	5	0.011	0.005	0.001	0.000	0.976	0.000	0.009
	6	0.004	0.014	0.004	0.000	0.004	0.972	0.002
	7	0.009	0.006	0.005	0.001	0.006	0.000	0.974

transition matrices for 2002 and 2003 reported in Table 6. According to the results, it is worth emphasizing that at least one mail to every customer should be sent yearly. In fact, looking at the first column in the top section of Table 6, customers in the direct mail intensity category "none" have high probabilities of switching into the churned segment; for untreated customers, the estimated probability of switching from segment 6 to 2 is 0.05. Since 19 of the 28 diagonal elements of the matrices shown in Table 6 are at least equal to 0.9, customers remain in the same segment over consecutive years. Many values in the off-diagonal elements of these matrices are below 0.01, indicating that households tend to develop their financial product portfolios and the corresponding assets over their entire lifecycle, implying that changes in financial product portfolios occur only over longer time-periods (Paas et al. 2007) in agreement with the theory proposed in Browning and Lusardi (1996) and Wärneryd (1999). Mails with high intensity sent to customers in segment 2 (“checking account only”) result in relatively high switching probabilities towards segments in which customers have higher probabilities for owning multiple products: 5.6% for switching into state 5 (“phone customers”), 2% for switching into segment 6 (“loan customers”), and 1.5% for switching into segment 7 (“actives”), and these percentages are higher than those corresponding to customers treated less intensively. High direct mail intensity does not enhance switching probabilities from segment 2 (“checking account only”) into segment 4 (“investors”), which includes mortgages. Perhaps other marketing communication channels are more effective when approaching savings accounts, mortgages, or investment trust prospects, considering, for example, that a mortgage is required when buying a house, which is the most important financial decision households will make during their lifecycle and based on an essential need.

The other transition probabilities displayed in Table 6 show that an intensive treatment reduces the probability that customers in segments 6 and 7 (“loan” and “active” customers) switch into segments 1 (“churned segment”) and 2 (“checking account only”). Furthermore, we notice that customers in segment 2 (“checking account only”) are more likely to switch to segments 5 (“phone customers”), 6 (“loan customers”), and 7 (“active customers”) when they receive at least one mail a year.

Results of the analyses suggest that managers may assess, through experiments, whether other marketing instruments than direct mailings (such as outbound call centers and personal sales channels) can be employed to enhance switches into segments characterized by a high number of products owned by the customers. Other salient managerial implications are: (i) ensure each customer receives at least one direct mailing every year to reduce churn; (ii) mail customers in segment defined as that of “loan customers” at least six times yearly to reduce their probability of terminating the usage of the loan at the bank; (iii) send at least six direct mailings each year to customers in segment defined as that of “checking account only” to enhance their probability to switch into more active segments, emphasizing loans, online phone service as well as credit cards since the acquisition of these financial products is mainly influenced by the direct mail channel.

4.2.1 Comparison with other hidden Markov model formulations

In the following, we compare the proposal by showing some additional results obtained with alternative model formulations. First, an HM model is estimated with static weights as in Eq. (6); second, a HM model is estimated without weighting individuals by inverse of estimated probability of treatment assignment including only the time-varying treatment as covariate as in Eqs. (9) and (10) described in Sect. 2.4; and third, the HM model is estimated without weights including both the time-varying treatment and covariates on the initial and transition probabilities with parameterizations in (11) and (12), as described in Sect. 3.

First, the causal HM model with static weights is estimated by using only the pre-treatment covariates observed at the first time occasion for the multinomial logit model as in Eq. (6). This model compared to that proposed in the previous section provides a different customer segmentation with respect to that reported in Table 4. Furthermore, the estimated ATEs at the beginning of the period for the highest treatment intensity (at least 6) are smaller for segments 5, 6, and 7 than those shown in Table 11. With reference to the estimated transition probabilities, they are more persistent for the treatment level "none" and less persistent for the other treatment levels. Full results are not reported, and they are available from the authors upon request.

Second, the HM model is estimated without weights including only the time-varying treatment as covariate influencing both initial and transition probabilities through the parameterization in (9) and (10) described in Sect. 2.4. The model is fitted with seven latent states to compare the results with those obtained with the proposed approach described in Sect. 2. Segments from 4 to 7 are different from those shown in Table 4. Segment 4 is that of phone service customers, segments 5 and 6 across Table 4 differ for loans and investment products. Segment 7 of active clients includes mortgages that are not in segment 7 of Table 4. These differences may be due to the missing adjustment for confounders in treatment assignment as illustrated through the simulation study in Sect. 3. Table 13 in "Appendix A.3" reports the estimated average transition probabilities for the unweighted model according to the number of received mails. The estimated percentages of customers in each of the seven segments at the beginning of the period are the following: 0%, 49%, 20%, 6%, 13%, and 4%. The percentage of "savers" (segment 3) is much higher than that obtained with the proposed model (20% vs. 5%).

Third, the HM model is estimated without weights including the time-varying treatment and covariates on the initial and transition probabilities with parameterizations in (11) and (12) as described in Sect. 3. In this case, the initial endogenous products of each customer are considered as exogenous. Table 14 in "Appendix A.3" reports the estimated average transition probabilities. As shown with the simulation study presented in Sect. 3, this model may lead to biased estimates.

5 Concluding remarks

We propose a formulation of the hidden Markov (HM) model to assess the causal effects of a dynamic treatment in a longitudinal observational study given observed confounders under suitable assumptions. The model conceives potential versions of discrete latent variables representing the features of interest. Treatment effects are estimated on these variables first considering the probability of treatment assignment estimated through multinomial logit models at each time occasion. Weights are combined to obtain an overall weight for each individual using inverse probability weighting, and estimation of the model parameters is carried out through a weighted maximum log-likelihood approach.

We show via simulation experiments that the model outperforms in terms of bias and standard deviation other competitive model formulations. We illustrate the proposed model with data related to a marketing campaign conducted repeatedly over time by a bank. Concerning the results of this illustrative application, we infer a positive effects of the direct mails of the bank on segments of customers' financial product portfolios leading to different managerial insights.

Among the main aspects of the proposal, we mention that the causal HM model may account for multivariate responses, which are binary in the application at hand; however, the model may be formulated for responses of each kind. It allows us to account for unobserved heterogeneity, assuming that the population is composed of a finite number of homogenous subpopulations not directly observable and identified by the states of the latent variables. Thus, treatment effects are estimated across subgroups of the population at the beginning of the observed period and on the following time occasions. In the application, the transition matrix is time homogenous; however, it can be time heterogeneous according to the observed data variability. This modelling approach can be easily applied since the required computational tools are available in statistical packages of the open source software R.

The proposal holds promising potential for addressing causal research questions, especially in marketing contexts where selection may also result from managerial decision-making, as in prospect selection for direct mailings or other channels, targeting customers for retention campaigns. In this case, the manager will generally rely on observed variables to realize a selection. Limitations of the approach are related to the unobserved endogeneity, which may occur frequently. For instance, those clients acquiring a loyalty card may have been more loyal; medical experts who have asked for information about a specific brand may prescribe this one more often because of their pre-existing loyalty toward the brand instead of the information provided. In this case, the assumption of unmeasured confounders cannot be validated, and there can be residual confounding by unmeasured variables, which needs to be accounted for. Another assumption that may be restrictive is that of the common support since when confounding is high, the estimated time-varying weights may be more variable, and sensitivity analyses is necessary to assess the tenability of the required assumptions.

Acknowledgements F. Pennoni acknowledges the financial support from the grant "Dynamic clustering for time-evolving phenomena: Hidden Markov, finite mixture models, and related machine learning methods" of the University of Milano-Bicocca (2021-ATE-0539).

Funding Open access funding provided by Università degli Studi di Milano - Bicocca within the CRUI-CARE Agreement.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Appendix A

The following tables mainly concern some additional features of the simulation study presented in Sect. 3, and of the application presented in Sect. 4.

Appendix A.1. Overview of related literature

Table 7 lists some of the main proposals that appeared in the marketing literature for estimating causal effects.

Appendix A.2. Results of the simulation studies

Tables 8 and 9 report the results of the study performed using $B = 1000$ samples generated as illustrated above under the HM models M_1 and M_2 described in Sect. 3: (i) unweighted; (ii) unweighted with covariate; and (iii) weighted.

Table 7 Summary of selected related literature with applications of hidden Markov models in marketing research

Paper	Context	Response	States	Marketing activity	Endogeneity correction
Moon et al. (2007)	Pharmaceutical Industry	Physician's prescription behavior	Promotional states of competing firms	Pharmaceutical marketing activities (detailing and sampling)	None
Netzer et al. (2008)	Universities and their alumni	Alumni donations	Relationship states	Alumni-university interactions (reunions)	None
Montoya et al. (2010)	Pharmaceutical Industry	Physicians' drug prescription behavior	Frequency of prescription	Pharmaceutical marketing activities (detailing and sampling)	Targeting process equations (Manchanda et al. 2004)
Kumar et al. (2011)	High-tech B2B context	Customer purchase volumes of high-tech products	Relationship states	Marketing budget (amount)	Trivariate Tobit Hidden Markov model
Li et al. (2011)	Financial services (banking)	Ownership of financial products	States of financial maturity	Received postal mails and e-mails	Targeting process equations (Manchanda et al. 2004)
Schweidel et al. (2011)	Telecom	Telecom. services	Relationship states based on customer product portfolios	Introductory promotion	None
Luo and Kumar (2013)	High-tech B2B context	Customer purchasing behavior (amounts)	Buyer-seller relationship states	B2B marketing contacts	Targeting process equations (Manchanda et al. 2004)
Mark et al. (2013)	Retail marketing activities	Number of orders, Total order value	Customer values based on purchase amounts	Retail promotions	None
Schweidel and Knox (2013)	Non-profit sector	Donation behavior	Long-term donation propensity	Number of direct marketing activities	Gaussian copulas (Park and Gupta 2012)
Zhang et al. (2014)	B2B relationships	Buying indicators, (quantity requested, quote accepted)	Behavior relationship states of trust	Product pricing of the seller	Bayesian control function approach (Park and Gupta 2012; Petrin and Train 2010)

Table 8 Simulation results in terms of parameter estimates with data generated for $B = 1000$ samples under model M_1

n	T	Model	β_{21}	β_{22}	β_{31}	β_{32}	δ_{12}	δ_{13}	δ_{22}	δ_{23}
5000	4	Rand: mean	0.431	1.653	1.659	3.306	0.496	1.636	1.635	3.265
		Rand: sd	0.146	0.193	0.152	0.188	0.143	0.172	0.108	0.136
		Unweighted: mean	0.813	2.420	2.426	4.846	0.872	2.356	2.354	4.708
		Unweighted: bias	0.382	0.767	0.768	1.540	0.376	0.721	0.719	1.443
		Unweighted: sd	0.143	0.246	0.204	0.267	0.151	0.186	0.115	0.155
		Covariate: mean	0.605	2.007	2.020	4.026	0.606	2.010	2.014	4.023
		Covariate: bias	0.174	0.354	0.361	0.720	0.110	0.375	0.378	0.758
		Covariate: sd	0.141	0.236	0.201	0.259	0.147	0.193	0.126	0.167
		Weighted: mean	0.499	1.814	1.817	3.625	0.577	1.764	1.765	3.524
		Weighted: bias	0.068	0.162	0.158	0.319	0.082	0.128	0.129	0.259
		Weighted: sd	0.216	0.338	0.265	0.347	0.224	0.268	0.160	0.209
5000	8	Rand: mean	0.424	1.646	1.645	3.291	0.487	1.626	1.629	3.260
		Rand: sd	0.141	0.184	0.146	0.180	0.091	0.104	0.071	0.090
		Unweighted: mean	0.811	2.422	2.426	4.849	0.867	2.354	2.354	4.699
		Unweighted: bias	0.388	0.776	0.781	1.558	0.380	0.727	0.725	1.438
		Unweighted: sd	0.138	0.247	0.205	0.260	0.092	0.124	0.079	0.104
		Covariate: mean	0.602	2.014	2.019	4.035	0.604	2.006	2.008	4.005
		Covariate: bias	0.178	0.368	0.374	0.745	0.117	0.380	0.379	0.744
		Covariate: sd	0.137	0.236	0.207	0.256	0.089	0.125	0.085	0.111
		Weighted: mean	0.525	1.892	1.889	3.785	0.611	1.831	1.827	3.645
		Weighted: bias	0.101	0.246	0.244	0.495	0.124	0.205	0.198	0.385
		Weighted: sd	0.239	0.374	0.318	0.387	0.158	0.200	0.125	0.157
10,000	4	Rand: mean	0.420	1.643	1.644	3.293	0.491	1.627	1.631	3.258
		Rand: sd	0.102	0.135	0.104	0.128	0.105	0.124	0.073	0.094
		Unweighted: mean	0.807	2.417	2.411	4.834	0.863	2.353	2.348	4.701
		Unweighted: bias	0.387	0.774	0.767	1.541	0.372	0.726	0.718	1.443
		Unweighted: sd	0.104	0.173	0.141	0.184	0.105	0.141	0.081	0.111
		Covariate: mean	0.596	2.006	2.003	4.017	0.595	2.005	2.004	4.010
		Covariate: bias	0.176	0.362	0.359	0.725	0.104	0.378	0.373	0.753
		Covariate: sd	0.098	0.163	0.141	0.176	0.102	0.146	0.088	0.118
		Weighted: mean	0.495	1.801	1.791	3.595	0.565	1.764	1.756	3.518
		Weighted: bias	0.075	0.157	0.147	0.302	0.074	0.136	0.125	0.261
		Weighted: sd	0.150	0.227	0.185	0.232	0.147	0.183	0.109	0.148

Table 8 continued

n	T	Model	β_{21}	β_{22}	β_{31}	β_{32}	δ_{12}	δ_{13}	δ_{22}	δ_{23}
10,000	8	Rand: mean	0.427	1.651	1.645	3.293	0.490	1.630	1.630	3.258
		Rand: sd	0.096	0.133	0.103	0.124	0.063	0.071	0.048	0.059
		Unweighted: mean	0.811	2.412	2.408	4.820	0.864	2.346	2.347	4.695
		Unweighted: bias	0.384	0.761	0.763	1.527	0.375	0.716	0.717	1.437
		Unweighted: sd	0.098	0.168	0.140	0.183	0.067	0.083	0.052	0.071
		Covariate: mean	0.601	2.007	2.002	4.008	0.600	1.999	2.002	4.003
		Covariate: bias	0.175	0.355	0.357	0.715	0.110	0.369	0.372	0.745
		Covariate: sd	0.097	0.163	0.140	0.183	0.064	0.086	0.056	0.073
		Weighted: mean	0.534	1.865	1.862	3.725	0.598	1.817	1.816	3.635
		Weighted: bias	0.108	0.213	0.216	0.431	0.109	0.187	0.186	0.377
		Weighted: sd	0.173	0.259	0.214	0.270	0.114	0.137	0.083	0.111

Rand: treatment assigned randomly; unweighted: the unweighted HM model; covariate: the unweighted HM model with covariates; weighted: the proposed causal HM model. Top panel: n = 5000, bottom panel n = 10,000. Average values (mean), average of the bias (bias) and standard deviation (sd)

Appendix A.3. Additional results of the application

In A5, we show the descriptive statistics for the covariates collected in 2000; in Table 11, we show the average treatment effects (ATEs) on the initial segment membership probabilities as in Eq. (9); and in Table 12, we show the estimated ATEs on the transition probabilities as in Eq. (10).

In the following, we also show some results of two different HM models. The first is formulated including as covariate only the time-varying treatment without weighting individuals by the inverse of estimated probability of treatment assignment; see Table 13. The second includes, along with the time-varying treatment, also the other available covariates without weighting individuals; see Table 14. The complete results are available from the authors upon request.

Table 9 Simulation results in terms of parameter estimates with data generated for $B = 1000$ samples under model M_2

n	T	Model	β_{21}	β_{22}	β_{31}	β_{32}	δ_{12}	δ_{13}	δ_{22}	δ_{23}
5000	4	Rand: mean	1.268	2.479	2.478	4.963	1.842	3.513	3.515	7.023
		Rand: sd	0.123	0.244	0.226	0.294	0.097	0.153	0.154	0.205
		Unweighted: mean	1.927	4.019	4.012	8.036	2.860	5.395	5.398	10.795
		Unweighted: bias	0.659	1.540	1.533	3.073	1.018	1.882	1.882	3.773
		Unweighted: sd	0.122	0.722	0.789	1.032	0.121	0.213	0.219	0.296
		Covariate: mean	1.791	3.736	3.722	7.467	2.681	5.269	5.266	10.542
		Covariate: bias	0.522	1.257	1.243	2.504	0.839	1.756	1.751	3.519
		Covariate: sd	0.133	0.541	0.472	0.705	0.130	0.227	0.227	0.310
		Weighted: mean	1.369	2.851	2.820	5.677	2.039	3.877	3.874	7.754
		Weighted: bias	0.101	0.372	0.342	0.714	0.197	0.364	0.358	0.731
		Weighted: sd	0.180	0.651	0.745	0.944	0.146	0.256	0.253	0.345
5000	8	Rand: mean	1.269	2.493	2.478	4.967	1.836	3.500	3.498	6.998
		Rand: sd	0.120	0.230	0.218	0.278	0.062	0.096	0.098	0.126
		Unweighted: mean	1.926	4.036	4.041	8.086	2.846	5.372	5.364	10.737
		Unweighted: bias	0.657	1.543	1.562	3.119	1.010	1.872	1.866	3.739
		Unweighted: sd	0.119	0.792	0.746	1.098	0.074	0.131	0.131	0.179
		Covariate: mean	1.793	3.728	3.742	7.479	2.665	5.228	5.223	10.454
		Covariate: bias	0.524	1.235	1.264	2.512	0.829	1.729	1.725	3.456
		Covariate: sd	0.129	0.474	0.581	0.755	0.076	0.136	0.144	0.194
		Weighted: mean	1.427	3.029	3.032	6.070	2.114	4.015	4.003	8.022
		Weighted: bias	0.157	0.537	0.553	1.103	0.278	0.516	0.504	1.024
		Weighted: sd	0.212	1.090	1.041	1.465	0.117	0.193	0.196	0.268
10,000	4	Rand: mean	1.268	2.460	2.471	4.936	1.837	3.503	3.498	7.002
		Rand: sd	0.087	0.166	0.150	0.193	0.068	0.103	0.109	0.134
		Unweighted: mean	1.933	3.932	3.935	7.869	2.857	5.383	5.379	10.761
		Unweighted: bias	0.666	1.471	1.463	2.933	1.020	1.880	1.882	3.759
		Unweighted: sd	0.086	0.337	0.350	0.465	0.082	0.149	0.152	0.205
		Covariate: mean	1.795	3.689	3.684	7.379	2.676	5.253	5.248	10.502
		Covariate: bias	0.527	1.229	1.213	2.443	0.839	1.749	1.751	3.499
		Covariate: sd	0.093	0.268	0.280	0.364	0.085	0.155	0.157	0.220
		Weighted: mean	1.378	2.762	2.782	5.543	2.035	3.863	3.860	7.720
		Weighted: bias	0.110	0.302	0.311	0.607	0.198	0.359	0.363	0.718
		Weighted: sd	0.123	0.336	0.322	0.418	0.105	0.180	0.178	0.246

Table 9 continued

n	T	Model	β_{21}	β_{22}	β_{31}	β_{32}	δ_{12}	δ_{13}	δ_{22}	δ_{23}
10,000	8	Rand: mean	1.268	2.460	2.462	4.925	1.835	3.495	3.498	6.993
		Rand: sd	0.083	0.163	0.148	0.188	0.043	0.069	0.069	0.094
		Unweighted: mean	1.932	3.956	3.936	7.891	2.847	5.363	5.366	10.730
		Unweighted: bias	0.664	1.496	1.474	2.966	1.012	1.868	1.868	3.737
		Unweighted: sd	0.088	0.340	0.309	0.429	0.052	0.097	0.095	0.132
		Covariate: mean	1.793	3.699	3.684	7.380	2.664	5.220	5.223	10.444
		Covariate: bias	0.525	1.238	1.222	2.455	0.828	1.725	1.725	3.451
		Covariate: sd	0.096	0.275	0.260	0.361	0.055	0.098	0.099	0.138
		Weighted: mean	1.434	2.913	2.876	5.790	2.114	4.001	4.001	7.999
		Weighted: bias	0.165	0.453	0.414	0.865	0.279	0.506	0.503	1.006
		Weighted: sd	0.150	0.453	0.360	0.519	0.081	0.140	0.134	0.186

Rand: treatment assigned randomly; unweighted: the standard HM model covariate: the unweighted HM model with covariates; weighted: the proposed causal HM model. Top panel: n = 5000, bottom panel n = 10,000. Average values (mean), average of the bias (bias) and standard deviation (sd)

Table 10 Descriptive statistics of the covariates referred to year 2000 according to the direct mail intensity in 2001: mean and standard deviation for the continuous variables, frequencies for the categorical variables

Mail intensity in 2001	0	1–2	3–5	≥ 6
Transactions mean	150.924	182.393	263.563	349.454
(s.d.)	(162.763)	(161.053)	(189.629)	(194.380)
Profits mean	53.634	44.257	91.643	228.740
(s.d.)	(354.684)	(349.983)	(501.505)	(679.966)
Age mean	52.793	42.182	41.663	46.400
(s.d.)	(20.396)	(19.775)	(15.965)	(13.626)
Money transferred: none	0.463	0.381	0.248	0.116
≤ 965	0.151	0.169	0.133	0.067
966–1188	0.187	0.218	0.184	0.143
1189–1883	0.107	0.137	0.223	0.265
≥ 1883	0.092	0.094	0.212	0.409
Loans	0.028	0.037	0.089	0.151
Credit cards	0.081	0.114	0.236	0.395
Checking accounts	1.000	1.000	1.000	1.000
Investment products	0.033	0.043	0.106	0.278
Mortgages	0.009	0.019	0.046	0.116
Savings accounts	0.256	0.309	0.419	0.575
Online phone service	0.084	0.100	0.188	0.300

Profits and money transferred are in Euro

Table 11 Estimates of the logit regression parameters for the initial probabilities under the causal HM model with k=7 latent states ([*]significant at 5%, [**]significant at 1%, the second state is taken as reference)

Mail intensity	Latent state (h)					
	1	3	4	5	6	7
None	−16.041**	−0.859**	−2.852**	−1.543**	−2.552**	−1.184**
1–2 versus none	−1.322**	0.160**	0.548**	0.058**	0.202**	0.055**
3–5 versus none	−1.122**	0.437**	0.879**	0.289**	0.548**	0.235**
$\geq$ 6 versus none	−1.024**	1.378**	2.304**	1.239**	1.643**	1.397**
3–5 versus 1–2	0.201*	0.277	0.331	0.232	0.346	0.180
$\geq$ 6 versus 1–2	0.302*	1.187**	1.781**	1.157**	1.452**	1.281**
$\geq$ 6 versus 3–5	0.101	0.934**	1.441**	0.936**	1.071**	1.122**

Table 12 Estimates of the logit regression parameters for the transition probabilities of the latent process under the causal HM model with k=7 latent states ([†]significant at 10%, [*]significant at 5%, [**]significant at 1%), the second latent state is considered as reference

Mail intensity	Latent state (h)					
	1	3	4	5	6	7
None $\bar{h}$=1	−18.501**	−0.023**	−0.954**	−0.691**	−1.052**	−0.581**
None $\bar{h}$=2	−1.821**	−3.554**	−21.618**	−4.195**	−5.552**	−5.261**
None $\bar{h}$=3	0.013**	3.132**	−17.991**	−1.209**	−4.157**	−2.082**
None $\bar{h}$=4	15.433**	−0.484**	17.631**	12.052**	−1.310**	11.759**
None $\bar{h}$=5	1.442**	−2.421**	−15.813**	3.865**	−15.477**	−0.618**
None $\bar{h}$=6	−0.718**	−1.446**	−16.582**	−2.831**	2.372**	−3.447**
None $\bar{h}$=7	0.982**	−0.491**	−3.195**	−1.499**	−16.286**	3.843**
1–2 versus none	−1.262**	−0.151**	1.116**	0.447**	0.153	−0.041
3–5 versus none	−0.832**	−0.011†	1.118**	1.015**	0.681†	0.702
$\geq$ 6 versus none	−0.602**	0.251*	1.941**	1.527**	1.845*	1.283
3–5 versus 1–2	0.443**	0.142**	0.000**	0.576**	0.539**	0.741**
$\geq$ 6 versus 1–2	0.682**	0.407**	−0.179**	1.073**	1.680**	1.321**
$\geq$ 6 versus 3–5	0.241**	0.263**	−0.162**	0.511**	1.160**	0.581**

Table 13 Estimated averaged transition probabilities according to increasing marketing stimuli under the HM model with $k = 7$ latent states without weighting individuals by the inverse of estimated probability of treatment assignment

Mail intensity	Latent state ($\bar{h}$)	Latent state (h)						
		1	2	3	4	5	6	7
none	1	1.000	0.000	0.000	0.000	0.000	0.000	0.000
	2	0.060	0.894	0.028	0.013	0.004	0.001	0.001
	3	0.027	0.054	0.894	0.018	0.006	0.001	0.001
	4	0.056	0.031	0.007	0.878	0.020	0.005	0.003
	5	0.038	0.046	0.014	0.010	0.880	0.007	0.004
	6	0.022	0.024	0.021	0.009	0.008	0.912	0.003
	7	0.017	0.047	0.015	0.006	0.007	0.005	0.903
1–2	1	1.000	0.000	0.000	0.000	0.000	0.000	0.000
	2	0.031	0.900	0.035	0.026	0.007	0.001	0.001
	3	0.011	0.044	0.906	0.029	0.007	0.001	0.001
	4	0.016	0.017	0.005	0.939	0.018	0.003	0.003
	5	0.013	0.030	0.011	0.013	0.925	0.005	0.003
	6	0.011	0.024	0.026	0.018	0.013	0.903	0.004
	7	0.006	0.034	0.013	0.008	0.008	0.004	0.926
3–5	1	1.000	0.000	0.000	0.000	0.000	0.000	0.000
	2	0.038	0.855	0.037	0.049	0.017	0.002	0.002
	3	0.012	0.038	0.879	0.049	0.016	0.003	0.003
	4	0.010	0.009	0.003	0.948	0.023	0.003	0.003
	5	0.007	0.012	0.005	0.011	0.957	0.005	0.004
	6	0.006	0.010	0.012	0.015	0.014	0.939	0.005
	7	0.003	0.014	0.006	0.007	0.008	0.004	0.958
≥6	1	1.000	0.000	0.000	0.000	0.000	0.000	0.000
	2	0.034	0.812	0.048	0.068	0.032	0.003	0.003
	3	0.009	0.028	0.880	0.053	0.024	0.004	0.003
	4	0.007	0.006	0.003	0.945	0.032	0.005	0.003
	5	0.003	0.006	0.004	0.008	0.972	0.005	0.003
	6	0.003	0.005	0.008	0.011	0.014	0.957	0.003
	7	0.002	0.010	0.006	0.007	0.011	0.006	0.959

Table 14 Estimated averaged transition probabilities according to increasing marketing stimuli under the HM model with $k = 7$ latent states only with time-varying treatment and the other covariates without weighting individuals by the inverse of estimated probability of treatment assignment

Mail intensity	Latent state ($\bar{h}$)	Latent state (h)						
		1	2	3	4	5	6	7
None	1	0.850	0.067	0.021	0.020	0.016	0.015	0.010
	2	0.101	0.803	0.025	0.024	0.018	0.018	0.012
	3	0.200	0.161	0.493	0.048	0.038	0.036	0.023
	4	0.205	0.162	0.050	0.483	0.039	0.037	0.023
	5	0.224	0.173	0.054	0.052	0.431	0.040	0.026
	6	0.230	0.179	0.055	0.054	0.045	0.411	0.026
	7	0.264	0.209	0.064	0.062	0.051	0.047	0.302
1–2	1	0.448	0.175	0.144	0.049	0.070	0.073	0.040
	2	0.021	0.807	0.067	0.023	0.030	0.034	0.019
	3	0.023	0.093	0.761	0.026	0.036	0.039	0.021
	4	0.043	0.168	0.139	0.473	0.067	0.071	0.039
	5	0.037	0.141	0.117	0.040	0.573	0.060	0.033
	6	0.036	0.138	0.114	0.039	0.057	0.585	0.032
	7	0.047	0.182	0.149	0.051	0.073	0.076	0.421
3–5	1	0.218	0.155	0.160	0.177	0.141	0.088	0.061
	2	0.010	0.707	0.074	0.080	0.061	0.040	0.028
	3	0.010	0.069	0.714	0.079	0.062	0.039	0.027
	4	0.009	0.065	0.067	0.738	0.059	0.037	0.025
	5	0.011	0.073	0.076	0.085	0.683	0.043	0.029
	6	0.014	0.096	0.099	0.111	0.090	0.552	0.038
	7	0.016	0.114	0.118	0.131	0.105	0.065	0.450
≥ 6	1	0.158	0.220	0.181	0.199	0.132	0.067	0.042
	2	0.006	0.777	0.065	0.070	0.044	0.024	0.015
	3	0.006	0.089	0.729	0.080	0.052	0.027	0.017
	4	0.006	0.083	0.069	0.751	0.050	0.025	0.016
	5	0.008	0.105	0.088	0.097	0.648	0.033	0.021
	6	0.011	0.148	0.123	0.136	0.092	0.461	0.029
	7	0.013	0.177	0.146	0.160	0.108	0.054	0.342

References

Angrist J, Imbens G, Rubin DB (1996) Identification of causal effect using instrumental variables. J Am Stat Assoc 91:444–472

Bacci S, Pandolfi S, Pennoni F (2014) A comparison of some criteria for states selection in the latent Markov model for longitudinal data. Adv Data Anal Classif 8:125–145

Bartolucci F, Farcomeni A, Pennoni F (2013) Latent Markov models for longitudinal data. Chapman and Hall, Boca Raton

Bartolucci F, Farcomeni A, Pennoni F (2014) Latent Markov models: a review of a general framework for the analysis of longitudinal data with covariates (with discussion). TEST 23:433–465

Bartolucci F, Pennoni F, Vittadini G (2016) Causal latent Markov model for the comparison of multiple treatments in observational longitudinal studies. J Educ Behav Stat 41:146–179

Bartolucci F, Pandolfi S, Pennoni F (2017) `LMest`: an `R` package for latent Markov models for longitudinal categorical data. J Stat Softw 81:1–38

Baum LE, Petrie T, Soules G, Weiss N (1970) A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. Ann Math Stat 41:164–171

Browning M, Lusardi A (1996) Household savings: micro theories and micro facts. J Econ Lit 34:1797–1855

Chang CW, Zhang JZ (2016) The effects of channel experiences and direct marketing on customer retention in multichannel settings. J Interact Mark 36:77–90

Conley TG, Hansen CB, Rossi PE (2012) Plausibly exogenous. Rev Econ Stat 94:260–272

Davison AC, Hinkley DV (1997) Bootstrap methods and their application. Cambridge University Press, Cambridge

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc B 39:1–38

Ehrenberg AS (1965) An appraisal of Markov brand-switching models. J Mark Res 2:347–362

Goodman LA (1974) Exploratory latent structure analysis using both identifiable and unidentifiable models. Biometrika 61:215–231

Holland PW (1986) Statistics and causal inference. J Am Stat Assoc 81:945–960

Joffe MM, Ten Have TH, Feldman HI, Kimmel SE (2004) Model selection, confounder control, and marginal structural models: review and new applications. Am Stat 58:272–279

Kappe E, Blank AS, DeSarbo WS (2018) A random coefficients mixture hidden Markov model for marketing research. Int J Res Mark 35:415–431

Knott A, Hayes A, Neslin SA (2002) Next product-to-buy models for crosselling applications. J Interact Mark 16:59–75

Kumar V, Sriram S, Luo A, Chintagunta PK (2011) Assessing the effect of marketing investments in a business marketing context. Mark Sci 30:924–940

Lanza ST, Coffman DL, Xu S (2013) Causal inference in latent class analysis. Struct Equ Model 20:361–383

Lazarsfeld PF (1950) The logical and mathematical foundation of latent structure analysis. In: Stouffer SA, Guttman EASL (eds) Measurement and prediction. Princeton University Press, New York

Lemmens A, Croux C, Stremersch S (2012) Dynamics in the international market segmentation of new product growth. Int J Res Mark 29:81–92

Li S, Sun B, Montgomery AL (2011) Cross-selling the right product to the right customer at the right time. J Mark Res 48:683–700

Luo A, Kumar V (2013) Recovering hidden buyer–seller relationship states to measure the return on marketing investment in business-to-business markets. J Mark Res 50:143–160

Manchanda P, Rossi PE, Chintagunta PK (2004) Response modeling with nonrandom marketing-mix variables. J Mark Res 41:467–478

Mark T, Lemon KN, Vandenbosch M, Bulla J, Maruotti A (2013) Capturing the evolution of customer-firm relationships: how customers become more (or less) valuable over time. J Retail 89:231–245

McCaffrey DF, Griffin BA, Almirall D, Slaughter ME, Ramchand R, Burgette LF (2013) A tutorial on propensity score estimation for multiple treatments using generalized boosted models. Stat Med 32:3388–3414

Montoya R, Netzer O, Jedidi K (2010) Dynamic allocation of pharmaceutical detailing and sampling for long-term profitability. Mark Sci 29:909–924

Moon S, Kamakura WA, Ledolter J (2007) Estimating promotion response when competitive promotions are unobservable. J Mark Res 44:503–515

Mor B, Garhwal S, Kumar A (2021) A systematic review of hidden Markov models and their applications. Arch Comput Methods Eng 28:1429–1448

Netzer O, Lattin JM, Srinivasan V (2008) A hidden Markov model of customer relationship dynamics. Mark Sci 27:185–204

Neyman J (1923) On the application of probability theory to agricultural experiments. Transl Stat Sci 5:465–480

Paas LJ, Vermunt JK, Bijmolt THA (2007) Discrete time, discrete state latent Markov modelling for assessing and predicting household acquisitions of financial products. J R Stat Soc Ser A 170:955–974

Papies D, Ebbes P, Van Heerde HJ (2017) Addressing endogeneity in marketing models. In: Leeflang P, Wieringa JE, Bijmolt THA, Pauwels KH (eds) Advanced techniques and methods to model markets. Springer, Cham

Park S, Gupta S (2012) Handling endogenous regressors by joint estimation using copulas. Mark Sci 31:567–586
Park CH, Park Y-H, Schweidel DA (2018) The effects of mobile promotions on customer purchase dynamics. Int J Res Mark 35:453–470
Petrin A, Train K (2010) A control function approach to endogeneity in consumer choice models. J Mark Res 47:3–13
Poulsen CA (1982) Latent structure analysis with choice modeling applications. Ph.D. Thesis, Aarhus School of Business Administration and Economics, Aarhus
Robins JM (1997) Causal inference from complex longitudinal data. In: Berkane M (eds) Latent variable modeling and applications to causality. Lecture notes in statistics, vol 120. Springer, New York, pp 69–117
Robins J, Rotnitzky A (1995) Semiparametric efficiency in multivariate regression models with missing data. J Am Stat Assoc 90:122–129
Robins JM, Hernán MA, Brumback B (2000) Marginal structural models and causal inference in epidemiology. Epidemiology 11:550–560
Rosenbaum PR (1987) Model-based direct adjustment. J Am Stat Assoc 82:387–394
Rosenbaum PR (2020) Modern algorithms for matching in observational studies. Ann Rev Stat Appl 7:143–176
Rosenbaum PR, Rubin DB (1983) The central role of the propensity score in observational studies for causal effects. Biometrika 70:41–55
Rubin DB (1990) Formal mode of statistical inference for causal effects. J Stat Plan Inference 25:279–292
Rubin DB (2005) Causal inference using potential outcomes: design, modeling, decisions. J Am Stat Assoc 100:322–331
Schwarz G (1978) Estimating the dimension of a model. Ann Stat 6:461–464
Schweidel DA, Knox G (2013) Incorporating direct marketing activity into latent attrition models. Mark Sci 32:471–487
Schweidel DA, Bradlow ET, Fader PS (2011) Portfolio dynamics for customers of a multiservice provider. Manag Sci 57:471–486
Skrondal A, Rabe-Hesketh S (2014) Handling initial conditions and endogenous covariates in dynamic/transition models for binary data with unobserved heterogeneity. J R Stat Soc Ser C 63:211–237
Stuart E (2010) Matching methods for causal inference: a review and a look forward. Stat Sci 25:1–21
Tullio F, Bartolucci F (2022) Causal inference for time-varying treatments in latent Markov models: an application to the effects of remittances on poverty dynamics. Ann Appl Stat 16:1962–1985
Vermunt JK, Paas L (2017) Mixture models. In: Leeflang P, Wieringa JE, Bijmolt THA, Pauwels KH (eds) Advanced techniques and methods to model markets. Science+Business Media, Cham, Switzerland
Visser I, Speekenbrink M (2022) Mixture and hidden Markov models with R. Science+Business Media, Cham, Switzerland
Wärneryd K-E (1999) The psychology of saving: a study of economic psychology. Edward Elgar Publishing, Northampton
Wedel M, Kamakura WA (2012) Market segmentation: conceptual and methodological foundations. Springer, New York
Welch LR (2003) Hidden Markov models and the Baum–Welch algorithm. IEEE Inf Theory Soc Newsl 53:1–13
Wiggins LM (1973) Panel analysis: latent probability models for attitude and behaviour processes. Elsevier, Amsterdam
Zhang JZ, Netzer O, Ansari A (2014) Dynamic targeted pricing in B2B relationships. Mark Sci 33:317–337
Zucchini W, MacDonald IL, Langrock R (2016) Hidden Markov models for time series: an introduction using R, 2nd edn. Chapman and Hall, New York

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.