Uncovering Path-to-Purchase Segments in Large Consumer Population using Clustered Multivariate Autoregression

Yicheng Song, Nachiketa Sahoo, Shuba Srinivasan, Chris Dellarocas

School of Management, Boston University

Abstract

We propose a novel method to identify predominant paths-to-purchase of retail consumers from activity level dataset collected in CRM systems. We verify the effectiveness of the proposed model on a simulated dataset. Following successful verification, we apply the model on a retail dataset from a major multi-channel, multi-brand North American Retailer. We uncover three different types of consumers based on how they respond to external stimuli over time: catalog driven shoppers, email driven shoppers, and holiday driven online shoppers. We also find significant activity across channels by these consumers. Finally, we use the path information in the segments to identify the groups that are most sensitive to a certain type of marketing contact. By analyzing the response of customers in different groups in a test dataset, we show that managers can optimize marketing budget allocation using our proposed segmentation approach.

1 Introduction

Consumers' path-to-purchase has been a topic of intense interest in recent years (e.g., David et al.2009, Xu et al.2014). It is recognized that during shopping consumers move through a sequence of states, such as, Awareness to Familiarity to Consideration to Purchase. By aligning a firm's marketing efforts with the consumer's journey on path-to-purchase one may move the customers along the path to purchase. Properly targeted marketing exposure can move the consumer forward through a sequence of pre-purchase activities (consideration to information search) to purchase behaviors (online and offline purchase) to potential after-purchase behaviors (reviews and ratings) (Figure 1).

Figure 1. Path-to-Purchase example.

Recent observation from the industry suggests that the consumers' journey in a digital age could be non-linear (David et al. 2009). E.g., an impulse purchaser might go from consideration to purchase quickly without going through a process of active evaluation. On the other hand, a deliberate consumer might return from evaluating one product to considering other products. So, for different consumers one would need to identify their shopping habits and craft appropriate marketing strategies to be effective. Despite the potential utility of the path-to-purchase few

papers in the literature have defined path-to-purchase in an empirically identifiable way, let alone determine them from the customer activity data. We aim to fill this gap in the current paper.

The goal of the current research is to develop an approach to identify the predominant paths-to-purchase taken by the consumers and segment them based on this characteristic. We accomplish this major goal in three steps. First, we propose a model of interaction between different activities of a consumer over time. We argue that such a model captures the consumer' path-to-purchase. Second, we develop a clustering algorithm that simultaneously identifies groups of consumers with similar path-to-purchase and the group specific model that represents those paths. Finally, we extract the predominant path for each cluster as a sequence of activities that the consumers in the cluster do leading to purchase.

The rest of the paper is organized as follows: We briefly discuss related work in Section 2. Then, we describe our application context and the dataset in Section 3. The clustering algorithm, called Clustered Zero Inflated Multivariate Autoregressive Poisson (CZMAP) is presented in Section 4. We validate CZMAP algorithm using simulated data in Section 5. We apply the algorithm to the retail dataset and discuss the findings in Section 6. Finally, we conclude the paper in Section 7.

2 Related Works

Given the paper's focus on segmenting consumers based on their characteristic Path-to-Purchase, we draw upon two research streams: literature on Path-to-Purchase and literature on consumer segmentation.

2.1 Path-to-Purchase

Path-to-Purchase has been widely studied in recent years. It has two notable features:

The first feature of Path-to-Purchase is sequential activities. David et al.(2009) found that there are multiple stages before consumer decide to purchase and Path-to-Purchase is considered as consumer decision journey, which describes consumer decision journey as "Awareness→ Familiarity→Consideration→Purchase→Loyalty". The funnel analogy journey suggests that consumers follow a sequence of activities as they weigh options, make decisions, and buy products. Thus, Path-to-Purchase requires using consumer time series data on their shopping behavior.

The second feature is that there is interaction between activities within the Path-to-Purchase. For example, marketing campaigns may not result in immediate purchases, but they will make consumers more likely to browse the product, followed by search for information about the product, and ultimately lead to purchases. Xu et al. (2014) model consumers' path-to-purchase with a Bayesian hierarchical framework to capture the dynamic interactions among sequential activities of consumers. They found that display advertisements have relatively low direct effect on purchase conversion, but they are more likely to stimulate subsequent visits through other advertisement formats. Li et al. (2013) makes use of a granular record of every touch point consumers have in the online purchase journey. They found that when taking dynamic interactions within the Path-to-Purchase into consideration, the relative contributions of different channels are significantly different from and more accurate than those found by other currently used metrics.

Based on the above analysis, we loosely define ***Path-to-Purchase as a sequence of activities leading to purchase that is characterized by interaction between these activities***. In order to model consumer's Path-to-Purchase, we need to utilize time series data to examine the sequence of activities leading to purchase, while capturing the interactions between these activities.

2.2 Consumer Clustering

Research on segmenting consumers based on their past purchase history can be categorized into two approaches.

The first approach (e.g., Baragona 2011) estimates individual level vector auto-regressive coefficients and groups consumers based on these coefficients using an clustering algorithm such as K-means (Sims 1980). A limitation of such approaches is that when faced with the sparse data with many zeros, it is difficult to get a robust estimation of individual-level vector autogression coefficients, leading to poor clustering solutions.

A second approach (e.g., Pauwels et al. 2011) clusters consumers based on their known characteristics such as demographics, followed by an estimation of a time-series model such as vector autogression for each group. The key assumption behind this approach is that similar demographics lead to the similar consumer purchasing behavior, but often this is not the case. Despite this shortcoming, this paper highlights the interest in segmenting customers and identifying the unique paths to purchase in different segments.

The proposed CZMAP approach is a superior approach over demographics based segmentation since it is based directly on the observed behavior. It also overcomes a key limitation in some of the approaches proposed in the literature by being able to deal with problem of data sparsity that is common in individual level activity data.

3 Data

We collect a dataset from a major multi-channel, multi-brand North American retailer through Wharton Customer Analytics Initiative (WCAI). The dataset includes marketing activity, customers' pre-purchase behavior, customers' purchases, as well as their post-purchase behavior. The dataset we used in this study includes 9,805 active consumers from the largest brand of the retailer. The dataset spans from 7/1/2010 to 6/30/2012. The notable feature of Path-to-Purchase is the sequential activity, necessitating the use of weekly time-series data for each consumer. For each week we construct the following variables: number of email and catalogs received, number of products browsed, number of keyword search performed, number of online and offline purchases. Because the consumer's activities are affected by marketing campaigns, such as catalogs and emails, they are considered endogenous variables while activities such as email campaigns, catalogs sent, holidays and promotions available are treated as exogenous variables. The descriptions the variables are listed in Appendix A.

4 Model Development

We begin with the first step outlined in Section 1, which is the modeling of multivariate time-series of activity data for each consumer.

4.1 Zero-Inflated Multivariable Autoregressive Poisson Model

We need a time-series model that captures the interaction among activities along the path to purchase. For this purpose we extend the Vector Auto-regression (VAR) model. Vector Autoregressive Models are used to model contemporaneous and over time correlation between components of multivariate time series data. However, a standard VAR model is unsuitable for our setting because all our endogenous variables are time-series of count data with abundance of zero values. The standard VAR model is suitable for continuous data where the error term can be assumed to follow a Normal distribution. We therefore need a time-series model that makes appropriate distributional assumptions for counts, accommodates the zeros, and estimates contemporaneous correlations and lagged effects.

We use a Zero Inflated Multivariate Autoregressive Poisson model (ZMAP) to accommodate discreteness, sparseness and both auto and cross-correlation in the count data. The ZMAP model is developed based on (e.g., Stephen and Galak 2012, Heinen and Rengifo 2007). It explicitly accommodates sparse events that generate excess zero counts in time series. Specifically, the ZMAP has three components: (1) Zero-Inflated-Poisson distributions for the marginal distributions of the components to accommodate sparse count data, (2) Autoregressive models for the occurrence of the zero values and expected mean of the Zero-Inflated-Poisson distributions to capture lagged effects of variables, (3) Multivariate Normal copula to connect the marginal distributions and capture the contemporaneous correlation between endogenous time series that not captured in lagged effects. Next, we introduce them one by one:

(1) Zero-Inflated-Poisson (ZIP). For count data, we need a discrete probability distribution over Natural numbers. Although Poisson or Negative Binomial distributions are typically used, when dealing with excessive zeros, a special treatment of these zeros can improve estimates and model fit. Thus, we adopt ZIP (Lambert 1992) to model the sparse count data. Specifically, $X_i^T(j)$ is the j *th* endogenous variable for i *th* consumer at T *th* week. The probability of generating $X_i^T(j)$ can be expressed as

$$f_{ZIP}\left(X_i^T(j)\right) = p + (1-p)Poisson(0|\lambda) \;\; if\; X_i^T(j) = 0$$
$$= (1-p)Poisson(X_i^T(j)|\lambda) \;\; if\; X_i^T(j) > 0 \qquad (1)$$

Where $p \in [0,1]$ is the mixture parameter of the zero-inflated model and λ is the expected mean of the Poisson distribution. The calculation of p and λ are based on the next step.

(2) Autoregression. We model λ (*the expected mean of the Poisson distribution*) and p (*mixture parameter of the zero-inflated model*) using a linear autoregressive model to capture own and cross variable lagged effects. Specifically,

$$Log(\lambda) = BM \; and \; Logit(p) = GM \qquad (2)$$

Where M is the matrix $\{X_i^{T-1}, \dots, X_i^{T-p}, Y_i^T\}$, which consists of p lagged endogenous variables and the exogenous variables. The parameters B and G capture the relationship of endogenous variable X_i^T with its lags $\{X_i^{T-1}, \dots, X_i^{T-p}\}$ and exogenous variables Y_i^T.

(3) Normal Copula. The last part of the model is a copula, which ties the endogenous variables together to capture contemporaneous correlation not captured in the autoregressive specification.

In order to accommodate multiple endogenous variables (Heinen and Rengifo 2007), we choose a Gaussian copula. Its density is given by:

$$c(u,\Sigma) = |\Sigma|^{-\frac{1}{2}} \exp\left(\frac{1}{2}(u'(I-\Sigma^{-1})u)\right) \tag{3}$$

where Σ is the covariance matrix that capture the contemporaneous correlations correlation between endogenous, and u is a vector of the normal quantiles of the probability integral transforms (PITs) (Heinen and Rengifo 2007) of the endogenous variable under the marginal ZIP densities. u measures the fitness of modeling X via ZIP. So, maximizing $c(u,\Sigma)$ via Σ will capture the correlation among endogenous variables that is not explained by the lagged effects.

After considering the above three components, the joint density of the ZMAP for all N consumers can written as the product of marginal Zero-Inflated-Poisson densities and the multivariate normal copula:

$$P(X|B,G,\Sigma) = \prod_{i=1}^{N} f_{ZIP}(X_i,B,G)\times c(u_i,\Sigma) \tag{4}$$

The parameters (B,G,Σ) could be estimated by maximizing the Expression (4) using a Quasi-Newton's method. These parameters determine the evolution of the consumer time series data. Specifically, dynamic interaction between activities within the Path-to-Purchase are captured through the cross/own variable lagged effects (B,G) and contemporaneous effects (Σ). Thus, we could use (B,G,Σ) to describe Path-To-Purchase for each consumer. But before that, we need to segment consumers based on (B,G,Σ).

4.2 Clustered ZMAP

We now extend ZMAP to mixture of ZMAP to accommodate the heterogeneity of shopping behavior and consumers' reaction to the marketing campaigns. The underlying data generating process for ZMAP mixture is:

1. Each consumer is drawn from one of K different clusters with certain probability.
2. The time series data for the consumer is generated by ZMAP model of the cluster.

The Bayesian Network of CZMAP model is described in Figure 2.

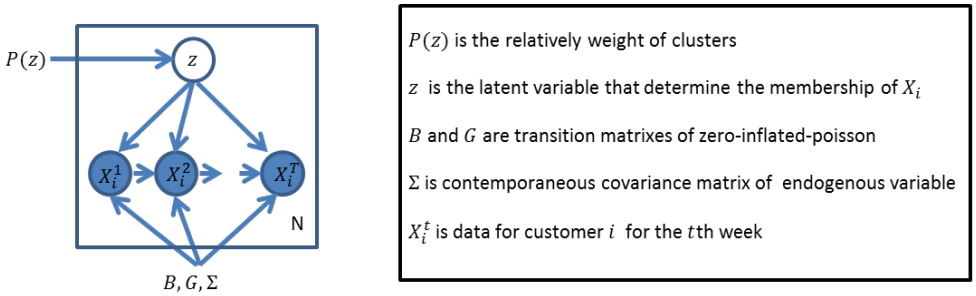


Figure 2. The Bayesian network of Clustered ZMAP model

The log likelihood of the ZMAP mixture is:

$$\sum_{i=1}^{N} In\left\{\sum_{k=1}^{K} P(z_k)P(X_i|B_k, G_k, \Sigma_k)\right\} \quad (5)$$

Because sum function is within the log function in (5), it is hard to get the closed form estimation of (B, G, Σ) to maximize (5). Then, we use an algorithm in the Expectation Maximization framework to estimate the proposed model. This algorithm has the following two iterating steps:

In E step, we estimate the membership of each consumer using the current estimate of the parameters:

$$r_{ik} = P(z_k = 1|X_i) = \frac{P(z_k)P(X_i|z_k = 1)}{\sum_{u=1}^{K} P(z_u)P(X_i|z_u = 1)} = \frac{P(z_k)P(X_i|B_k, G_k, \Sigma_k)}{\sum_{u=1}^{K} P(z_u)P(X_i|B_u, G_u, \Sigma_u)} \quad (6)$$

where r_{ik} represents the probability of assigning consumer i to cluster k.

In M step, we find the values of (B_k, G_k, Σ_k) that maximize the following expectation of the complete log likelihood.

$$\sum_{i=1}^{N}\sum_{k=1}^{K} r_{ik}\log\left(f_{ZIP}(X_i, B_k, G_k)\right) + \sum_{i=1}^{N}\sum_{k=1}^{K} r_{ik}\log\big(c(u_i, \Sigma_k)\big) + \sum_{i=1}^{N}\sum_{k=1}^{K} r_{ik} InP(z_k) \quad (7)$$

Thus, B_k, G_kcan be derived from optimizing $\sum_{i=1}^{N} r_{ik}\log\left(f_{ZIP}(X_i, B_k, G_k)\right)$ from (7) for each k, which is equivalent to estimating a weighted (r_{ik} *is the weight*) Zero-Inflated-Poisson regression. On the other hand, Σ_k can be estimated via maximizing $\sum_{i=1}^{N} r_{ik}\log\big(c(u_i, \Sigma_k)\big)$ for each k, which is equivalent to solving the Gaussian Mixture Model. According to (Bishop 2006), the maximal likelihood estimation for the he Gaussian Mixture Model is:

$$\hat{\Sigma}_k = \frac{\sum_{i=1}^{N} r_{ik} u_i u_i'}{\sum_{i=1}^{N} r_{ik}} \quad (8)$$

where u_i is the normal quantiles of the probability integral transforms of the endogenous variable X_i under the marginal ZIP densities. The EM algorithm executes E step and M step iteratively until the membership assignment/estimated parameters converge. After that, we get two results: the membership assignment is accomplished in E step and we can get parameters $B_k, G_k, \Sigma_k$for each cluster in M step, which will be used to describe Path-to-Purchase for each cluster later.

4.3 Path-to-Purchase Construction

As mentioned above, the dynamic interaction features of Path-to-Purchase are captured through the cross/own variable lagged effects and contemporaneous effects. Although, these information exist within (B_k, G_k, Σ_k), it is hard to interpret the parameter B_k, G_k, Σ_k directly (Sims 1980). We will summarize instead the over-time impact of a unit increase (shock) in a variable over its baseline on all endogenous variables. This is typically done using orthogonalized impulse response approach (Stephen and Galak 2012), which takes into account contemporaneous correlations captured within Σ_k. In so doing, this allows shocks to any variable to be simultaneously accompanied by shocks to other variables as dictated by their contemporaneous correlations. We follow the similar procedure in (Stephen and Galak 2012) to get orthogonalized impulse response and employ bootstrap (Sims et al. 1999) to get its confidence interval.

We illustrate the Path-to-Purchase construction with an example of the offline purchase response to a catalog, as shown in Figure 3. When faced with catalogs, the response of offline purchase reaches its peak in the second period. Thus, we focus on the offline purchase node at second period and find which of the possible pathways, from receiving a catalog to increase in offline purchase in second period, contributes most to the increase in purchase by tracing back from the offline purchase node[1]. We name this "predominant" path as the Path-to-Purchase. We list these Paths-to-Purchase for each cluster in Section 6.2.

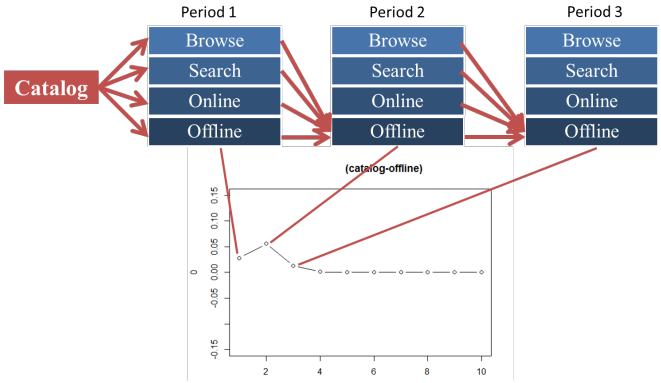


Figure 3. Example of Impulse Response.

5 Validation using Simulated Data

In order to validate the effectiveness of the CZMAP model, we need a dataset with known cluster membership so that we can compare the result of CZMAP clustering with it. Since we don't have access to such a dataset, we simulate one with cluster membership based on a mixture of ZMAP model. Then we apply the proposed algorithm to uncover the clusters and estimate the parameters.

We generate a dataset with 4 clusters, 500 customers in each, 5 endogenous variables, 4 exogenous variables, and a lag of 2. We generated different datasets spanning T=100, 200 and 500 time periods. For each cluster, we randomly set the B_k, G_k, Σ_k by drawing from a Normal Distribution (mean=0, variance=0.15). For each consumer, we generate the first two periods of endogenous data with a uniform distribution (min=0, max=2). We also generate exogenous data for each consumer based on the same uniform distribution. Based on the first two periods endogenous data, exogenous data and the corresponding transition matrixes B_i, G_i and covariance matrix Σ_k, we generate the endogenous data of remaining T-2 periods for that consumer.

	T=100	T=200	T=500
CZMAP	0.967	0.982	0.989
K-Means	0.743	0.832	0.959

Table 1. Clustering performance comparison between CZAMP and K-Means.

We compare the accuracy of the clusters identified by CZMAP to those identified by K-means (estimated ZMAP coefficients separately for each customer, then clustered consumers based on

[1] The procedure of tracing back is shown in Appendix B.

these coefficients). We obtain 5 runs for each case and use the averages across these runs. We employ purity (Manning et al. 2008), measured as the *fraction of the consumers assigned to the cluster containing majority of their original group members,* to measure the clustering accuracy. The results are shown in Table 1. We find that:

- K-means does not performance well when T=100 or 200. This is because the estimation of individual-level ZMAP coefficients is inaccurate due to limited amount of individual-level data and the large number of parameters that need to be estimated, leading to a decrease in the clustering performance. In some situations with extreme sparseness, we are unable to estimate the ZMAP using data of individual consumers.
- When the length of time-series data is longer, the performance of K-means improves. In contrast, CZMAP performs well even with shorter time-series data. The reason is that when estimating the coefficients, CZMAP uses weighted data of all consumers that belong to a cluster rather than individual data. So, it doesn't suffer from the sparsity issue. Overall, CZMAP outperforms K-means for all three simulated datasets.

6 Empirical Results and Application

6.1 Validation by out-of-sample prediction

To further validate the CZMAP model using real world data, we measure its out-of-sample prediction ability. We compare it to with two alternative approaches: 1) Estimated a ZMAP for **each** consumer without clustering based only on the consumer's own data. 2) A ZMAP model for all consumers where everyone is assumed to have a set of common parameter (CZMAP with number of clusters = 1 or C_1ZMAP). We divide the 106 weeks data of all consumers into training set (first 96 weeks) and test set (last 10 weeks), we train CZMAP with K=3, determined by AIC and BIC) and two alternative approaches using the training data. Then we predict the number of online purchase and offline purchase in each period of the test set using the trained models and the observed exogenous variables in the same period. The Mean Square Error of CZMAP is 0.74 compared to 1.20 for C_1ZMAP and 1.13 for individual ZMAP. This shows that the CZMAP predicts much better than alternative approaches, hence a more appropriate model for our data.

6.2 Path-to-Purchase for each cluster

After successful validation, we apply the CZMAP algorithm to the retailer dataset to identify the predominant paths to purchase. The number of clusters is determined using the AIC and BIC criteria. On searching over $K \in \{2,3,4,5\}$ the selection criteria lead us to $K = 3$. The relative size of three clusters are $\{56.4\%, 35\%, 8.5\%\}$. Following the procedure in Section 4, we get the predominant Path-to-Purchase for each cluster, which is listed in Table 2. We can see that predominant Path-to-Purchase in different clusters are quite different from each other. To further describe our cluster the demographic composition for each cluster are given in Appendix C.

Start	End	Cluster 1	Cluster 2	Cluster 3
Email	Online Purchase	***Email***-online purchase	***Email***-online purchase	
Email	Offline Purchase		***Email***-browse-offline purchase	
Catalog	Online Purchase	***Catalog***-search-online purchase		
Catalog	Offline Purchase	***Catalog***-offline purchase	***Catalog***-search-offline purchase	
Holiday	Online Purchase	***Holiday***-online purchase		***Holiday***-browse-browse-online purchase
Holiday	Offline Purchase			***Holiday***-offline purchase

Table 2. Predominant Path-To-Purchase comparison for different clusters[2]

Cluster 1. Catalog Driven shoppers. This segment comprises a majority of the population (56.4%). Catalogs are very effective for consumers in this cluster. In response, they engage in offline shopping during that week and they tend to search more products online, which results in increase online shopping next week. Beside catalog, these customers respond to email and holidays, through increased online shopping. In summary, these customers respond to catalogs and are more likely to engage in online shopping.

Cluster 2. Email Driven shoppers. This group consists of 35.1% of the population. In contrast to the consumers in cluster 1, consumers in this cluster are more sensitive to email. When they receive an email they engage in online shopping. Moreover, they browse more products when they receive email. These products enter their consideration set, resulting in subsequent *offline* purchase. Finally, we find there are more young consumers (age<=40) in this cluster than cluster 1 and 3, which indicates that young consumers are more sensitive to email.

Cluster 3. **Holiday shoppers.** This is a relatively small group of the population (8.5%). In contrast to the customers in the first two clusters, the consumers in this cluster are not sensitive to either email or catalogs. However, they are active offline shoppers during holidays. In addition, they browse more products online during and the week after holidays leading to online purchases up to two weeks later. Examining the demographics, we find that the proportion of older consumers (age>40) in this cluster are much larger than that in cluster 1 and 2 and they are the majority of this cluster. This seems to indicate that older consumers are more active during holidays.

6.3 Application to target marketing

In this section we show that we can use the proposed algorithm to identify the groups of customers who are most and least sensitive to a particular type of marketing. This can be used to identify the right type of advertisement for those customers.

We generate three clusters based on data in the first 96 weeks. The remaining 10 weeks were set aside as test data. For each cluster, we computed the cumulative orthogonalized impulse response (Sims 1980), shown in Table 3. This measures the cumulative long term effects of a unit increase (shock) in an exogenous variable on overall purchase (*both online and offline*) for

[2] The empty Path-To-Purchase indicate insignificant effect on the type of purchase.

each cluster. Based on the results in Table 3 we predict that catalogs would be most effective for cluster 1, followed by clusters 2 and 3. Email would be the most effective marketing campaign for cluster 2, closely followed by cluster 1, while cluster 3 is least sensitive to email.

	Cluster 1	Cluster 2	Cluster 3
Catalog	0.16072	0.08754	0.08258
Email	0.01076	0.01752	0.00525
Holiday	0.06871	0.062941	0.08064

Table 3. long term effects of exogenous on overall purchase for each cluster.

To verify our predictions, we divide each cluster of consumers into two parts: those who actually received more than median number of emails in the last 10 weeks or test data (sub-cluster H) and those who received less than the median (sub-cluster L). According to our prediction, the ***difference*** between the cumulative purchase of these H and L groups should be *maximum for the most sensitive cluster (cluster 2) and minimum for the least sensitive cluster (cluster 3)*. We report the average weekly online and offline purchase for the H and L sub-clusters in Table 6. We find that difference of average weekly purchase between the H and L sub-clusters for emails is the highest in cluster 2 and lowest in cluster 3. Similarly the difference between the H and L sub-clusters for catalogs is highest for cluster 1 and lowest for cluster 3. These observations match our predictions.

	Effect of Email			**Effect of Catalog**		
	Sub-cluster L	**Sub-cluster H**	**Difference**	**Sub-cluster L**	**Sub-cluster H**	**Difference**
Cluster 1	0.2243	0.2094	**0.0149**	1.2756	1.0431	**0.2325**
Cluster 2	0.2854	0.2681	**0.0173**	0.9345	0.7864	**0.1481**
Cluster 3	0.1958	0.1897	**0.0061**	0.8915	0.7568	**0.1347**

Table 6. Predict the average weekly overall purchase for two sub-clusters for each cluster.

This shows that the CZMAP can be used to identify the groups of customers for whom a particular marketing strategy could be most effective.

7 Conclusion

In this paper**,** we propose an approach, called CZMAP, to identify predominant paths-to-purchase from consumer-activity data and segment consumers based on these paths. We verify the effectiveness of the proposed approach in clustering using simulated data and in prediction using real-world data. We then identify different consumers' path to purchase from a large real world dataset collected from a CRM system of a retailer. Finally, we show an application of the proposed approach to targeted advertisement.

The proposed approach could be used by a retailer to understand the shopping behavior of its customers. It could also guide the retailer in tailoring the marketing mix for different customer segments based on different paths they take to purchase.

References

Baragona, R. (2011). Clusters of Multivariate Stationary Time Series by Differential Evolution and Autoregressive Distance. Pattern Recognition and Machine Intelligence. S. Kuznetsov, D. Mandal, M. Kundu and S. Pal, Springer Berlin Heidelberg. 6744: 382-387.

Bishop, C. M. (2006). Pattern recognition and machine learning. New York, Springer.

David Court, D. E., Susan Mulder, and Ole Jørgen Vetvik (2009). The consumer decision journey. McKinsey Quarterly.
Heinen, A. and E. Rengifo (2007). "Multivariate autoregressive modeling of time series count data using copulas." Journal of Empirical Finance 14(4): 564-583.
Lambert, D. (1992). "Zero-inflated Poisson regression, with an application to defects in manufacturing." Technometrics 34(1): 1-14.
Li, H. and P. K. Kannan (2013). "Attributing Conversions in a Multichannel Online Marketing Environment: An Empirical Model and a Field Experiment." Journal of Marketing Research 51(1): 40-56.
Manning, C. D., P. Raghavan and H. Schütze (2008). An introduction to information retrieval. Cambridge, Cambridge University Press.
Pauwels, K., P. S. H. Leeflang, M. L. Teerling and K. R. E. Huizingh (2011). "Does Online Information Drive Offline Revenues?: Only for Specific Products and Consumer Segments!" Journal of Retailing 87(1): 1-17.
Sims, C. (1980). "Macroeconomics and Reality." Econometrica 48(1): 1-48.
Sims, C. A. a. Z. T. (1999). "Error Bands for Impulse Responses." Econometrica 67(5): 1113--1155.
Stephen, A. T. and J. Galak (2012). "The Effects of Traditional and Social Earned Media on Sales: A Study of a Microlending Marketplace." Journal of Marketing Research 49(5): 624-639.
Xu, L., J. A. Duan and A. Whinston (2014). "Path to Purchase: A Mutually Exciting Point Process Model for Online Advertising and Conversion." Management Science 60(6): 1392-1412.

Appendix A

Type	Variables	Definition	Min	Mean	Median	75%	Max
Endogenous	Browse	number of product browsed online	0	1.321	0	0	854
	Search	number of searches within the web site	0	0.00325	0	0	13
	Online Purchase	number of purchase via online channel	0	0.04043	0	0	125
	Offline Purchase	number of purchase via offline channel	0	0.1756	0	0	86
	Return	number of returns	0	0.03384	0	0	50
Exogenous	Catalog	number of marketing campaigns received via catalog	0	0.1067	0	0	2
	Email	number of marketing campaigns received via email	0	0.9339	0	2	20
	Holiday	Holiday dummy	0	0.1635	0	0	1
	Promotion	the number of available promotions	1	7.596	5	9	34

Description of weekly endogenous and exogenous variables in the retailer dataset.

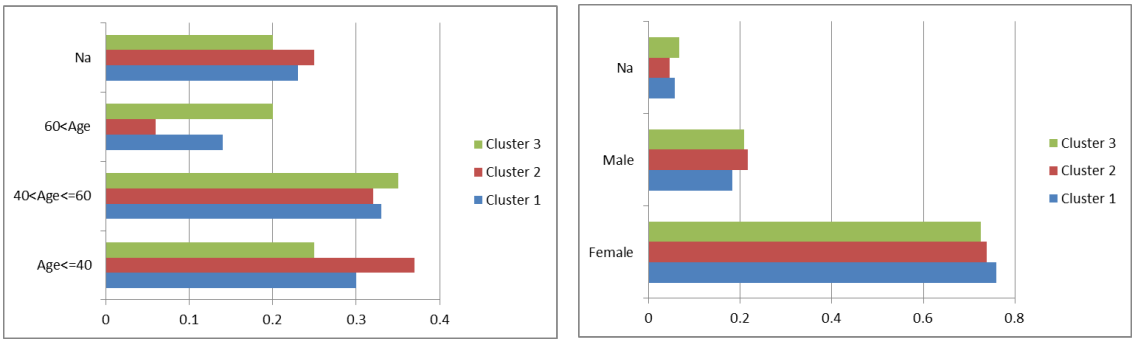
Appendix B

The pseudo code of finding predominant Path-to-Purchase by tracing back:

1. **For each** endgeoneous varible e
   a) **For** exogenous variable v
      I. Get the impulse response IR_{e_v} [1:10] from v to p from week 1 to week 10
   b) **End For**
2. **End For**

3. **For each** endgeoneous varible e
   a) **For** p in [offline purchase, online purchase]
      I. Get the maximal $IR_{ep}[t']$ from IR_{ep} [1,10]
      II. $Path[t'] = p$
      III. t=t'
      IV. While t>0
         I. For each exogenous variable v at *t-1* week
            i. Get IR_{e_v}[t-1]
            ii. Calculate the contribution $C[v]$ from IR_{e_v}[t-1] to $IR_{e\_Path[t]}$[t] based on formula (1) and (2)
         II. End For
         III. Get the maximal $C[v]$
         IV. $Path[t-1] = v$
         V. $t = t - 1$
      V. End While
      VI. Print $Path[1:t']$ form e to p
   b) **End For**
4. **End For**

Appendix C



The numerical proportion of age and gender for 3 clusters.