Causality patterns of a marketing campaign conducted over time: evidence from the latent Markov model

Effetti causali di una campagna di marketing protratta nel tempo attraverso un modello con processo di Markov latente

Fulvia Pennoni, Leo Paas and Francesco Bartolucci

Abstract Many statistical methods currently employed to evaluate the effect of a marketing campaign in dealing with observational data advocate strong parametric assumptions to correct for endogeneity among the participants. In addition, the assumptions compromise the estimated values when applied to data in which the research expects endogeneity but this is not realized. Based on the recent advances in the literature of causal models dealing with data collected across time, we propose a dynamic version of the inverse-probability-of-treatment weighting within the latent Markov model. The proposal, which is based on a weighted maximum likelihood approach, accounts for endogeneity without imposing strong restrictions. The likelihood function is maximized through the Expectation-Maximization algorithm which is suitably modified to account for the inverse probability weights. Standard errors for the parameters estimates are obtained by a nonparametric bootstrap method. We show the effects of multiple mail campaigns conducted by a large European bank with the purpose to influence their customers to the acquisitions of the addressed financial products.

Keywords: causal latent Markov model, customer relationship management, direct marketing, Expectation-Maximization algorithm

---

Fulvia Pennoni
Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Milano, Italy, e-mail: fulvia.pennoni@unimib.it

Leonard J. Paas
School of Communication, Journalism and Marketing, Massey University, Auckland, New Zealand, e-mail: L.J.Paas@massey.ac.nz

Francesco Bartolucci
Department of Economics, University of Perugia, Perugia, Italy, e-mail: francesco.bartolucci@unipg.it

1 Background

Firms operationalizing relationship marketing strategies often require insights into long-term developments of interactions with customers. Marketing response models have been developed to relate the customer decisions to the marketing efforts (Manchanda et al., 2004). For example, Li et al. (2005, 2011) propose to assess the consumer financial product portfolios at a bank, by considering key customer lifetime value indicators, such as customer profitability and retention. Latent Markov (LM) models (Bartolucci et al., 2013) have been extensively applied to incorporate switching of customers between segments over time (see, among others, Paas et al., 2007; Schweidel et al., 2011). They have been employed to potentially identify cross-sell opportunities.

In this context, the portfolios development can be influenced by the marketing activities operated by the firm or the bank that may include, among many strategies, also the direct mail channel. This kind of activity is often strategically based on customer characteristics or behaviors. When they are incorporated as covariates in an LM model, endogeneity can occur, since the bank is using the information on, for example, customer demographic characteristics and product ownership when targeting customers for campaigns. Among the models which have been applied to cope with this feature, the Gaussian copulas (Park and Gupta, 2012) and the latent instrumental variable models (Ebbes et al., 2005) are suitable to cope with endogeneity. However, these alternative models need strong assumptions which are not often satisfied. The first one assumes a Gaussian distribution for the errors and non-Gaussian distribution for the endogenous regressors. The latent instrumental variable models assume a discrete component among the endogenous regressors.

We propose an alternative model which is an extension of that proposed by Bartolucci et al. (2016) to address causal effects within observational studies. Consistently with the potential outcome (PO) framework (Rubin, 2005), the causal effects are considered under a counterfactual scenario because individuals cannot be observed under multiple treatments, but only for the treatment they effectively receive. We apply the inverse-probability-of-treatment weighting (IPW), based on the reciprocal of the probability for an individual of receiving the treatment s/he effectively received to alleviate the potential bias due to endogeneity (see also, Skrondal and Rabe-Hesketh, 2014). This accounts for dynamic counterfactuals and sequential endogeneity since treatments at different time occasions may induce different responses. Therefore, we propose to apply a weighted maximum likelihood approach to obtain a consistent estimator for the causal effect of interest.

The remainder of the paper proceeds as follows. In Section 2 we illustrate the proposed causal LM model. In Section 3 we present the empirical application related to a repeated direct mail campaign of an important bank to enlarge the customers' financial product portfolios. In Section 4 we report the results and in Section 5 we offer some concluding remarks.

2 The proposed causal LM model

We consider an observational longitudinal study involving n customers of a firm or a bank. Let $\boldsymbol{Y}_{it}$ be the observed binary vector of responses for customer i, $i = 1,\dots,n$, at each time occasion t, $t = 1,\dots,T$, which is a vector indicating the product portfolio of the customer owned at time t. Let $\boldsymbol{X}_{it}$ be a column vector of the available time-varying customer's covariates, some of which may have been selected by the bank to establish the marketing campaign.

Under the potential outcome framework we define a discrete variable to denote the time-varying treatment Z_{it} for customer i and time occasion t that is an ordinal variable with levels in $\{1,\dots,l\}$. We denote by $H_{it}^{(z)}$ the discrete latent variable that is customer- and time-specific when s/he has received treatment z. Its distribution has support $\{1,\dots,k\}$ and the support points are discrete finite values as each customer has many potential outcomes which are "potential versions" associated with each marketing intensity z administered at each time occasion. We also allow that the treatment can be absent when the first level of z is set equal to zero. For instance, $H_{it}^{(3)}$ corresponds to the latent state of prospect i at time t if s/he has received the treatment intensity equal to three marketing stimuli up to time t. The time sequence of these variables for customer i is collected in vector $\boldsymbol{H}_i^{(z)} = (H_{i1}^{(z)},\dots,H_{iT}^{(z)})'$.

In the following, we denote by lower case letters the realized values of the variables. Accordingly, the time-specific vectors of responses are denoted by $\boldsymbol{y}_{i1},\dots,\boldsymbol{y}_{iT}$ and we assume that every customer has a positive probability $p(Z_{it} = z|\boldsymbol{x}_{it})$ of receiving a marketing stimulus at each time period. We also assume that the marketing stimuli at each time period are independent of the potential outcomes given the pre-treatment covariates, that is, $Z_{it} \perp\!\!\!\perp H_{it}^{(z)}|\boldsymbol{X}_{i,t-1}$ for $i = 1,\dots,n,\ t = 1,\dots,T$.

The distribution of each $H_{it}^{(z)}$ variable is defined according to a of first-order Markov chain, where the initial probabilities of the latent chain define the potential products portfolios according to the received treatment at the first occasion. These are modeled by the following baseline-category logit model because the latent states do not show a precise order:

$$\log \frac{p(H_{i1}^{(z)} = h)}{p(H_{i1}^{(z)} = 1)} = \alpha_h + \boldsymbol{d}(z)'\boldsymbol{\beta}_h, \quad h = 2,\dots,k. \tag{1}$$

In the previous expression, α_h is the state specific intercept, $\boldsymbol{\beta}_h = (\beta_{h2},\dots,\beta_{hl})'$ is a column vector of $l-1$ parameters, and $\boldsymbol{d}(z)$ is a column vector of $l-1$ zeros with the $(z-1)$-th element equal to 1 if $z > 1$. The element β_{hz} of $\boldsymbol{\beta}_h$ for $z > 1$, corresponds to the effect at the first occasion of the z-th treatment with respect to the effect of the first level of the treatment ($z = 1$). This coefficient is the average effect of the corresponding treatment for the population of interest at the beginning of the campaign.

A similar parameterization, which is not reported here, is employed for the transition probabilities. The parameters referred to the distribution of every observed response variable conditional to the POs identify the products owned by customers at each time occasion and they are useful to cluster the customers into homogenous segments. If required, the latter assumption may be relaxed in different ways (see Bartolucci et al., 2013, for further details).

In order to introduce the IPW estimator, we estimate the following multinomial logit model

$$\log \frac{p(Z_{it} = z|\boldsymbol{x}_{it-1})}{p(Z_{it} = 1|\boldsymbol{x}_{it-1})} = \eta_z + \boldsymbol{x}'_{it-1}\boldsymbol{\lambda}_z, \quad z = 1, \dots, l, \quad t = 1, \dots, T,$$

where η_z and $\boldsymbol{\lambda}_z$ are the intercept and the regression parameters referred to each treatment level. On the basis of the parameter estimates, we compute the customer weights referred to each time period as

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

and the overall individual weights are provided by products of the weights concerning each year of the campaign

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

Given the observed data, model estimation is performed by maximizing the weighted log-likelihood

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \hat{\boldsymbol{w}}_i \log \ell_i(\boldsymbol{\theta}), \quad \ell_i(\boldsymbol{\theta}) = \log p(\boldsymbol{y}_{i1}, \dots, \boldsymbol{y}_{iT}|z_{it}),$$

where $\boldsymbol{\theta}$ is the overall vector of parameters. This log-likelihood is maximized by using a modified version of the Expectation-Maximization (EM) algorithm (Baum et al., 1970; Dempster et al., 1977). The algorithm is based on the complete data log-likelihood $\ell^*(\boldsymbol{\theta})$ which involves the weighted frequencies of each latent states (see also Bartolucci et al., 2013, for details). The steps of the algorithm are the following:

- **E-step**: it computes the conditional expected value of each frequency involved in the complete data log-likelihood. It requires to compute the posterior probabilities of the latent variable given the weights, the observed responses, and the treatment sequence;
- **M-step**: it maximizes the complete-data log-likelihood where the frequencies are obtained by the corresponding expected values calculated at the E-step.

A non-parametric bootstrap algorithm (Davison and Hinkley, 1997) is employed to obtain the standard errors for the model parameters. We resample customers, with their observed pre-treatment covariates and outcomes, a suitable number of times; then we estimate the model for each generated sample in accordance with the estimated customer's time-varying weights. The number of latent states is selected by the Bayesian Information Criterion (BIC; Schwarz, 1978).

3 Application

We analyze a sample of 49,967 customers aged 18 years and older, provided by a large anonymous European bank that conducted a multiple direct mail campaign over a long period of time. We evaluate the effectiveness of the efforts made by the bank in terms of developments of the customers's portfolios. The following products could be affected by the marketing campaign: *loans, credit cards, checking accounts, investment products, mortgages, savings accounts, and a paid phone service enabling customers to gain insights into their account balances*.

Direct mail was the dominant channel for making product offers to customers in the 2000-2003-period to which the data refer. Table 1 shows the proportions of customers involved in each year of the campaign. We consider four treatment levels varying from none to more than six mails sent to each customer. As shown in Table 1, the sequential treatments have been administered according to an increasing intensity to the customers.

Direct mail intensity	2001	2002	2003
none	0.317	0.237	0.186
1-2	0.311	0.323	0.248
3-5	0.222	0.211	0.245
≥ 6	0.150	0.229	0.321

Table 1 *Observed proportions of direct mail intensity by year.*

The available time-varying covariates influencing the treatment probability are the following: customer's age, money s/he has transferred each year on the account, number of transactions made annually, and annual bank' profits on each customer. It is important to note that the endogeneity arises mainly because the customer selection to which address the campaign made by the managers is not done by randomized methods. Instead the bank's managers choose by simple common sense or by employing logistic regression models or tree-based algorithms.

4 Results

The BIC index favors the model with seven latent states when we estimate the causal LM model for a number of latent states ranging from 1 to 8. The model-based approach defines the following customer segments corresponding to the latent states and characterized on the basis of the conditional response probabilities given the latent states:

1. “none” (0%)
2. “checking account only” (40%)
3. “savers’segment” (5%)

4. “investors” (23%)
5. “phone service customers” (11%)
6. “loans customers” (16%)
7. “actives” (5%)

where the reported percentages are referred to the proportion of customers in the corresponding segment after the first year of the campaign. In this way, we also identify a segment (denoted with 1) as that of individuals which have churned the bank during the first period.

In Table 2 we show the estimated initial probabilities which are averaged according to the intensity of the treatment (see Table 1). They are obtained by considering the estimated average treatment effect on the initial segments as in equation (1). From these results we notice that if the campaign is not conducted, or it is conducted with a low mail intensity, there is a higher probability for a customer to be allocated in segment 2 “checking account only” at the beginning of the period respect to a more intensive campaign. Note that, except for segment 2, all the probabilities referred to intensities (3-5) and (≥ 6) are higher than those of lower treatment levels. This indicates that the treatment enhances the probability to be more active at bank, especially to become “investor” (segment 4).

	Latent state (h)						
Direct mail intensity	1	2	3	4	5	6	7
none	0.00	0.48	0.21	0.03	0.10	0.04	0.15
1-2	0.00	0.44	0.22	0.04	0.10	0.04	0.15
3-5	0.00	0.38	0.25	0.05	0.11	0.05	0.15
≥6	0.00	0.18	0.30	0.11	0.13	0.07	0.21

Table 2 *Estimated initial probabilities for each treatment level under the causal LM.*

	Latent state (h)						
$\bar{h}$	1	2	3	4	5	6	7
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

Table 3 *Estimated transition probabilities of the causal LM model with 7 latent states.*

In Table 3 we report the estimated transition probabilities for customers in the highest treatment level (intensity ≥6 mails). The other three tables concerning the transition probabilities of the latent states of lower mails intensity are not reported here. Based on all relevant tables, we notice that when treatment is intensive, the customers in segment 2 “checking account only” show a relative high probability to switch towards the segments in which customers own multiple products, that is, 5.6% for switching into latent state 4 “investment inclined customers”, 2.0% for switching into segment 6 “loan customers”, and 1.5% for switching into segment 7 of “active customers”.

From the other estimated matrices relative to a low mail intensity or no treatment at all, we notice that mainly for the customers in segment 5 “loan customers ” and 6 “mortgage customers” there is a higher probability to switch into segment 1 of “inactive customers” or segment 2 of “checking account only” compared to those observed in Table 3.

5 Conclusions

The findings gained by the results illustrated in Section 4 provide a number of salient managerial implications. We provide a short summary through the following three advices:

- ensure that each customer receives at least one direct mailing each year to reduce churn probabilities;
- perform an intensive campaign towards customers in segment 6 “lenders” to reduce the probability of terminating the usage of the loan at the same bank;
- send at least six direct mails each year to customers in segment 2 “checking account only”, to enhance their probability to switch into the more active states, emphasizing the loan and the online phone service, as the acquisition of these financial products is most strongly influenced by the direct mail channel.

Direct mail mostly affects forward switching probabilities for customers with only a checking account (segment 2) into multiple segments not characterized by high ownership of savings accounts and mortgages. This leads to two additional suggestions: (*i*) assess through experiments whether innovative direct mailings can enhance switching probabilities forward that are currently low. Also, other channels can be employed to assess whether these low transition probabilities can be affected (e.g., the personal sales channel); (*ii*) find how the direct mail channels or other marketing communication channels can be employed to effectively market mortgages.

The innovative approach we propose to address endogeneity has important features which may help managers to make better decisions on how many direct marketing mailings each individual customer should receive. Therefore, the proposed causal LM model may be fruitfully applied in many other potential marketing contexts over that illustrated in the applicative example.

References

Bartolucci, F., Farcomeni, A., and Pennoni, F. (2013). *Latent Markov Models for Longitudinal Data*. Chapman and Hall/CRC press, Boca Raton.

Bartolucci, F., Pennoni, F., and Vittadini, G. (2016). Causal latent Markov model for the comparison of multiple treatments in observational longitudinal studies. *Journal of Educational and Behavioral Statistics*, 41:146–179.

Baum, L., Petrie, T., Soules, G., and Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Annals of Mathematical Statistics*, 41:164–171.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and their Application*. Cambridge University Press, Cambridge, MA.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society - Series B*, 39:1–38.

Ebbes, P., Wedel, M., Böckenholt, U., and Steerneman, T. (2005). Solving and testing for regressor-error (in)dependence when no instrumental variables are available: With new evidence for the effect of education on income. *Quantitative Marketing and Economics*, 3:365–392.

Li, S., Sun, B., and Montgomery, A. L. (2011). Cross-selling the right product to the right customer at the right time. *Journal of Marketing Research*, 48:683–700.

Li, S., Sun, B., and Wilcox, R. T. (2005). Cross-selling sequentially ordered products: An application to consumer banking services. *Journal of Marketing Research*, 42:233–239.

Manchanda, P., Rossi, P. E., and Chintagunta, P. K. (2004). Response modeling with nonrandom marketing-mix variables. *Journal of Marketing Research*, 41:467–478.

Paas, L. J., Vermunt, J. K., and Bijmolt, T. H. (2007). Discrete time, discrete state latent Markov modelling for assessing and predicting household acquisitions of financial products. *Journal of the Royal Statistical Society - Series A*, 170:955–974.

Park, S. and Gupta, S. (2012). Handling endogenous regressors by joint estimation using copulas. *Marketing Science*, 31:567–586.

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

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

Schweidel, D. A., Bradlow, E. T., and Fader, P. S. (2011). Portfolio dynamics for customers of a multiservice provider. *Management Science*, 57:471–486.

Skrondal, A. and Rabe-Hesketh, S. (2014). Handling initial conditions and endogenous covariates in dynamic/transition models for binary data with unobserved heterogeneity. *Journal of the Royal Statistical Society - Series C*, 63:211–237.