

Uncovering Characteristic Paths to Purchase of Consumers

Yicheng Song

(ycsong@bu.edu)

PhD Candidate in Information Systems

Nachiketa Sahoo

(nachi@bu.edu)

Assistant Professor of Information Systems

Shuba Srinivasan

(ssrini@bu.edu)

Adele and Norma Barron Professor in Management, Boston University

Chrysanthos Dellarocas

(dell@bu.edu)

Richard Shipley Professor in Management, Boston University

Boston University, Questrom School of Business

595 Commonwealth Ave, Boston, MA 02215

January 20, 2016

Uncovering Characteristic Paths to Purchase of Consumers

Abstract

A digital consumer's purchase journey, referred to as the path to purchase, is non-linear and heterogeneous. Despite a strong interest in this concept, there are few published approaches to empirically extract consumers' path to purchase (in terms of a sequence of different types of activities leading to purchase), especially in settings where consumers engage in multiple simultaneous activities in each period. We address this gap by proposing a methodology that identifies consumers' paths to purchase from commonly available CRM touch point data. We propose a generalized multivariate autoregressive (GMAR) model to capture the interactions among distinct but potentially simultaneous activities of a consumer over time. Using the proposed model we show how to attribute parts of the purchase volume to consumer activity sequences, or paths, starting from an initial marketing stimulus leading to the maximal purchase response. We embed the GMAR model in a clustering framework that endogenously identifies segments of consumers who exhibit similar paths to purchase. We apply the methodology to a dataset from a multi-channel North American Specialty Retailer to uncover the distinct paths of five consumer segments: loyal and engaged shoppers, digitally-driven offline shoppers, holiday shoppers, infrequent offline shoppers, and frequently targeted occasional shoppers. Using out-of-sample forecasts, we demonstrate improved predictions of future purchases compared to extant methods. Finally, we perform policy simulations to show that managers can use the uncovered path information to dynamically optimize marketing campaigns for each segment.

Keywords: Path to Purchase, Generalized Multivariate Autoregressive (GMAR) model, Market Segmentation, Dynamic Programming, Optimal Marketing Policy.

Uncovering Characteristic Paths to Purchase of Consumers

1 Introduction

Managers have long been interested in understanding the consumers' purchase decision process so that they can make effective marketing decisions (Strong Jr 1925, Mullen 1956). When shopping was primarily done offline, a purchase funnel where customers sequentially proceed through the stages of awareness, familiarity, consideration, liking and purchase, served as a useful model of this process (Barry 1987). However, in the era of online shopping, customers gather information, evaluate, and seek opinions on products from a variety of easily accessible online and offline sources in a manner that best fits their unique needs (Nunes et al. 2012). In addition, consumers are exposed to various marketing stimuli (such as online advertisements, email marketing, etc.) that constantly alter their shopping process. As a result, the concept of the purchase funnel has evolved to that of *path to purchase* to describe shopping behavior that is non-linear (Court et al. 2009, Wolfersberger and Monteleone 2013) and heterogeneous across consumers (Brown 2013, Google 2014) (see Figure 1). One can no longer rely on a universal model of purchase decision process. To be truly effective, there is a need to identify the paths to purchase of different consumers and develop personalized marketing strategies based on them (Lovejoy 2014).

We posit that different groups of consumers have distinct *characteristic paths* that describe their most common behavior over time leading to purchase in response to marketing stimuli. For example, after receiving a promotional email from a retailer, an Internet savvy user might prefer to spend time comparing products online and reading reviews before buying online or offline, whereas a user who does not trust online reviews might prefer to visit a retail store to evaluate products directly before buying. To the extent that a key objective of identifying paths to purchase is to design effective marketing strategies based on them, it is important to develop methods to segment consumers based on their characteristic paths.

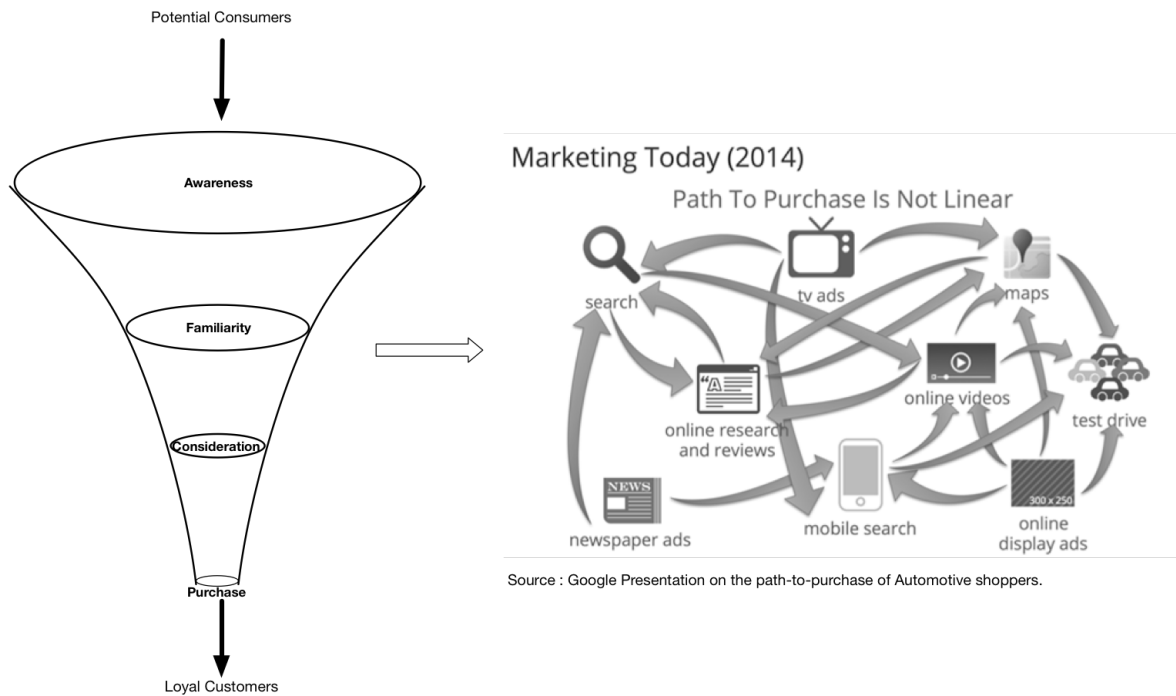


Figure 1. From Purchase Funnel to Path to Purchase.

In recent years, researchers have analyzed the transition of online shoppers through retailer webpages (Bucklin et al. 2002, Moe 2003, Montgomery et al. 2004) as well as the movement of offline shoppers through physical stores (Farley and Ring 1966, Larson et al. 2005). However, few have developed methods that attempt to identify the characteristic sequence of the activities that lead to purchase—which is typically how path to purchase is conceptualized (Nunes et al. 2012, Kimelfeld 2013).

Existing papers on path to purchase consider only one action of the consumer in each step of the journey (Montgomery et al. 2004, Larson et al. 2005, Xu et al. 2014). But, often consumers are observed to do more than one type of activities in the same period, or to start actions that last for several periods and overlap with other actions in future periods. For example, in e-commerce environments consumers are often observed to begin product browsing sessions whose results they take several periods to fully examine. In the meantime they perform simultaneous, and potentially correlated, keyword searches and purchases over multiple periods. In such settings, an activity in a period may lead to subsequent activities in the future, but these effects are observed combined with the effects of other activities from the same time period. Existing methods do not attempt to uncover consumer’s path to purchase through such simultaneous, and potentially correlated, activities across multiple categories.

Our objective is to fill the gap by providing an approach to identify paths to purchase of different segments of consumers in the form of sequence of activities that the subsequent purchases can be attributed to. We accomplish our objective in three steps. First, we propose a model of interactions between distinct, and potentially simultaneous, types of activities of a consumer over time, through which we capture the distribution of the consumer's activity sequences (paths) that ensue between a starting impulse (e.g. receipt of a catalog) and the maximal purchase activity. Second, we provide a method that extracts a consumer's characteristic paths (i.e. the paths that explain the most of the consumer's purchase response to the impulse) from the previous representation. Third, we develop a clustering algorithm that simultaneously identifies segments of consumers with distinct paths and the segment-specific models that represent those paths.

The rest of the paper is organized as follows: we discuss the related literature in Section 2. Then, we describe our proposed approach, which we call Clustered Generalized Multivariate Autoregressive Model (CGMAR), in Section 3. In Section 4, we apply the CGMAR algorithm to a real world dataset and discuss the identified paths of segments of consumers. In Section 5, we highlight two managerial applications of the proposed approach: for target marketing by evaluating out-of-sample predictions of purchase and for optimizing marketing mix by performing policy simulations. Finally, in Section 6, we conclude by summarizing the contributions and outlining opportunities for future research.

2 Literature Review and Key Contributions

Understanding the shopping processes of online consumers has been a topic of interest in Information Systems and Marketing disciplines for several years (Gefen et al. 2003, Xu et al. 2014, Ghose and Todri 2015). Given our focus on uncovering path to purchase and segmenting consumers based on their paths, we review three related research streams: analysis of consumer's paths, consumer behavior-based market segmentation, and time-series clustering.

2.1 Consumer's Path-related Research

Paths in physical shopping environments have been studied as early as in 1960s (Farley and Ring 1966). Since then researchers have studied individual consumer's movements through physical environments of

supermarkets and malls (Heller 1988, Underhill 2005). Some of the latest work in this space use RFID technology to track the movement of shopping carts (Sorensen 2003, Hui et al. 2009) and have applied time series clustering techniques to gain insight into the purchase behavior of the customers (Larson et al. 2005). In these studies each step of a consumer on the path through the physical store is taken to denote the presence of the consumer at a particular zone—each a proxy for certain type of purchases by the consumer.

Research on customers' paths through webpages in an E-commerce environment has been growing recently due to the availability of clickstream data (Bucklin and Sismeiro 2003, Montgomery et al. 2004). Montgomery et al. (2004) provides a dynamic choice model of shoppers' browsing behavior. Although, they provide an elegant model of how shoppers transition from one category of web-page to another that results in improved predictions of purchase conversions, it does not attempt to uncover paths of the users—with a well-defined start point and a sequence of activities leading to purchase. Recently, Xu et al. (2014) propose a point process that captures the effect of consumer's past clicks on advertisements and purchases on future activities aiming to attribute a purchase to various types of advertisements that the consumer was exposed to prior to purchase. For example, they find that while display advertisements have relatively low direct effect on purchase conversion, they are more likely to stimulate subsequent visits through other advertisement formats. But, extracting the complete paths of the consumers is not one of their objectives.

All the existing papers in path analysis only consider the scenario that users carry out one isolated activity at any step of the consumer's journey. In contrast, our proposed method can extract paths when users are observed to carry out many types of, possibly correlated, activities in each step of their path to purchase. To the best of our knowledge we are the first to propose such a method.

2.2 Market Segmentation Approaches

Segmentation based on consumers' behavior such as their response to marketing offers and their choice of brand, products, and services has been studied at depth over the last three decades (Wedel and Kamakura 2000). Typically, they propose a mixture of statistical models of a static consumer behavior to capture the heterogeneity among consumers and segment them into groups with similar models (Wedel et al. 1993, Ramaswamy et al. 1994, Wedel and DeSarbo 1995, Deb and Trivedi 1997, Bucklin et al. 1998).

Segmentation based on the evolution of consumers' behavior over time, however, is relatively rare. Reimer et al. (2014) segment a set of consumers based on their individual short-term response to marketing and then study the long-term evolution of average behavior, such as weekly spend and number of coupons used, of each segment using a Vector Autoregressive (VAR) model. The underlying assumption is that consumers who are similar in their short-term response to marketing will have similar evolution in long-term aggregate shopping behavior, which may not be generally true. While there is ongoing interest in segmenting based on behavior and in identifying groups of consumers whose shopping behavior evolves in a similar way, there remains a gap in the literature in segmenting consumers directly based on how their activities evolve.

2.3 Time-Series Clustering Approaches

Research on market segmentation is often done based on consumer's static properties. However, the movement of consumers on a path implies dependence between their past and future positions in a space of interest. In other words, the path data are time series observations of consumers' positions. Three types of approaches has been proposed in statistics, computer science, and management science research to cluster time series data (Liao 2005).

Raw-data based approaches work directly with time series data by defining a similarity measurement between them. For example, (Larson et al. 2005) use a k-medoid clustering algorithm with Euclidian distance between the paths to analyze movement of shopping carts in supermarkets. *Feature-based approaches* convert individual raw time series into a feature vector of lower dimension on which a conventional clustering algorithm is applied. The 2-step process avoids the dimensionality problem of comparing arbitrarily long-time series. Kalpakis et al. (2001) estimate autoregressive model coefficients for univariate series data and then apply partition-based clustering on the estimated coefficients. Baragona (2011) estimates a Vector Autoregressive (VAR) model for each multivariate series and then group the time-series based on their VAR coefficients using a K-means clustering algorithm. VAR models rely on Gaussian error assumption, which is readily met when the variables are group-level statistics but fails to be met when

modeling individual customer behavior over time.¹

Model-based approaches propose statistical mixture model-based data generation process for the observed times-series data. By fitting the model to the data, researchers endogenously determine the assignment of different time series to different clusters and the model parameters of each cluster. Mixture models avoid the sparsity problem that one would run into in the reduction step of feature-based approaches while estimating separate parameters of individual time-series, especially when time series are short. One example of such approach is Oates et al. (2001) who use a mixture of Hidden Markov Model (HMM) to cluster multivariate time-series data. Xiong and Yeung (2004) cluster univariate time-series data by using a mixture of Autoregressive–Moving-Average (ARMA). Fong et al. (2007) extend the mixture model to cluster multivariate time series by replacing the ARMA model with the Vector Autoregressive (VAR) model. They apply their model to 1 and 3 year US Treasury constant maturity rates where the variables follow a log-Normal distribution, hence the distributional assumptions of the VAR model are satisfied.

To sum up, there are two limitations of such time-series based methods. First, they are typically based on aggregate data and not on individual level data, which is necessary to segment customers based on their individual paths to purchase. Second, observations of individual customer behaviors in any time period rarely follow the Gaussian distribution. Accordingly, a more flexible distribution to model individual customer's activity is called for.

2.4 Contribution of the Current Work

Our paper contributes to the above three streams in several ways. First, we contribute to the literature on path to purchase by providing a method to empirically extract characteristic paths to purchase in settings where consumers may engage in several simultaneous, and potentially correlated, activities in each time period. Moreover, we provide a way to graphically show the paths and their relative weights so that an analyst may readily identify the important characteristic paths of consumers. Second, this research contributes to the literature on market segmentation by providing a method to segment based on consumer's characteristic path

¹ Reimer et al. (2014) avoid this problem by modeling the evolution of aggregate properties of the segment obtained through another means (i.e. their short-term response coefficients) instead of modeling the evolution of individual consumers and segmenting consumers based on that.

to purchase. At the modeling level, our approach captures the heterogeneity in individual consumers' paths. At the same time, it overcomes the sparsity problem that would occur if one were to estimate the consumer's path for each individual separately. Third, we contribute to the literature in time-series clustering by providing a method to cluster multivariate (vector) time series when the observations do not follow the Gaussian distribution necessary to apply the VAR model. Our proposed approach can incorporate any exponential family distribution or variants of those found in the Generalized Linear Model (GLM) family, which opens it up to applications in a wide variety of settings. For example, as shown in the current paper, by using a zero inflated Poisson distribution the model can be applied to customer activity time-series data that is readily available in many CRM systems. Finally, at the application level for practitioners, we provide both a forecasting tool that allows managers to predict purchases of individual customers and a dynamic optimization tool for optimizing marketing mix.

3 Model Development

3.1 Context and Concept of Path to Purchase

There are many events that can initiate a consumer on her path to purchase, e.g., receipt of a catalog or a promotional email, a holiday, or a personal reason such as an anniversary or a friend's birthday. Given our focus on empirically identifying paths from available data, we consider starting impulses that are observable to the firm such as the marketing communications the firm sends to its consumers. How consumers respond to promotional catalogs and email messages is a topic of intense interest for the managers (Zantedeschi et al. 2013). A catalog might get a consumer to look for product information online. In doing so, the consumer might better realize the characteristics of the products that she is interested in and perform keyword searches to find similar products. Finally, the consumer may go to a physical store to purchase the identified products. Therefore, purchase as a response to an impulse could occur after several time periods. Such purchase response observed for a group of homogenous consumers would be observed to peak after a certain number of time periods. We use this peak or maximal purchase response as the end point of the consumer's path. We define the consumers' path to purchase as a sequence of activities that start with the marketing impulse and

end with the maximal purchase response (Figure 2).

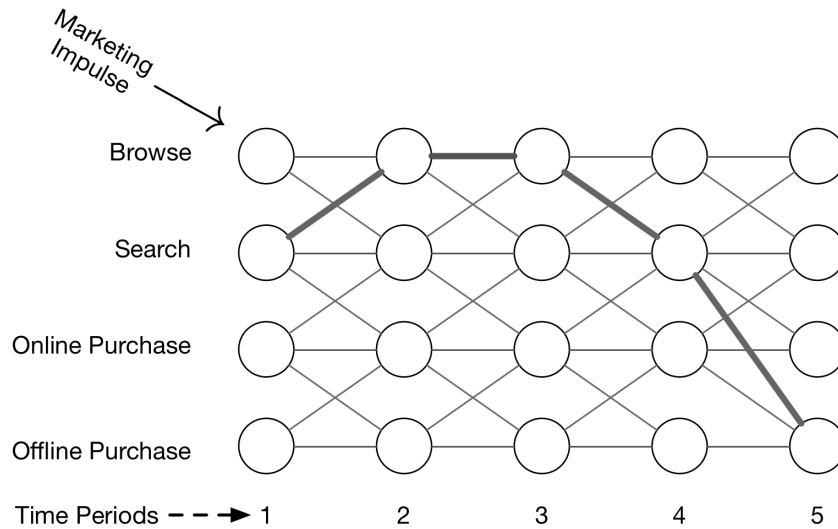


Figure 2. Illustration of path to purchase of a consumer through sequence of activities.

Although a given consumer may follow a variety of paths at different times, we hypothesize that each consumer can be associated with a *characteristic path behavior*, defined as a distribution over the consumers' paths to purchase. Our objective is not to identify the path of a consumer to a particular purchase in our dataset, because the exact sequence of steps leading to a particular purchase made by a consumer will have little value for policy making in the future. Instead we focus on identifying for a group of consumers the typical sequence of activities after the impulse that the subsequent peak purchase activity can be credited to. Although we allow for different consumers to have their own characteristic paths, we assume that the path of a consumer does not change over the data collection period. In addition, our model accommodates many possible paths of a consumer and estimates the relative frequency with which the consumer takes each of them, but we don't model the selection of a path as dependent upon consumer's objectives and type of products purchased, because of the difficulty in associating these variables to consumer's every shopping related activity.

The rest of this section will describe the components of our method. Deriving a consumer's path to purchase in response to a starting impulse requires a model of transition between consumer-activities over time. In Section 3.2 we develop such a model, starting from the well-known Vector Autoregressive (VAR) model, and extending it to account for non-Gaussian errors and the sparsity of individual-level data that

characterizes most settings of relevance to this paper. Once we have a suitable multivariate autoregression model in place, in Section 3.3 we explain how a consumer’s characteristic paths to purchase can be extracted from it. Finally, in Section 3.4 we describe how the preceding model infrastructure forms the basis of a clustering algorithm that simultaneously identifies groups of consumers with distinct path to purchase behavior and the group-specific models that represent those paths.

3.2 Generalized Multivariate Autoregression for Modeling Path to Purchase

Vector Autoregressive (VAR) models (Sims 1980) have been used to model parallel time series of aggregate data, ranging from the evolution of music sales, buzz and radio play to (Dewan and Ramaprasad 2014) to evolution of stock market indices (Tsay 2005). The VAR model is given as

$$\mathbf{x}^t = \mathbf{B}_1 \mathbf{x}^{t-1} + \mathbf{B}_2 \mathbf{x}^{t-2} + \dots + \mathbf{B}_p \mathbf{x}^{t-p} + \mathbf{C} \mathbf{y}^t + \boldsymbol{\epsilon}^t \quad (1)$$

where, \mathbf{x}^t is a $J \times 1$ vector of endogenous variables that contains the J components of the time series observed in period t . \mathbf{y}^t is a $J_e \times 1$ vector of exogenous variables. p is the number of lagged values of the endogenous variables considered, also known as the order of the autoregressive model. \mathbf{B}_p is a $J \times J$ matrix of coefficients for endogenous variables. \mathbf{C} is a $J \times J_e$ matrix of coefficients for the exogenous variables, and $\boldsymbol{\epsilon}^t \sim \mathbf{N}(\mathbf{0}, \boldsymbol{\Sigma})$ is a multivariate Gaussian error. Advantages of using VAR model include the ability to model the temporal dynamics among the endogenous variables, ability to estimate contemporaneous correlation through the off-diagonal elements of error covariance matrix, and the availability of summary tools such as impulse response functions and error variance decompositions to generate empirical insights (Lütkepohl 2007). However, the assumption of Gaussian error in the VAR makes it unsuitable when the data does not follow such a distribution (Heinen and Rengifo 2007, Stephen and Galak 2012). In our setting, the incidence of the different types of consumer activities across time periods is sparse, rendering the Gaussian error assumption inappropriate. To overcome this challenge, we propose a Generalized Multivariate Autoregressive (GMAR) Model based on a non-Gaussian error assumption, which we outline below.

3.2.1 Model for the components of the time series

We start by specifying a model for each component of the time series:

$$x_{ij}^t \sim f(\mathbf{M}_i^t; \boldsymbol{\theta}_j) \quad (2)$$

where x_{ij}^t is the j 'th component of the endogenous variable of the i 'th consumer in period t . \mathbf{M}_i^t is the $(J \times p + J_e) \times 1$ vector that consists of lagged values of J endogenous variables over p previous time periods $\{x_i^{t-1'}, x_i^{t-2'}, \dots, x_i^{t-p'}\}'$ and J_e other exogenous variables, $\mathbf{y}_i^{t'}$, hypothesized to affect the user i at time period t . $\boldsymbol{\theta}_j$ is a $(J \times p + J_e) \times 1$ coefficient vector. To make the implementation more concrete and use existing estimation techniques we assume $f(\cdot)$ to be a probability distribution in the exponential family or a variant, such as zero-inflated model, found in the Generalized Linear family of models (GLM). In particular, we assume that the mean of the endogenous variable $\mu_{ij}^t = E(x_{ij}^t) = g^{-1}(\boldsymbol{\theta}_j' \mathbf{M}_i^t)$ where $g(\cdot)$ is the GLM link function that converts the linear combination of the predictors into the mean of the outcome variable. This component-specific model can be estimated using the Iteratively Reweighted Least Squares method (Nelder and Wedderburn 1972).

3.2.2 Copula to capture the contemporaneous correlation

The above model captures the co-variation of the components of the multivariate time-series that are explained by the lagged endogenous variables and exogenous variables. However, there could be residual contemporaneous correlation among the endogenous variables not explained by the lags. To build a joint model of the multivariate time series from the marginal distributions shown in Eq. (2) that captures such contemporaneous correlations, we use the Normal copula technique (Trivedi and Zimmer 2005, Heinen and Rengifo 2007). According to this technique the Normal quantiles of the probability integral transform (PITs) of the endogenous variables are calculated separately for each series using the following formula:

$$x_{ij}^{*t} = \Phi^{-1} \left(F(x_{ij}^t | \mathbf{M}_i^t; \boldsymbol{\theta}_j) \right) \quad (3)$$

where, F is the cumulative density function (CDF) of f and Φ^{-1} is the inverse function of the CDF, or Quantile function, of a standard Normal distribution. It can be readily verified that x_{ij}^{*t} is zero when x_{ij}^t is equal to the median of $f(x_{ij}^t | \mathbf{M}_i^t; \boldsymbol{\theta}_j)$, positive when x_{ij}^t is higher than the median, and negative when it is lower than the median. By construction, x_{ij}^{*t} follows a Normal distribution. Therefore, x_{ij}^{*t} corresponds to the part of the endogenous variable that is not explained by the autoregressive part, similar to the error term in

Eq. (1). Under Normal copula the vector \mathbf{x}_i^{*t} follows the copula density (Song 2000):

$$c(\mathbf{x}_i^{*t}, \boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{x}_i^{*t'})'(I - \boldsymbol{\Sigma}^{-1})\mathbf{x}_i^{*t}\right) = c(\mathbf{x}_i^t; \boldsymbol{\Sigma}, \boldsymbol{\theta}) \quad (4)$$

The covariance matrix $\boldsymbol{\Sigma}$ captures the covariance of the endogenous variables that are not explained by the lagged endogenous variables and exogenous variables. The density assigned by the copula to the dependence of the original variables x_{ij}^t is a function of the parameter $\boldsymbol{\theta}_j$ because according to Equation (3), x_{ij}^{*t} is a function of model parameter $\boldsymbol{\theta}_j$. The copula density is defined as shown in Equation (4) when the x_{ij}^t are continuous variables. When they are not, such as counts of consumer activities, continued extensions of discrete variables proposed by Heinen and Rengifo (2007) could be used to first transform the x_{ij}^t to continuous variables before using the Equation (3). Using Equation (2) and (4), the joint density for all I consumers can be written as the product of marginal densities and the multivariate Normal copula:

$$P(X|\boldsymbol{\theta}, \boldsymbol{\Sigma}) = \prod_{i=1}^I P(X_i; \boldsymbol{\theta}, \boldsymbol{\Sigma}) = \prod_{i=1}^I \prod_{t=p+1}^{T_i} c(\mathbf{x}_i^{*t}; \boldsymbol{\Sigma}) \prod_{j=1}^J f(x_{ij}^t | \mathbf{M}_i^t; \boldsymbol{\theta}_j) \quad (5)$$

The parameters $(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ can be estimated using maximum likelihood estimation. These parameters determine the evolution of the consumer's time series. Specifically, dynamic interactions between activities are captured through the cross/own variable lagged effects $(\boldsymbol{\theta})$ and the contemporaneous effects $(\boldsymbol{\Sigma})$.

3.3 Extraction of Path to Purchase from GMAR model

As discussed in Section 3.1, we consider paths that start from a marketing impulse, such as email or catalog activity, and end with a purchase. As the coefficients of this multivariate time series model are difficult to interpret directly, due to the recursive nature of dependence, we compute the activities response of the consumers to a marketing impulse using the following function:

$$\boldsymbol{\mu}^t = m(\mathcal{M}_{mean}^t) = g^{-1}(\boldsymbol{\theta}' \mathcal{M}_{mean}^t) \quad (6)$$

where $m(\cdot)$ is the mean response function of the GMAR, $\mathcal{M}_{mean}^t = \{\boldsymbol{\mu}^{t-1'}, \boldsymbol{\mu}^{t-2'}, \dots, \boldsymbol{\mu}^{t-p'}\}'$ is the mean response to the impulse in the p previous periods. The values of $\boldsymbol{\mu}$ for each previous time period can be computed using a recursive call to the function $m(\cdot)$.

In our setting we are interested in how changes in the variables that model our starting impulses (email and catalog activities) affect purchases. Due to the effect of an impulse on all the endogenous variables and their effects on the endogenous variables in the subsequent periods, the purchase response to an impulse is often observed to peak after several time periods (Lütkepohl 2007). We refer to the peak response as the *maximal purchase response*. We employ the bootstrap approach outlined in Sims and Zha (1999) to obtain confidence intervals and use the maximal purchase response as the end point of the path only if the response is significant (p-value < 0.01). When there is no significant response on purchase, no meaningful path is considered to exist between the impulse and the purchase activity.

A feature of our setting is that it allows multiple activities to take place simultaneously in each time period. Between the start point of marketing impulse (y_{impulse}^0) and the end point of maximal purchase response (μ_{purchase}^m), which occurs after m time periods from the impulse, there are many possible paths that can be plotted among the activities in the intermediate time periods (Figure 2). However, not all of them contribute equally to the maximal purchase response. The contribution of each path can be obtained by separating the effect of the impulse on μ_{purchase}^m over each possible path. When the link function, $g(\cdot)$ in Eq. (6) is the Identity function, the effects corresponding to each path linearly separate. However, in case of GLMs with non-identity link functions such as Log or Logit, the contributions of each path are not linearly separable as presented in Equation (6). However, since the function $m(\cdot)$ is differentiable over real number space \mathbb{R}^{JP} , the response after m time periods of a unit increase in impulse can be expressed as

$$\Delta\mu_{\text{purchase}}^m = \int_0^1 \frac{\partial m_{\text{purchase}}(\mathcal{M}_{\text{mean}}^m)}{\partial \mu_{\text{impulse}}^0} d\mu_{\text{impulse}}^0 \quad (7)$$

The integrand in Equation (7) is the change in response μ_{purchase}^m due to an infinitesimal change in impulse μ_{impulse}^0 . It can be written for arbitrarily long activity sequences using the chain rule as:

$$\frac{d(\mu_{\text{purchase}}^m)}{\sum_{l=1}^{\min(p,m-1)} \sum_{j=1}^J \frac{\partial m_{\text{purchase}}(\mathcal{M}_{\text{mean}}^m)}{\partial \mu_j^{m-l}} \sum_{l'=1}^{\min(p,m-l-1)} \sum_{j'=1}^J \frac{\partial m_j(\mathcal{M}_{\text{mean}}^{m-l})}{\partial \mu_{j'}^{m-l-l'}} \dots \sum_{l''=1}^{\min(p,1)} \sum_{j''=1}^J \frac{\partial m_{j''}(\mathcal{M}_{\text{mean}}^2)}{\partial \mu_{j''}^{2-l''}} \frac{\partial \mu_{j''}^{2-l''}}{\partial \mu_{\text{impulse}}^0} d\mu_{\text{impulse}}^0 \quad (8)$$

Each summand in Equation (8) corresponds to a single path from impulse to maximal purchase. Each factor

in each summand is a partial derivative, such as $\frac{\partial m_{purchase}(\mathcal{M}_{mean}^m)}{\partial \mu_j^{m-l}}$, that captures the change to purchase in the m 'th time period per unit change in the j 'th endogenous variable, μ_j , in the $m - l$ 'th time period after impulse. Thus, the summand corresponding to each path such as:

$$\frac{\partial m(\mathcal{M}_{mean}^m)}{\partial \mu_j^{m-1}} \frac{\partial m_j(\mathcal{M}_{mean}^{m-1})}{\partial \mu_{j'}^{m-2}} \cdots \frac{\partial m_{j''}(\mathcal{M}_{mean}^2)}{\partial \mu_{j'''}^1} \frac{\partial \mu_{j'''}^1}{\partial \mu_{impulse}^0} d\mu_{impulse}^0 \quad (9)$$

captures the change in $\mu_{purchase}^m$ due to change in the previous endogenous variable in the sequence, μ_j^{m-1} , that is caused by the changes in variable before that, $\mu_{j'}^{m-2}$, and so on up to the initial impulse, $\mu_{impulse}^0$. The summand therefore captures the total increase in $\mu_{purchase}^m$ that is caused through changes in this particular sequence of activities resulting from the marketing impulse $\mu_{impulse}^0$ ². From the form of Equation (8) we can compute the weight of each path as a portion contributed to the maximal response. This allows us to identify and focus on the most important paths taken by the consumers. The combinations of the top few sequences starting from the impulse leading to maximal response would explain most of the maximal purchase. Jointly these few sequences would capture the consumers' characteristic path behavior. We provide an illustrative example of extracting path to purchase in Section 4.2.

3.4 Market Segmentation Using a Mixture Model

We embed the GMAR in a mixture model to uncover the heterogeneity in shopping behavior. The underlying data generating process for the mixture of GMAR is as follows. First, each consumer is drawn from one of K clusters according to a categorical distribution. Second, the time series data for the consumer is generated from her p initial activities according to the GMAR model of the assigned cluster, where p is the order of the GMAR model. The Bayesian Network of the model is shown in Figure 3. We call it the Clustered Generalized Multivariate Autoregressive (CGMAR) model.

² The expression in Equation (9) corresponds to a path in an autoregressive model of order 1. But, the discussion readily applies to any arbitrary model of order p .

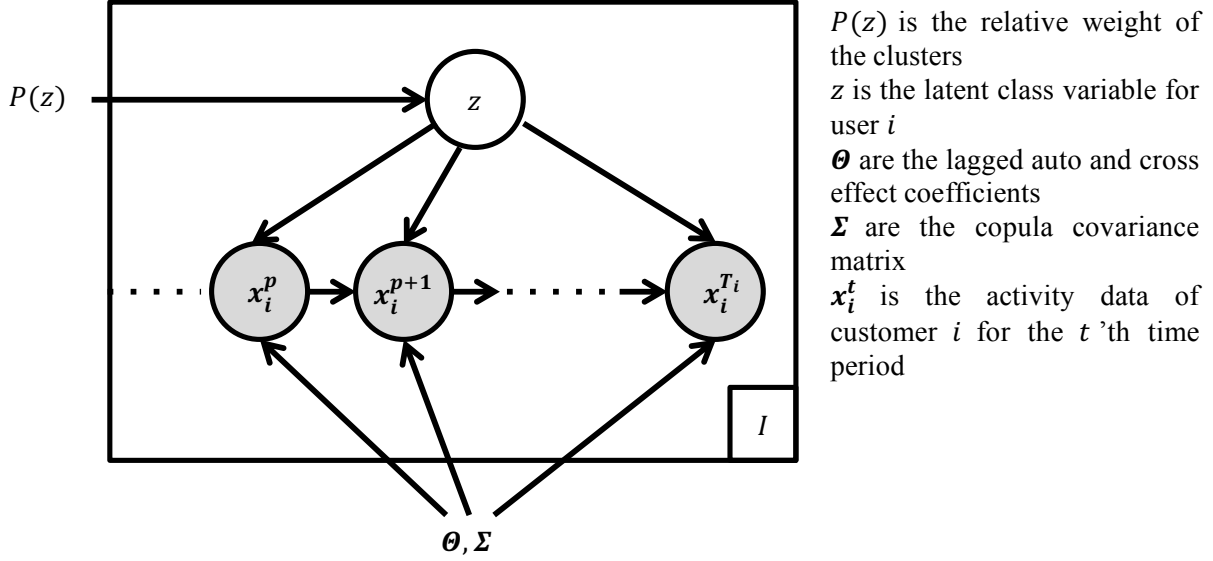


Figure 3. The Bayesian network of mixture of GMAR model.

The log likelihood of the CGMAR is:

$$\sum_{i=1}^I \log \left\{ \sum_{k=1}^K P(z_k) P(\mathbf{X}_i | {}_k\theta, {}_k\Sigma) \right\} \quad (10)$$

where ${}_k\theta$ and ${}_k\Sigma$ are the parameters of the GMAR model for cluster k . $P(z_k)$ is the probability of a random user belonging to cluster k . $P(\mathbf{X}_i | {}_k\theta, {}_k\Sigma)$ is the probability of generating time series data \mathbf{X}_i based on parameter ${}_k\theta$ and ${}_k\Sigma$, which are defined in Eq. (5). Because the sum is within the log function in Eq. (10), it is not possible to estimate the parameters (${}_k\theta, {}_k\Sigma$) separately for each cluster. We develop an algorithm under the Expectation Maximization framework to estimate these parameters. This has two iterative steps. In the **E-step**, we calculate the membership of each consumer to each cluster using the current estimate of the parameters:

$$r_{ik} = P(z_k = 1 | \mathbf{X}_i) = \frac{P(z_k) P(\mathbf{X}_i | z_k = 1)}{\sum_{k'=1}^K P(z_{k'}) P(\mathbf{X}_i | z_{k'} = 1)} = \frac{P(z_k) P(\mathbf{X}_i | {}_k\theta, {}_k\Sigma)}{\sum_{k'=1}^K P(z_{k'}) P(\mathbf{X}_i | {}_{k'}\theta, {}_{k'}\Sigma)} \quad (11)$$

where r_{ik} represents the probability of assigning consumer i to cluster k . In the **M-step**, we estimate the values of the cluster specific parameters (${}_k\theta, {}_k\Sigma$) that maximize the following expectation of the complete data log likelihood.

$$\sum_{i=1}^I \sum_{k=1}^K r_{ik} \log (f(\mathbf{X}_i; {}_k\theta)) + \sum_{i=1}^I \sum_{k=1}^K r_{ik} \log (c(\mathbf{X}_i^*; {}_k\Sigma)) + \sum_{i=1}^I \sum_{k=1}^K r_{ik} \log P(z_k) \quad (12)$$

where, \mathbf{X}_i^* is the entire set of Normal quantiles of PITs for the user i . $f(\mathbf{X}_i; \mathbf{\theta}_k)$, the probability of the user i 's data according to the marginal distributions of the time series components under the cluster k , is given as:

$$f(\mathbf{X}_i; \mathbf{\theta}_k) = \prod_{t=p+1}^{T_i} \prod_{j=1}^J f(x_{ij}^t | \mathbf{M}_i^t; \mathbf{\theta}_k) \quad (13)$$

$\mathbf{\theta}_k$ can be estimated by maximizing $\sum_{i=1}^I r_{ik} \log (f(\mathbf{X}_i; \mathbf{\theta}_k))$ from Eq. (12) for each k , which is equivalent to estimating K separate weighted GLM with r_{ik} as the weight. $\mathbf{\Sigma}_k$ can be estimated via maximizing

$\sum_{i=1}^I r_{ik} \log (c(\mathbf{X}_i^*, \mathbf{\Sigma}_k))$ separately for each k , which is equivalent to solving the Gaussian Mixture Model.³

According to Bishop (2006), the maximum likelihood estimation of covariance matrix for the Gaussian Mixture Model is:

$$\widehat{\mathbf{\Sigma}}_k = \frac{\sum_{i=1}^I r_{ik} \sum_{t=p+1}^T \mathbf{x}_i^{*t} \mathbf{x}_i^{*t'}}{\sum_{i=1}^I r_{ik} \times (T - p)} \quad (14)$$

Finally, the third summand $\sum_{i=1}^I \sum_{k=1}^K r_{ik} \log P(z_k)$ is maximized by $\hat{P}(z_k) = \sum_{i=1}^I r_{ik} / I$. This is the relative weight of each cluster. The EM algorithm executes the E- and M-steps iteratively until the parameters converge. It has two outcomes: assignment of consumers to the clusters, r_{ik} , or a segmentation, and the parameters $\mathbf{\theta}_k, \mathbf{\Sigma}_k$ for each cluster that will be used to obtain the path to purchase for each cluster. An advantage of estimating the $\mathbf{\theta}_k$ and $\mathbf{\Sigma}_k$ at the group level in an endogenous manner, as opposed to estimating the parameters separately for each consumer and then clustering them based on the similarity between the parameters, is the ability to use more data in estimating the parameters. In maximizing the first summand of Eq. (12) we estimate the parameters of each cluster using the data from *all* the consumers—each consumer gets a weight equal to the probability of belonging to a cluster. Although the cluster level parameters are estimated from the data on all the consumers, as each consumer's degree of membership to the clusters is unique, we capture their unique activity-to-activity progression as a combination of typical

³ \mathbf{X}_i^* is a function of $\mathbf{\theta}$. Hence the second summand is not strictly separable from the first. However, Patton (2006) has shown that a 2-step procedure that separately estimates $\mathbf{\theta}$ and $\mathbf{\Sigma}_k$ leads to consistent estimates of the parameters. The first step estimates the coefficient $\mathbf{\theta}$ only on the basis of the marginal data, or the first summand in Eq. (7). In the second step the estimated $\mathbf{\theta}$ is used to compute the \mathbf{x}_i^{*t} , which is used to estimate $\mathbf{\Sigma}_k$. It is same as the process we follow in each M-step.

behavior of consumers in K clusters. Thus, the algorithm mitigates the sparsity problem while capturing the uniqueness of each consumer's path.⁴

4 Empirical Analysis

After verifying from a simulation study that the proposed approach could effectively recover the cluster membership (r_{ik}) and the parameters for each cluster (${}_k\boldsymbol{\theta}, {}_k\boldsymbol{\Sigma}$), we apply it to identify the paths to purchase of consumers at a multi-channel specialty retailer selling women's clothing, accessories, and home items. The retailer sells the products only through its offline and online stores. Within its brand customers vary in their channel usage, their responses to emails and catalogs, and the nature of online research they conduct prior to purchase. These features make the setting a prime case for identifying consumers' paths to purchase.

The retailer does catalog and promotional email campaigns, however it does little mass media advertising and price promotions. Catalogs are mailed to customers approximately once a month. Emails are targeted to customers based on their purchase behavior and, occasionally, to promote new products. Unlike some online retailers (e.g. Newegg and TigerDirect), the retailer does not use other consumer activities, such as browsing, searching, or carting behavior, to target promotional emails. This setting therefore presents an opportunity for improving marketing communication strategy incorporating other evidences from the customers' path to purchase.

4.1 Data Description

We collect a detailed customer activity and transactional dataset from the retailer through Wharton Customer Analytics Initiative (WCAI) spanning a two-year period from 7/1/2010 to 6/30/2012. The dataset used in this study includes 9,805 active consumers, i.e., consumers with at least five purchases, from the largest brand of the retailer. We have information at the consumer level on emails and catalogs received, online searches, product browses, and online and offline purchases with the date of the event. Note that more than one of the aforementioned events can occur for a consumer in a given day. We also have several consumer characteristics such as age, gender, tenure with the brand, distance from the nearest store, and store loyalty card membership. As this is a specialty retailer the retailer sells products only through its exclusive

⁴ We illustrate these advantages of the mixture model through a simulation study. The details are given in Appendix B.

channels and not through popular stores such as Amazon or Walmart. So, when one of the observed customers buys a product by this retailer within the observation period, we observe the purchase and the relevant pre-purchase activities as part of our dataset. This reduces the concern that the paths taken by the customers could be taking them out of our scope of observation.

We use the weekly time-series data for each consumer to construct four variables from each consumer’s activities: number of products browsed, number of keyword searches performed, number of online purchases and offline purchases in each week.⁵ In addition, we have the number of emails and the number of catalogs sent to a consumer in a week. We also created a dummy variable indicating the weeks that include one of the ten major holidays to capture the increased shopping during holidays.⁶ Table 2 provides the variable definitions and summary statistics. We can see that the variables are quite sparse. For all consumer-initiated behaviors the 3rd quartiles of the variables are 0—indicative of zero inflation.

[Insert Table 2 about here]

4.2 A Specific Implementation of CGMAR using Zero-Inflated Poisson Regression

We consider paths of customers through the activities outlined in Table 2. These variables are counts with an abundance of zeros, rendering a standard VAR model unsuitable. Accordingly, as a specific implementation of the CGMAR described in Section 3.2, we develop a model to accommodate discreteness, sparseness and both auto and cross-correlation in the count data called Zero Inflated Multivariate Autoregressive Poisson model (ZMAP). It has three parts:

(1) Zero-Inflated-Poisson (ZIP) for the marginal distributions of the components to explicitly accommodate excessive occurrence of zeros (Lambert 1992).⁷ The probability of generating x_{ij}^t , the j 'th endogenous variable for i 'th consumer at t 'th period, is

$$f_{ZIP}(x_{ij}^t) = \begin{cases} \rho_{ij}^t + (1 - \rho_{ij}^t)Poisson(0|\lambda_j^t), & x_{ij}^t = 0 \\ (1 - \rho_{ij}^t)Poisson(x_{ij}^t|\lambda_j^t), & x_{ij}^t > 0 \end{cases} \quad (15)$$

⁵ We use the weekly activities instead of daily to reduce the sparsity problem and improve parameter estimation.

⁶ The ten major holidays we include are: New Year, Martin Luther King Day, President’s Day, Memorial Weekend, 4th of July, Labor Day, Columbus Day, Veterans Day, Thanksgiving, and Christmas.

⁷ Although Negative Binomial distributions are often used when dealing with count data with many zeros we find that a special treatment of these zeros improves our ability to estimate the model and the model fit.

where $\rho_{ij}^t \in [0,1]$ is the mixture probability of the zero-inflated model and λ_{ij}^t is the mean of the Poisson distribution for x_{ij}^t .

(2) Autoregression for the occurrence of the zero values and the mean of the Poisson distributions to capture own and cross variable lagged effects. Specifically,

$$\begin{aligned} \text{Log}(\lambda_{ij}^t) &= B_j \mathbf{M}_i^t, \text{ and} \\ \text{Logit}(\rho_{ij}^t) &= G_j \mathbf{M}_i^t \end{aligned} \quad (16)$$

where \mathbf{M}_i^t is the data vector containing the lagged endogenous and exogenous variables as defined in Section 3.2.1. The parameters $\boldsymbol{\theta} = (\{B_j\}, \{G_j\})$ capture the relationship of endogenous variable \mathbf{x}_i^t with its lag $\{\mathbf{x}_i^{t-1}, \mathbf{x}_i^{t-2}, \dots, \mathbf{x}_i^{t-p}\}$ and exogenous variable vector \mathbf{y}_i^t .

(3) Multivariate Normal Copula to connect the marginal distributions of endogenous variables and capture the contemporaneous correlations between them that are not captured by the autoregressive specification in Equations (15) and (16). Note that for the particular implementation, the $F(\cdot)$ in Eq. (3) is the CDF of the ZIP distribution. Since the endogenous variables are integers, we follow Heinen and Rengifo (2007)'s continued extension of the discrete random variables to obtain the PITs for the Zero Inflated Poisson model that are uniform random variables in $[0,1]$ as required of the copulas (See Appendix C.1). Diagnostics of the copula model, e.g., correlation between the Normal Quantiles of the PITs and histograms of the PITs, are shown in Appendix C.2.

The above model is incorporated into a mixture model and estimated using the EM algorithm described in Section 3.4 to obtain segments of the consumers with distinct path to purchase. We name this Clustered Zero-Inflated Multivariate Autoregressive Poisson (CZMAP) model.

4.3 Findings on Path to Purchase Segmentation

Before applying the CZMAP model we need to decide which of the weekly variables are endogenous and which are exogenous. Using the ZMAP model in Granger Causality test (Granger 1969), the consumer activities (including browse, search, online purchase, offline purchase and marketing email events) are treated as endogenous variables. Catalogs and holiday dummy are treated as exogenous variables. Treating the emails sent to the consumer, as an endogenous variable might seem surprising at first, but makes sense

for the following reason. Many retailers, including our industry partner, send promotional emails to customers based on their recent purchase activities. In contrast, the schedule of mailing of printed catalogs does not change based on day to day activities of individual consumers. Therefore, they appear as exogenous variables from the Granger causality test. Following Zantedeschi et al. (2013) we show the distribution of number of emails/catalogs each customer gets in a week to provide the visual clue to the categorization (Appendix A).

The number of segments is set to 5 and the order of the autoregressive part is set to 1 for the CZMAP model based on Bayesian information criterion (BIC). Table 3 compares the CZMAP model and a set of alternative models without zero-inflation and Normal copula features. We also fit the VAR Mixture model, suggested in Fong et al. (2007), which is a closely related approach available in the literature. The proposed CZMAP model (Model 4) has the highest log-likelihood and the lowest BIC in the compared set, suggesting that it best describes how the data are generated among the compared approaches.

[Insert Table 3 about here]

4.3.1 Path to Purchase Construction

We focus on the paths of the consumers through the activity space that start from a marketing impulse such as email or catalog and end with a maximal online or offline purchase response. To identify the period with maximal purchase response we calculate the purchase response based on Equation 6. The maximal and cumulative purchase responses to the impulse of one email or catalog are shown in Table 4. For example, in the case of Segment 3, the row labeled Catalog-Offline in Table 4 reports that an additional catalog in the first week results in a 0.0712 peak increase in offline purchase in the third week—15% of the average weekly offline purchase for the segment. The cumulative increase in offline purchase due to the additional catalog received by the customer across all weeks is 0.1668.

[Insert Table 4 about here]

For illustration purposes, we next work through the derivation of characteristic paths between receiving a catalog and the maximal response to it in offline purchase for segment 3. We denote the maximum response by $\Delta\mu_{offline}^3$ where 3 is the indicator for time and *offline* is the index of the variable for offline purchase.

We can decompose $\Delta\mu_{offline}^3$ as the sum of weight of different paths. Here $\mu_{offline}^3 = m_{offline}(\mathcal{M}_{mean}^3)$, where, $m_{offline}(\cdot)$ is the conditional mean function of the offline sales variable under the ZMAP model. As described in Section 3.3, $\mathcal{M}_{mean}^3 = \{\mu_{browse}^2, \mu_{search}^2, \mu_{online}^2, \mu_{offline}^2, \mu_{email}^2\}'$ since lag is 1. Because $\Delta\mu_{offline}^3$ is the result of increasing $y_{catalog}^1$ from zero to one, we have, using the chain rule of derivative:

$$\Delta\mu_{offline}^3 = \int_0^1 d\mu_{offline}^3 = \int_0^1 \sum_{j=1}^5 \frac{\partial m_{offline}(\mathcal{M}_{mean}^3)}{\partial \mu_j^2} \sum_{j'=1}^5 \frac{\partial m_j(\mathcal{M}_{mean}^2)}{\partial \mu_{j'}^1} \frac{\partial m_{j'}(\mathcal{M}_{mean}^1)}{\partial y_{catalog}^1} dy_{catalog}^1 \quad (17)$$

Based on above equation, $\Delta\mu_{offline}^3$ is the sum of 25 terms and each of them corresponds to a single path that starts from catalog and ends with the offline node in the third period in Figure 4.

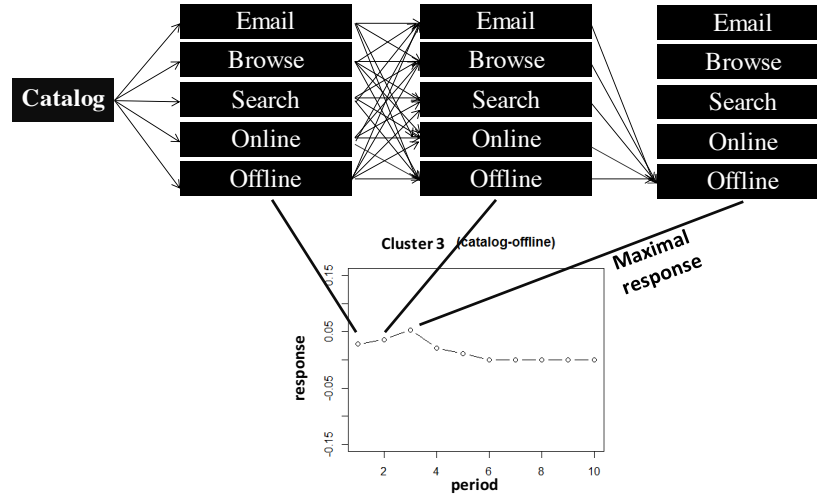


Figure 4. Example of a catalog leading to maximal offline purchase response.

For example, the weight of a single path such as “catalog—browse—search—offline_purchase” is:

$$\int_0^1 \frac{\partial m_{offline}(\mathcal{M}_{mean}^3)}{\partial \mu_{search}^2} \frac{\partial m_{search}(\mathcal{M}_{mean}^2)}{\partial \mu_{browse}^1} \frac{\partial m_{browse}(\mathcal{M}_{mean}^1)}{\partial y_{catalog}^1} dy_{catalog}^1 \quad (18)$$

Appendix C.3 shows how the mean function $m_j(\cdot)$ for the j 'th component and its derivative could be computed for discrete marginal variables. Following the above procedure, we calculate the weights of the paths for each segment. We list the paths in Table 5. If there are multiple paths, we rank them in the descending order of their weights and show only the top three. A blank path indicates that the maximal purchase response to the impulse is not significant (at p-value of 0.01), i.e. there is no significant purchase response for the segment to draw a path to.

[Insert Table 5 about here]

To understand how to interpret each path, let us consider the path “catalog—browse—search—offline purchase” for the segment 3.⁸ It indicates that when a consumer receives a catalog impulse in the first week, the customer browses more in the same week which leads to more searches for information in the second week and finally purchases at an offline store in the third week. The weight of the path is 0.0192, i.e. receiving one catalog in the first week leads to a 0.0192 increase in offline purchases in the third week via the path. Based on the maximal and cumulative response shown in Table 4, the path contributes 27% (or $0.0192/0.1137$) of the maximal increase in offline purchase, or 11.5% (or $0.0192/0.1668$) of the cumulative increase offline purchase as a response to a single catalog. Comparing this to the second most important path, “catalog—search—search—offline purchase,” we see that expected occurrences of both browse and search increase when the consumer receives a catalog. However, the expected increase in search in the second period due to the increase in catalog in the first period is higher than that due to the increase in the search in the first period. The larger increase in expected search volumes in the second period in the first path leads to a larger increase in expected purchases at the end of the first path. The expected increase in purchase at the end of first path is 27% of the maximal purchase; where as the expected increase in purchase at the end of second path is 21% of the maximal purchase.

4.4 Descriptions of the Path to Purchase Segments

We discuss the characteristics of the five segments, each of which has a different set of characteristic paths to purchase. These paths are shown in Figure 5. The demographics and average behavioral characteristics of the five segments are shown in Table 7 as an aid to their interpretation.

⁸Since catalog is an exogenous variable, its effects are observed starting from the same period in which it is received. However, email is treated as an endogenous variable and since the endogenous variables are, its initial effects are observed in the period after the email was received.

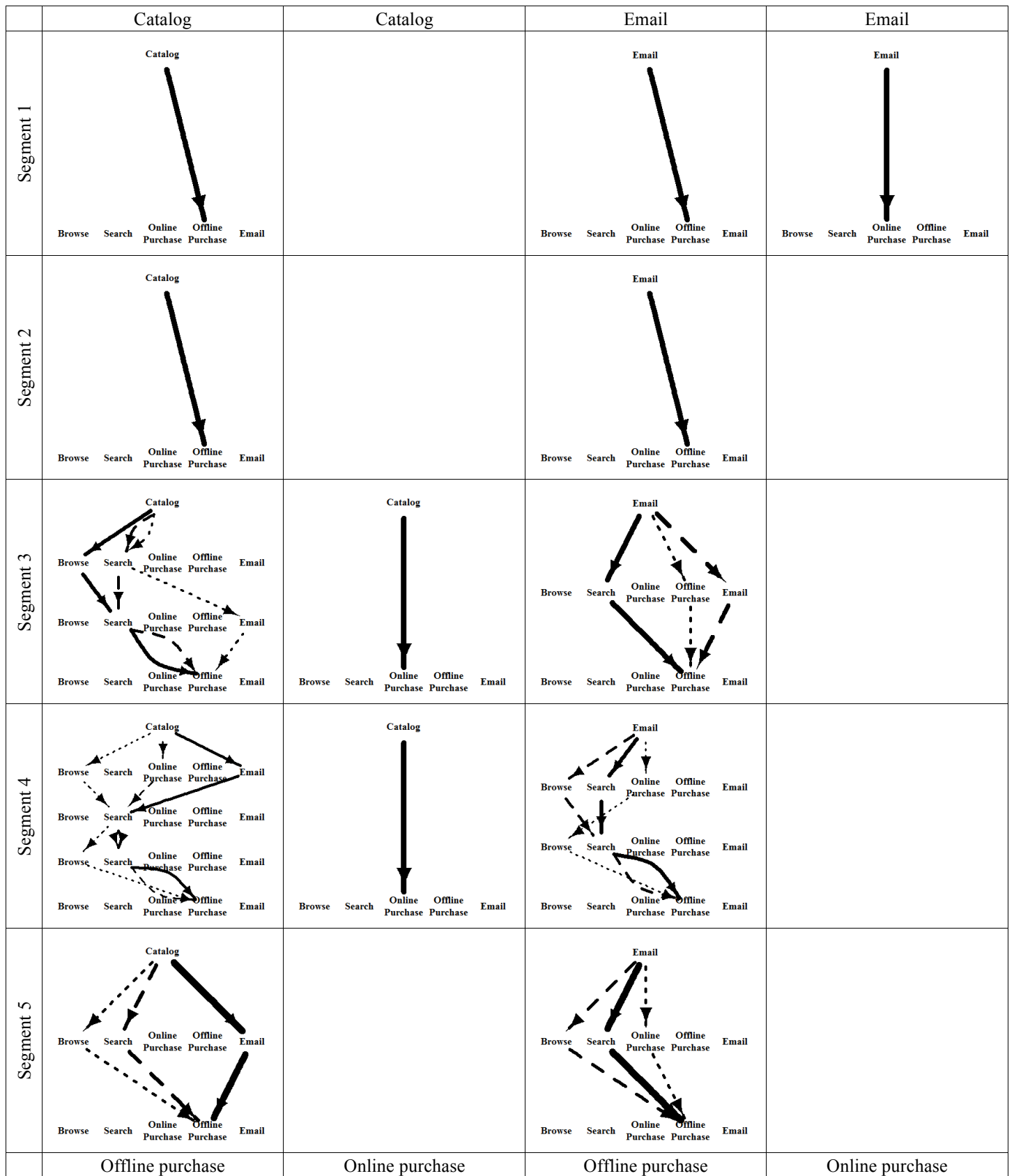


Figure 5. Visualization of the characteristic paths for each segment. Each row corresponds to a segment. Each column corresponds to a marketing impulse and purchase response pair between which the paths are drawn. Width of each path is proportional to the fraction of the increase in purchase attributed to the path. Empty cells represent that there was no significant purchase response for the impulse; hence there was no path.

It is interesting to note that different sequences of activities explain the eventual purchase to a varying degree. The weights of the paths in Figure 5 show their relative contribution. The percentages of the maximal purchase response explained by each possible path are shown in Table 6. The number of possible paths increases exponentially with the time to maximal purchase response. However, as we can see from the histograms in Table 6, in most cases majority of the purchases can be attributed to a small number of paths, indicating that the consumers in a segment have a small number of characteristic paths from marketing impulse to eventual purchase. As shown in Figure 5 the combination of these small number of paths capture the simultaneous activities the consumers are observed to engage in between the impulse and the maximal purchase response. Next we examine the characteristic paths in each of the five identified segments.

[Insert Table 6 about here]

Loyal and Engaged Shoppers. This group of customers' (Segment 4) purchases is the highest among all segments. The proportion of consumers having the store loyalty card is also the highest among all segments. Based on the date of the first contact with the brand, these consumers' tenure with the brand is among the longest. Interestingly, they have the longest path among all groups of three/four weeks to offline purchase. After receiving an email or a catalog, these customers take several weeks to browse and search online and sometimes make a purchase online before going to a store to make the maximal purchase. Their average distance to the nearest physical store is the largest at 15.2 miles, which might explain why they take much longer than the other groups to purchase in a store. In contrast to offline purchases, the duration of their path to maximal online purchase is only one week. The responses to emails and catalogs are also the highest for these consumers (Table 4). As we can see from Table 6 this segment of consumers have 25 and 125 possible sequences of activities leading to maximal offline purchases starting from Email and Catalogs respectively. However, the top three paths explain 64% of the maximal purchase as a response to emails and 57% of the maximal purchase in response to the catalogs.

Digitally Driven Offline Shoppers. These consumers (Segment 3) exhibit several interesting paths leading to offline purchases. Two of the longer paths that start with a catalog and end with an offline purchase are “*Catalog(1)-browse(1)-search(2)-offline(3)*” and “*Catalog(1)-search(1)-email(2)-offline(3)*”. In the first

path, the after receiving the catalog we see that this group of consumers do higher level of browse and search in the first and second week respectively before doing most of their purchases in the third. The second path captures the phenomenon where a consumer, after receiving the catalog, is subsequently encouraged by an email message in the second week to move forward to purchase. From Table 5, we see that this path, shaped by the combination of marketing campaigns, lead to 15% of the maximal purchases of these consumers.

Holiday Shoppers. For these customers (Segment 5), 46.2% of their online purchases and 49.1% of their offline purchases occur during the holidays. From Table 7 we see that these consumers are the oldest of the five groups, yet are among the newest to the brand. Although there are significant paths between catalog/email and offline purchase, their weights are small. Overall, the cumulative increase in purchases as a response to emails and catalogs are the smallest of all segments indicating that these customers are not as sensitive to email and catalog communications and respond to holidays instead.

Infrequent Offline Shoppers. For this group (Segment 1) the number of marketing communications received is lower than other segments and the average number of purchases is small. These customers are among the least active in terms of purchases and are the least marketed to. They live close to the physical stores and we find that most of their purchases are done offline. The length of path to offline purchase is zero for these customers suggesting that the maximal response occurs in the same week that the email or catalog is received. We do not observe any significant online purchase response to marketing communication for this segment.

Frequently Targeted Occasional Shoppers. This group (Segment 2) has many similarities with Segment 1. They live relatively close to the store and are not active shoppers. Their only response to email or catalog is observed in the in-store purchases. Based on the cumulative impulse results from Table 4, they are the one of the least sensitive to emails and catalogs—only second to the Holiday Shoppers. They receive more catalogs and email than segment 1, however, their average online and offline purchase is lower in comparison.

[Insert Table 7 about here]

5 Managerial Applications

We next explore two applications where the proposed CZMAP model is helpful for guiding a manager's

decision. The first application predicts consumers' purchase in a future test period. The second application derives the optimal marketing mix over a planning horizon using the customers' path information via dynamic programming.

5.1 Prediction of Purchase Behavior based on Path to Purchase Segments

From a managerial perspective, it will be helpful to establish if our proposed approach to uncover path to purchase improves the forecasts of consumer purchases given the user activities and marketing activities. We therefore compare the ability of the CZMAP model to predict out-of-sample purchases by individual customers with that of a number of alternative models. We longitudinally divide the dataset into training set (first 96 weeks) and test set (last 10 weeks). We first estimate the parameters of the models using the training data. Using the estimated model we predict the number of online and offline purchases for each consumer in each week of the test period and compare it to the observed purchases. We use the observed exogenous variables in the test period for our prediction, but do not use the observed endogenous variables. Instead, we use the estimated values of the endogenous variables for predicting the endogenous variables in the subsequent periods. This allows us to evaluate the accuracy of each model in predicting the cumulative response to the exogenous shocks in the presence of the other exogenous variables. We compare the CZMAP algorithm to the following alternative approaches.

Approaches that do not take advantage of the sequence information in consumer activities

1. **The average of the training data:** For each consumer, we compute the average number of online and average number of offline purchases per week during the training period. For each consumer, this average value is used as the prediction for each week in the test period.
2. **Purchases at the last training period:** For each consumer, we record the number of online and number of offline purchases made during the last training period (96th period). These numbers for each consumer are used as the predictions for the online and offline purchases in each week of the test period.
3. **Random selection:** For each consumer, in making a prediction for each week in the test period we randomly select the online and offline purchases made in one of the weeks in the training period and use them as the prediction.

4. **Regression on RFM:** For all the user and week pairs in the second year⁹ of the training period we compute the recency (r), frequency (f), and monetary value (m) of the consumer based on her purchases. We also compute the recency and frequency of the searching and browsing activity of the consumer to provide a comparison using the same type of data available to the CZMAP approach. Then a Zero-Inflated-Poisson regression of the following form is estimated

$$\begin{aligned} \text{onlinepurchase}(i, t) \sim & r_{\text{purchase}}(i, t) + f_{\text{purchase}}(i, t) + m_{\text{purchase}}(i, t) + r_{\text{search}}(i, t) \\ & + f_{\text{search}}(i, t) + r_{\text{browse}}(i, t) + f_{\text{browse}}(i, t) \end{aligned}$$

where recency, $r_{\text{purchase}}(i, t)$ is the negative of number of days before period t the user i made a purchase. Frequency, $f_{\text{purchase}}(i, t)$, is the number of purchases the user i has made in one year before t . Recency and frequency for search and browse activities were computed in a similar manner. $m_{\text{purchase}}(i, t)$ is the dollars spent by the user i in one year before t . These variables summarize all the activities of the user that are used by CZMAP, but ignore the sequence information as done by the popular RFM metrics. We estimate a similar equation for the offline purchase. After estimating the eight coefficients for each model, including one for the intercepts, we use the explanatory variables computed in the last week of the training period to predict the weekly online and offline purchases for each user in the test period.

Approaches that take advantage of the sequence information in consumer activities

These approaches are categorized by the clustering algorithms used for segmentation and the autoregressive specifications used for the individual customer's time-series observations.

1. **No-segmentation:** We estimate the parameters of individual level models with all consumers in one group.
2. **Demographics segmentation:** We use the customer demographic variables (age, gender, tenure with the brand, distance from the nearest store, and store loyalty card membership) to create five customer segments using the k-means clustering algorithm.
3. **Segmentation by short-term response:** We follow Reimer et al. (2014) to segment customers based on their short-term response to marketing communication using a mixture of logistic regressions. The number

⁹ We focus on the second year in order to be able compute reliable estimates of recency, frequency, and monetary values using data from the preceding one year.

of clusters was set to 8 following BIC criterion.

4. **Recency-Frequency-and-Monetary value (RFM) segmentation:** We classify each customer using the quintiles of their Recency, Frequency, and Monetary value computed from their purchase activities (Hughes 2006). The sorted set of 125 segments is then divided into 5 groups again by their quintiles to create final customer segments.
5. **VAR Mixture** (Fong et al. 2007). We estimate a mixture of VAR model similar to CZMAP, where segments and segment specific parameters of the time series are simultaneously determined. The only difference is that the customer's time series are modeled using a VAR instead of a ZMAP.

We estimate *individual level* ZMAP and VAR models for each segment under the approaches 1 to 4 above. For the approach 3 (i.e. short-term response based segmentation) we also estimate a VAR model from the average level of activity for each segment as done by Reimer et al. (2014). Each of these models is then used to predict each customer's purchase activity in the test period. The Mean Absolute Errors per customer per week prediction in the test period are listed in Table 8.

[Insert Table 8 about here]

The prediction performance of the CZMAP (ZMAP with mixture model) is better than all other compared models. Among the methods that do not use the sequence information, the Zero Inflated Poisson regression using RFM scores performs best. However, when we use the sequence information by using an autoregressive component in the models the performance improves substantially, suggesting that there is considerable information in consumers' activity sequence that can be exploited to more accurately predict purchase conversion.

Among the methods that do use an autoregressive specification, the largest improvement is obtained when we move from using one model for all customers to segmenting customers and using different models for each homogenous group. Segmenting by behavioral attributes such as RFM and short-term response leads to better forecasts than segmenting based on demographic attributes. The last row in Table 8 shows that segmenting consumers endogenously based on the evolution of their behavior leads to improved predictions. And finally, within each segmentation strategy using a zero-inflated Multivariate Poisson model offers

superior predictions of individual customer's purchase behavior in the test period (see column titled "ZMAP" in Table 8).

5.2 Resource Allocation based on Path to Purchase Segments

Knowing the consumers' path to purchase, a manager can sequence the marketing messages to achieve a predefined objective. To illustrate this, we perform a policy simulation where the goal is to achieve 10% increase in sales over a planning horizon by sending emails and catalogs at the optimal time. This is often how sales targets are set in practice (Saghafi 1987) and used in the literature as a planning scenario (Roy et al. 1994, Hanssens et al. 2014). We compute the average weekly online and offline purchase for each customer during the training period. By adding 10% to the average weekly online/offline purchases in the training period, we get the targets for each customer in the testing period, i.e., period 97 to 106. The objective is to derive the optimal marketing communication sequence for each of the 5 segments to reach the target as closely as possible. We use the following squared deviations from the target as our loss function for each customer and each test period:

$$L(x_{online}^t, x_{offline}^t) = (x_{online}^{t*} - x_{online}^t)^2 + (x_{offline}^{t*} - x_{offline}^t)^2 \quad (19)$$

where x_{online}^{t*} and $x_{offline}^{t*}$ are the target values for *online and offline purchase* at period t and x_{online}^t and $x_{offline}^t$ are the realized online and offline purchase at period t . The loss over the entire planning horizon is expressed as:

$$Loss = \sum_{t=97}^{106} \theta^t [L(x_{online}^t, x_{offline}^t)] \quad (20)$$

where the discount factor θ is set to 0.9 following Hanssens et al. (2014). After obtaining the CZMAP parameters using the training data, we find the optimal marketing sequence for each segment using the estimated parameters for that segment via Dynamic Programming. Here we minimize the *Loss* function by choosing the number of emails and number of catalogs to send (optimal marketing decision or OMD^t) at time t . The optimization problem can be written as the following finite horizon Bellman equation:

$$V^t(x_{online}^t, x_{offline}^t) = \min_{OMD^t} \{L(x_{online}^t, x_{offline}^t) + \theta V^{t+1}(x_{online}^{t+1}, x_{offline}^{t+1})\} \quad (21)$$

The equation can be solved by backward recursion (Miranda and Fackler 2002). The pseudo-code is

provided in Appendix D.

Both email and catalog are firm controlled activities, thus are considered as decision variables for policy simulation. In order to use email as a decision variable we need to treat it as an exogenous variable in the CZMAP model. Applying Sequential Bayesian Cut theory we can do so and still get unbiased estimates of the parameters (Bauwens et al. 2000, Zantedeschi et al. 2013). Emails are endogenous variables based on the Granger causality test (Section 4.1). This is consistent with the firm policy of sending some of the emails to the customers based on their recent purchase activity. In our simulations using both email and catalog as firm controlled exogenous variables leads to a lower loss than using catalog alone, suggesting that there is a better way to schedule these emails than the current process if the objective is to reach 10% increase in sales.

The catalog and email decisions pair $[y_{email}^t, y_{catalog}^t]$ in each period is defined in a feasible set as $y_{catalog}^t \in [0,1], y_{email}^t \in [0,3]$. The state space consists of four consumer behaviors $[x_{browse}^t, x_{search}^t, x_{online}^t, x_{offline}^t]$. Each state variable is discretized into 10 intervals for using backward recursion.

Figure 6 displays the optimal marketing mix sequence over 10 periods to achieve the targeted +10% sales for the five segments and for the scenario where no clustering was done. Note that despite the same improvement in sales target, the optimal marketing policies vary across segments. For example for the loyal and engaged shoppers (i.e. Segment 4), which has one of the most extended path to purchase, the optimal policy involves sending a total of 13 emails and 4 catalogs at regular intervals interspersed by one or two periods where no catalog is sent. Since the effect of emails and catalogs occur over several time periods according to the consumers' paths, the optimal policy would involve reducing or not sending marketing communication at certain times so as not to overshoot the set target. Such knowledge of consumers' path through the activity space allows a manager to anticipate the response to a marketing communication over several future periods and adjust the next marketing stimulus to arrive as close to the sales objective of each period as possible.

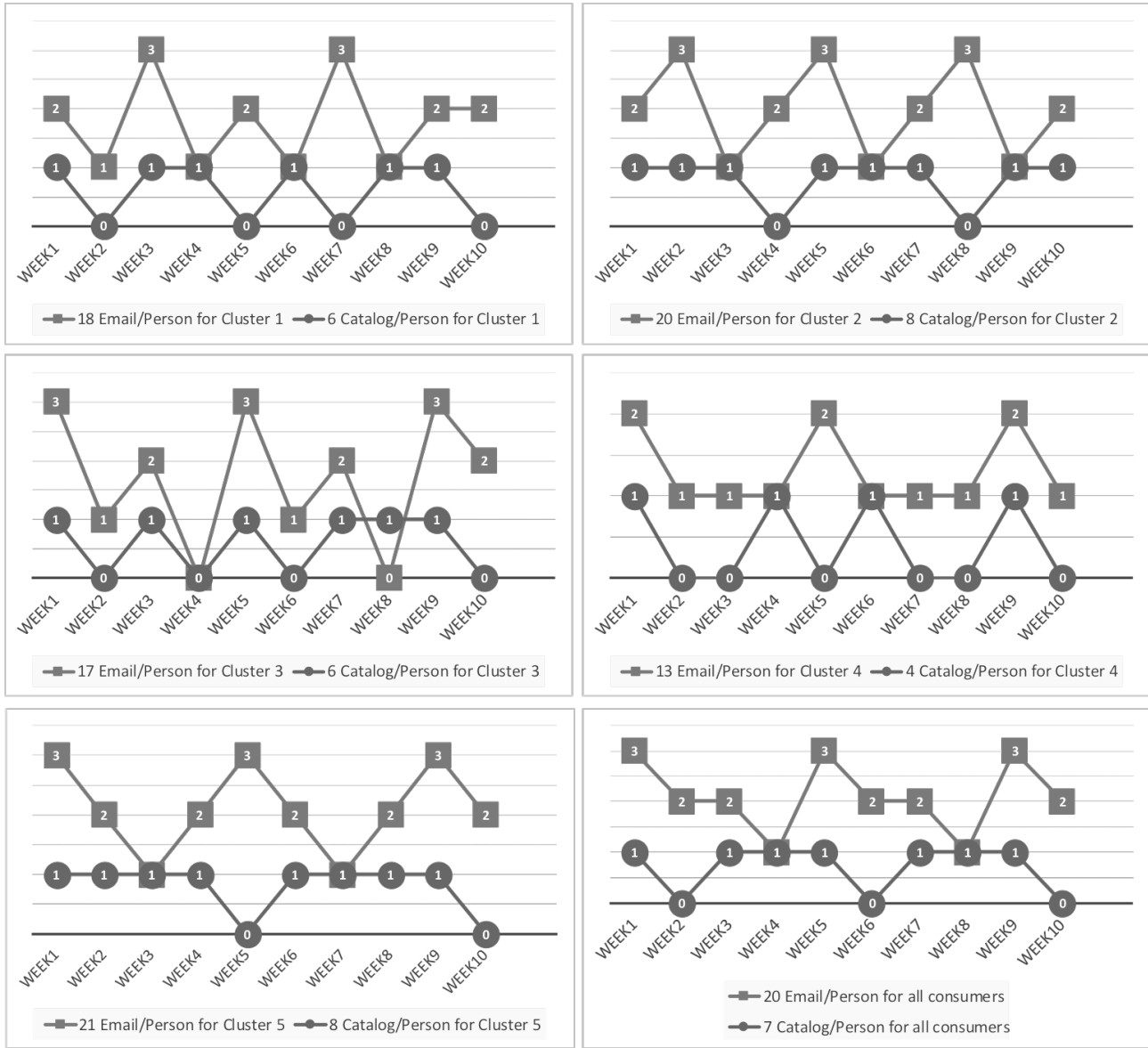


Figure 6. Optimal email and catalog policies for the five segments

In contrast, for a customer segment that is not as sensitive to marketing communications one would require more marketing activity. For example, for Segment 2, that is not as sensitive to catalogs, one would need to send 8 catalogs as compared to only 4 for Segment 4. We also find that the magnitude of the Loss relative to the per customer target for the test period for those segments are higher than the losses for segments that have more elaborate paths and are relatively more sensitive to emails and catalogs (Segment 3 and 4 in Table 9).

We compare the optimal marketing mix policy and the loss for the five segments with the scenario without segmentation. The optimal policy is shown in the last panel of Figure 6 and the loss is given in Table

10. The much larger loss (0.54) in the no segmentation scenario indicates that segmenting and designing separate marketing strategies for different groups based on their path to purchase will meet the purchase target more precisely than treating all customers as one group. Finally, we apply CVAR (Fong et al. 2007) to perform the same analysis. We find that with the paths uncovered using the CVAR the loss is higher, which can be attributed to the poor fit of the model to the individual consumer activity data.

6 Conclusion

We propose an approach called Clustered Generalized Multivariate Autoregression (CGMAR) to empirically uncover the paths to purchase through the activities of the consumers and to segment consumers based on these paths. In contrast to prior work, our method can extract paths from settings where consumers engage in multiple simultaneous and correlated activities (browsing, searching, reading emails, etc.) at each time period. Our empirical application to a customer touch-point dataset collected from a large multi-channel retailer uncovers five distinct consumer segments: loyal and engaged shoppers, digitally-driven offline shoppers, holiday shoppers, infrequent offline shoppers, and frequently targeted occasional shoppers.

The characteristic paths to purchase of the segments provide several key insights into the shopping behavior of the customers at this particular retailer. First, we find that the “loyal and engaged” shopper segment has long paths to purchase involving considerable online research. After receiving an email or a catalog, these customers take several weeks to browse and search online and sometimes make a purchase online before going to a physical store to make the maximal purchase. This suggests that the retailer must focus on keeping these customers engaged with online content in the weeks following the mailing of catalogs. Second, the “digitally driven offline” shoppers’ segment exhibits several interesting paths to offline purchase involving considerable online activity. Overall, the combination of retailer marketing campaigns and digital online activity characterizes these consumers path to purchase. Third, the segment of “holiday shoppers” are not as sensitive to email and catalog communications but are most responsive to holidays, providing recommendations to the retailer for holiday oriented marketing programs for this segment. Promotional emails occur in the path to offline purchase for more than one segment of customers, suggesting that it is a means for retailers to proactively shape customers’ path to purchase. Most importantly, the study’s

insights reveal that there is considerable heterogeneity in the path to purchase behavior that needs to be considered in developing an understanding of shopping behavior at the individual level.

Beyond the insight into customers' behavior, we provide two managerial applications. We show that our proposed Clustered Zero-inflated Multivariate Poisson (CZMAP) model is better able to predict individual customer's purchases in future test periods. Finally, through policy simulations we show that the knowledge of the paths to purchase of the customer segments can help a manager in designing optimal marketing mix over time for each segment of customers to achieve a predefined sales target.

The proposed general methodology can be applied to other settings where entities are observed and measured over time along multiple dimensions. If an interesting path can be conceptually defined through these dimensions, the CGMAR approach can be applied to cluster the entities based on their progress over time and extract significant paths between an impulse and a response variable of interest. One such setting is in healthcare where our approach can be applied to medical records that contain patients' vital statistics and hospital visits over time to uncover the paths of different groups of patients from a particular intervention (e.g., a consultation or a treatment) to a health outcome of interest through the space observed attributes about the patients.

The CGMAR approach can be extended in a number of ways. Future research should incorporate non-linearity to model saturation in response to exogenous variables for deriving optimal marketing sequences. Interactions between exogenous and endogenous variables could also be incorporated to model the state dependent response to exogenous impulses. One can also relate the uncovered paths to purchase to the demographics of consumers to understand the role that demographics play in explaining their shopping activities. Finally, a consumer's path through endogenously identified states that govern her shopping activities can also be modeled using a hidden Markov model. Consumers may be segmented based on their transitions through the hidden states using a mixture of HMMs. This will allow us to empirically identify paths through what resemble cognitive states of shoppers.

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.
- Barbiero, A. and P. A. Ferrari (2015). "Simulation of Correlated Poisson Variables." Applied Stochastic Models in Business and Industry **31**(5): 669-680.
- Barry, T. E. (1987). "The Development of the Hierarchy of Effects: An Historical Perspective." Current issues and Research in Advertising **10**(1-2): 251-295.
- Bauwens, L., M. Lubrano and J.-F. Richard (2000). Bayesian Inference in Dynamic Econometric Models, Oxford University Press.
- Bishop, C. M. (2006). Pattern Recognition and Machine Learning, springer New York.
- Brown, M. (2013). "Navigating the New Path to Purchase." Retrieved 5/18, 2015, from [https://www.millwardbrown.com/docs/default-source/global-solutions-downloads/channel-optimization/Navigating the New Path to Purchase MBD.pdf](https://www.millwardbrown.com/docs/default-source/global-solutions-downloads/channel-optimization/Navigating%20the%20New%20Path%20to%20Purchase%20MBD.pdf).
- Bucklin, R., J. Lattin, A. Ansari, S. Gupta, D. Bell, E. Coupey, J. C. Little, C. Mela, A. Montgomery and J. Steckel (2002). "Choice and the Internet: From Clickstream to Research Stream." Marketing Letters **13**(3): 245-258.
- Bucklin, R. E., S. Gupta and S. Siddarth (1998). "Determining Segmentation in Sales Response across Consumer Purchase Behaviors." Journal of Marketing Research **35**(2): 189-197.
- Bucklin, R. E. and C. Sismeiro (2003). "A Model of Web Site Browsing Behavior Estimated on Clickstream Data." Journal of Marketing Research **40**(3): 249-267.
- Court, D., D. Elzinga, S. Mulder and O. J. Vetvik (2009). "The Consumer Decision Journey." McKinsey Quarterly.
- Deb, P. and P. K. Trivedi (1997). "Demand for Medical Care by the Elderly: A Finite Mixture Approach." Journal of Applied Econometrics **12**(3): 313-336.
- Dewan, S. and J. Ramaprasad (2014). "Social Media, Traditional Media, and Music Sales." MIS Q. **38**(1): 101-122.
- Farley, J. U. and L. W. Ring (1966). "A Stochastic Model of Supermarket Traffic Flow." Operations Research **14**(4): 555-567.
- Fong, P., W. Li, C. Yau and C. Wong (2007). "On a Mixture Vector Autoregressive Model." Canadian Journal of Statistics **35**(1): 135-150.
- Gefen, D., E. Karahanna and D. W. Straub (2003). "Trust and Tam in Online Shopping: An Integrated Model." MIS Q. **27**(1): 51-90.
- Ghose, A. and V. Todri (2015). "Towards a Digital Attribution Model: Measuring the Impact of Display Advertising on Online Consumer Behavior." Social Science Research Network Working Paper Series.
- Google. (2014). "The Customer Journey to Online Purchase." Retrieved 5/18, 2015, from <https://www.thinkwithgoogle.com/tools/customer-journey-to-online-purchase.html>.
- Granger, C. W. (1969). "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods." Econometrica: Journal of the Econometric Society: 424-438.
- Hanssens, D. M., K. H. Pauwels, S. Srinivasan, M. Vanhuele and G. Yildirim (2014). "Consumer Attitude Metrics for Guiding Marketing Mix Decisions." Marketing Science **33**(4): 534-550.
- Heinen, A. and E. Rengifo (2007). "Multivariate Autoregressive Modeling of Time Series Count Data Using Copulas." Journal of Empirical Finance **14**(4): 564-583.
- Heller, W. (1988). "Tracking Shoppers through the Combination Store." Progressive Grocer **67**: 47-62.
- Hughes, A. M. (2006). Strategic Database Marketing, McGraw-Hill New York: 101-129.
- Hui, S. K., E. T. Bradlow and P. S. Fader (2009). "Testing Behavioral Hypotheses Using an Integrated Model of Grocery Store Shopping Path and Purchase Behavior." Journal of consumer research **36**(3): 478-493.
- Kalpakis, K., D. Gada and V. Puttagunta (2001). Distance Measures for Effective Clustering of Arima Time-Series. Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on.
- Kimelfeld, Y. (2013) "Marketers Can Map the Consumer's Journey to Purchase." Compete Pulse.
- Lambert, D. (1992). "Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing." Technometrics **34**(1): 1-14.
- Larson, J. S., E. T. Bradlow and P. S. Fader (2005). "An Exploratory Look at Supermarket Shopping Paths." International Journal of Research in Marketing **22**(4): 395-414.
- Liao, T. W. (2005). "Clustering of Time Series Data—a Survey." Pattern Recognition **38**(11): 1857-1874.
- Lovejoy, J. (2014). "The Customer Journey: Explaining the Path to Purchase." <http://www.brandwatch.com/2014/10/path-to-purchase/> Accessed 5/18 2015.
- Lütkepohl, H. (2007). New Introduction to Multiple Time Series Analysis, Springer: 53-54.

Manning, C. D., P. Raghavan and H. Schutze (2008). Introduction to Information Retrieval, Cambridge University Press.

Miranda, M. J. and P. L. Fackler (2002). Applied Computational Economics and Finance, MIT press.

Moe, W. W. (2003). "Buying, Searching, or Browsing: Differentiating between Online Shoppers Using in-Store Navigational Clickstream." Journal of Consumer Psychology **13**(1-2): 29-39.

Montgomery, A. L., S. Li, K. Srinivasan and J. C. Liechty (2004). "Modeling Online Browsing and Path Analysis Using Clickstream Data." Marketing Science **23**(4): 579-595.

Mullen, J. J. (1956). "Reviewed Work: Effective Advertising Copy." Journal of Marketing **21**(2): 251-253.

Nelder, J. A. and R. W. M. Wedderburn (1972). "Generalized Linear Models." Journal of the Royal Statistical Society. Series A (General) **135**(3): 370-384.

Nunes, P. F., O. Schunck and R. E. Wollan (2012). "Serving the Nonstop Customer." Outlook **3**.

Oates, T., L. Firoiu and P. Cohen (2001). Using Dynamic Time Warping to Bootstrap Hmm-Based Clustering of Time Series. Sequence Learning. R. Sun and C. L. Giles, Springer Berlin Heidelberg. **1828**: 35-52.

Patton, A. J. (2006). "Estimation of Multivariate Models for Time Series of Possibly Different Lengths." Journal of Applied Econometrics **21**(2): 147-173.

Ramaswamy, V., E. W. Anderson and W. S. DeSarbo (1994). "A Disaggregate Negative Binomial Regression Procedure for Count Data Analysis." Management Science **40**(3): 405-417.

Reimer, K., O. J. Rutz and K. Pauwels (2014). "How Online Consumer Segments Differ in Long-Term Marketing Effectiveness." Journal of Interactive Marketing **28**(4): 271-284.

Roy, A., D. M. Hanssens and J. S. Raju (1994). "Competitive Pricing by a Price Leader." Management Science **40**(7): 809-823.

Saghafi, M. M. (1987). "Market Share Stability and Marketing Policy: An Axiomatic Approach." Research in Marketing **9**: 267-284.

Sims, C. A. (1980). "Macroeconomics and Reality." Econometrica **48**(1): 1-48.

Sims, C. A. and T. Zha (1999). "Error Bands for Impulse Responses." Econometrica **67**(5): 1113-1155.

Song, P. X.-K. (2000). "Multivariate Dispersion Models Generated from Gaussian Copula." Scandinavian Journal of Statistics **27**(2): 305-320.

Sorensen, H. (2003). "The Science of Shopping." Marketing Research **15**(3): 30-35.

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.

Strong Jr, E. K. (1925). "Theories of Selling." Journal of applied psychology **9**(1): 75.

Trivedi, P. K. and D. M. Zimmer (2005). "Copula Modeling: An Introduction for Practitioners." Foundations and Trends® in Econometrics **1**(1): 1-111.

Tsay, R. S. (2005). Analysis of Financial Time Series, John Wiley & Sons.

Underhill, P. (2005). Call of the Mall: The Geography of Shopping, Simon and Schuster.

Wedel, M. and W. DeSarbo (1995). "A Mixture Likelihood Approach for Generalized Linear Models." Journal of Classification **12**(1): 21-55.

Wedel, M., W. S. Desarbo, J. R. Bult and V. Ramaswamy (1993). "A Latent Class Poisson Regression Model for Heterogeneous Count Data." Journal of Applied Econometrics **8**(4): 397-411.

Wedel, M. and W. A. Kamakura (2000). Market Segmentation: Conceptual and Methodological Foundations, Kluwer Academic Publishers.

Wolfersberger, J. and J. P. Monteleone. (2013). "The Digital Consumer Journey the Six Types of Digital Consumers; Amazon's Increasing Influence in the Purchase Pathway." Retrieved 5/18, 2015, from <https://blog.compete.com/2013/05/30/the-6-types-of-digital-consumers-and-their-paths-to-purchase/>.

Xiong, Y. and D.-Y. Yeung (2004). "Time Series Clustering with Arma Mixtures." Pattern Recognition **37**(8): 1675-1689.

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.

Zantedeschi, D., E. M. Feit and E. T. Bradlow (2013). Measuring Multi-Channel Advertising Effectiveness Using Consumer-Level Advertising Response Data, Working paper, The Wharton School, University of Pennsylvania, Philadelphia.

Tables and Figures

<i>Symbol</i>	<i>Explanation</i>
T	<i>length of time series data</i>
K	<i>number of clusters</i>
I	<i>number of consumers</i>
y_{ij}^t	<i>the exogenous variable j for consumer i at period t</i>
x_{ij}^t	<i>the endogenous variable j for consumer i at period t</i>
λ	<i>the mean of the Poisson distribution</i>
B	<i>autoregressive parameter to calculate λ</i>
ρ	<i>mixture parameter of the Zero-Inflated model</i>
G	<i>autoregressive parameter to calculate ρ</i>
θ	<i>autoregressive parameter in GLM</i>
Σ	<i>covariance matrix that captures the contemporaneous correlation</i>
u	<i>vector of the normal quantiles of the probability integral transforms of the endogenous variable under the marginal distribution</i>
z	<i>latent variable that indicates the membership of consumers</i>
r_{ik}	<i>the probability of assigning consumer i to cluster k</i>

Table 1. Variable Notations.

Variables	Definition	Min	Mean	Median	3rd quartile	Max
Browse	Number of product browsed online	0	2.78	0	0	1273
Search	Number of searches on the web site	0	0.01	0	0	26
Online Purchase	Number of online purchases	0	0.09	0	0	97
Offline Purchase	Number of offline purchases	0	0.36	0	0	94
Email	Number of emails received	0	1.39	0	3	20
Catalog	Number of catalogs received	0	0.14	0	0	2

Table 2. Description of customer-level weekly variables (based on 9,805 customers).

	VAR Mixture (Fong et al. 2007)	Model 1	Model 2	Model 3	Model 4 (CZMAP)
<i>Gaussian (G) / Poisson (P)</i>	G	P	P	P	P
<i>Zero-Inflation</i>	×	×	√	×	√
<i>Contemporaneous Correlation</i>	√	×	×	√	√
<i>Log-Like</i>	-3,701,671	-4,101,021	-3,675,528	-4,095,966	-3,671,275
<i>BIC</i>	7,400,769	8,200,158	7,347,334	8,189,359	7,338,138

Table 3 Likelihoods and BICs comparison for different multivariate autoregressive models. (√ indicates that the feature is included in the model. × indicates that the feature is absent.)

Impulse-Response	Response Type	Segment 1	Segment 2	Segment 3	Segment 4	Segment 5
Email-Online	Maximal Response	0.0104***	0.0038**	0.0041**	0.0042**	0.0036**
	p-Value	0.004	0.022	0.024	0.023	0.023
	% of the Weekly Purchase	61%	15%	3%	1%	1%
	Time of Maximal Response	2	2	3	2	2
	Cumulative Response	0.0182	0.0095	0.0113	0.0093	0.0077
Email-Offline	Maximal Response	0.0201***	0.0193***	0.0381***	0.0528***	0.0115***
	p-Value	0.003	0.003	0.001	0.001	0.003
	% of the Weekly Purchase	7%	7%	9%	9%	1%
	Time of Maximal Response	2	2	3	4	3
	Cumulative Response	0.0420	0.0363	0.0568	0.0814	0.0366
Email-Overall	Cumulative Response	0.0602	0.0458	0.0681	0.0907	0.0443
Catalog-Online	Maximal Response	0.0058**	0.0032	0.0310***	0.0269***	0.0028
	p-Value	0.023	0.132	0.001	0.002	0.163
	% of the Weekly Purchase	34%	13%	22%	7%	1%
	Time of Maximal Response	2	2	1	1	2
	Cumulative Response	0.0163	0.0087	0.0482	0.0482	0.0086
Catalog-Offline	Maximal Response	0.1021***	0.0421***	0.0712***	0.2144***	0.0335***
	p-Value	0.001	0.003	0.001	0.001	0.003
	% of the Weekly Purchase	33%	12%	15%	39%	4%
	Time of Maximal Response	1	1	3	4	2
	Cumulative Response	0.2235	0.0843	0.1668	0.2677	0.0703
Catalog-Overall	Cumulative Response	0.2398	0.093	0.215	0.3159	0.0789

Table 4. Cumulative and maximal purchase response for each segment. “p-Value” shows the statistical significance of the maximal response. “% of the Weekly Purchase” shows the relative magnitude of the maximal purchase response with respect to the average weekly purchase of a customer in the segment (see Table 7) in the corresponding purchase channel. Significance codes are 0 ‘*’ 0.01 ‘**’ 0.05 ‘*’ 0.1**

Start (Impulse)	Segment 1 46%	Segment 2 21%	Segment 3 20%	Segment 4 8%	Segment 5 6%	End (Purchase)
Email	(0.0104, 100%)					Online Purchase
Email	(0.0201, 100%)	(0.0193, 100%)	Search (0.0132, 34.75%)	Search-Search (0.0144, 27.45%)	Search (0.0043, 38.05%)	Offline Purchase
			Email (0.0101, 26.67%)	Browse-Search (0.0111, 21.31%)	Browse (0.0023, 20.35%)	
			Offline (0.0073, 19.42%)	Online-Browse (0.0079, 15.04%)	Online (0.0018, 15.92%)	
Catalog			(0.0310, 100%)	(0.0269, 100%)	Online Purchase	
Catalog	(0.1021, 100%)	(0.0421, 100%)	Browse-Search (0.0192, 26.91%)	Email-Search-Search (0.0582, 27.19%)	Email (0.0146, 43.58%)	Offline Purchase
			Search-Search (0.0147, 20.69%)	Online-Search-Search (0.0337, 15.76%)	Search (0.0092, 27.56%)	
			Search-Email (0.0108, 15.19%)	Browse-Search-Browse (0.0301, 14.08%)	Browse (0.0060, 17.62%)	

Table 5. Path-To-Purchase comparison for the five segments. There are two numbers associated with each path, the first number is the weight of the path and second number is the percentage of the path contributing to the maximal purchase response. When there are more than 3 significant paths, only the top 3 are shown. Therefore, the percentage attributed to the three paths may not add up to 100%.

Segment	Catalog	Catalog	Email	Email
1				
2				
3				
4				
5				
	Offline purchase	Online purchase	Offline purchase	Online purchase

Table 6. Distribution over all possible paths between the starting impulse and the maximal online/offline purchase response. The Y-axis shows the % of the maximal purchase explained by the path whose index is shown on the X-axis. We don't have any paths for the scenario under which there is no statistically significant maximal response. When the maximal response occurs within one time period we have only one possible path from the impulse directly to the maximal purchase response. Therefore, under these two scenarios distribution over the paths are not available.

Cluster Index	Relative Size	Demographic						Firm Behavior	
		Average distance to the nearest store	Median value of distance to the nearest store	Average Age	Average tenure with the brand ¹⁰ (days)	Proportion of female consumers	Proportion of Loyalty users	Average number of email (2 years)	Average number of catalog (2 years)
1	0.46	8.97	4.31	41.90	572.43	0.91	0.18	0.90	0.13
2	0.21	8.83	4.65	40.25	591.02	0.91	0.06	1.60	0.14
3	0.20	12.81	5.41	38.67	808.02	0.91	0.22	2.04	0.14
4	0.08	15.21	6.69	39.51	806.80	0.88	0.24	2.21	0.13
5	0.06	10.77	5.06	42.83	630.19	0.90	0.19	1.13	0.14
Overall	1.00	10.24	4.94	40.79	644.20	0.91	0.17	1.39	0.14
Cluster Index	Relative Size	Consumer Behavior							
		Average online purchase (2 years)	Average offline purchase (2 years)	Holiday online purchase/All online purchase	Holiday offline/All offline purchase	Average number of browse (2 years)	Average number of search (2 years)		
1	0.46	0.017	0.279	0.158	0.176	0.230	0.002		
2	0.21	0.025	0.296	0.184	0.181	0.538	0.001		
3	0.20	0.138	0.431	0.176	0.173	3.954	0.015		
4	0.08	0.371	0.519	0.170	0.180	13.608	0.021		
5	0.06	0.298	0.762	0.462	0.491	12.896	0.010		
Overall	1.00	0.09	0.36	0.172	0.189	2.78	0.01		

Table 7. Demographics and Weekly Behavior of Consumers in Five Segments.

¹⁰ Measured as the number days between the first purchase by the customer and the start of the data collection period.

Approaches using no sequence information		
The average of the training data	0.4	
Purchases at the last training period	0.4253	
Random selection	0.4045	
Regression on RFM	0.3814	
Approaches using sequence information		
Segmentation framework	Autoregressive specification	
	VAR	ZMAP
No segmentation	0.26523	0.24041
Demographics segmentation	0.22300	0.21533
Segmentation by short term response	0.21015 0.28122 (average VAR, Reimer et al. (2014))	0.20565
RFM	0.20850	0.20514
Mixture model	0.20145 (Fong et al. 2007)	0.19021

Table 8. Mean absolute error of out of sample prediction performance for individual purchase. Average number of purchases per customer per week is 0.44.

	Average target (per customer for the entire 10 week period)	\sqrt{Loss}
Segment 1	3.65	0.41
Segment 2	2.02	0.46
Segment 3	3.02	0.41
Segment 4	20.77	0.40
Segment 5	18.98	0.47

Table 9. The average loss of optimal marketing policy for the five segments

	\sqrt{Loss} Average sales target (per customer for the 10 week period) = 4.44
ZMAP (No Segmentation)	0.54
CVAR	0.47
CZMAP	0.44

Table 10. The average loss of optimal marketing policy based on the three models.

Appendix A: Plot of the catalog and email deliveries

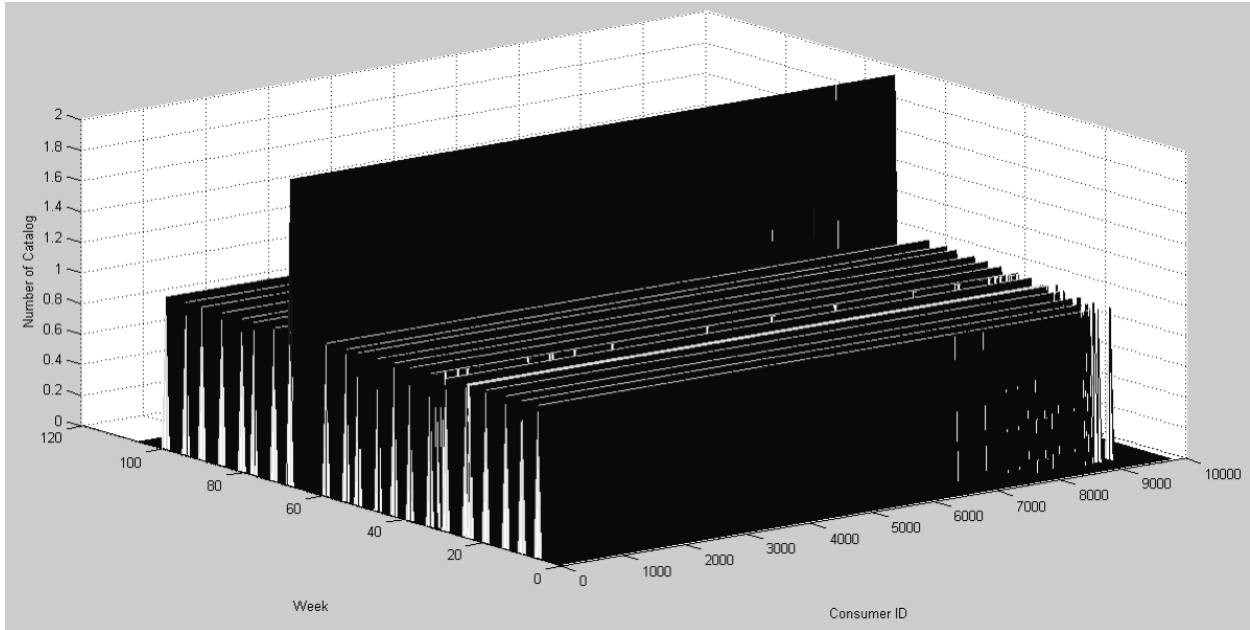


Figure A.1. Catalog distribution of all 9,805 consumers over 106 weeks.

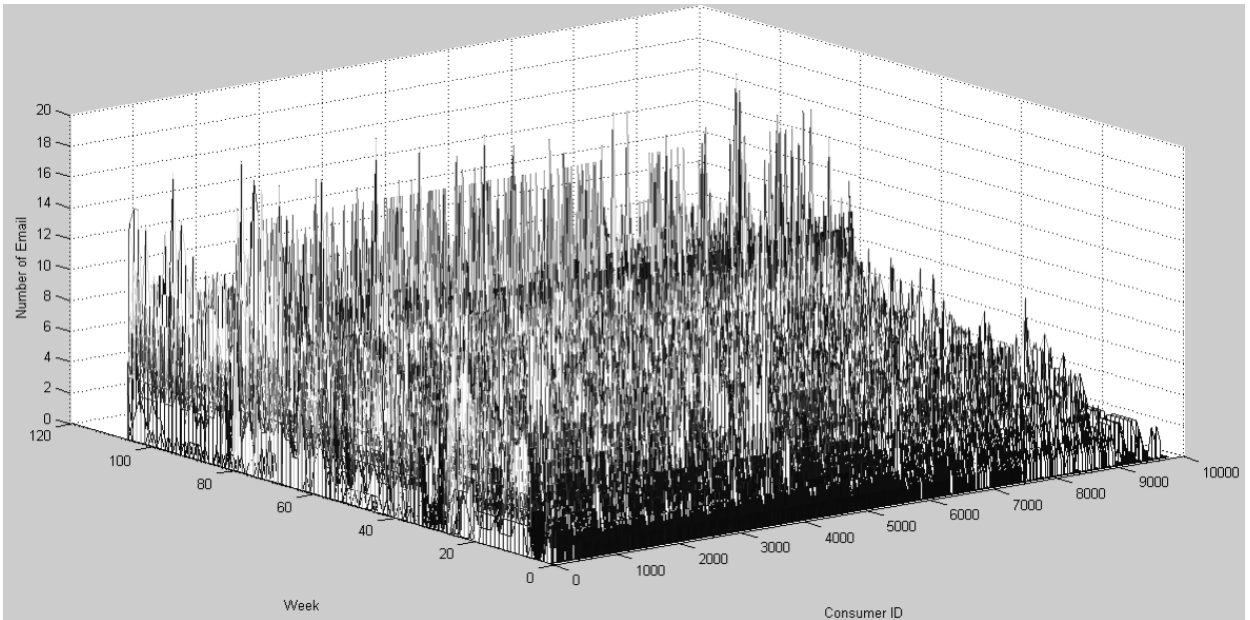


Figure A.2. Email distribution of all 9,805 consumers over 106 weeks.

Appendix B: Model Validation Using a Simulated Dataset

Before applying our model to a real world dataset, we validate its effectiveness in uncovering the existing clusters and parameters in each cluster by applying CZMAP to a simulated dataset with known cluster

memberships and parameters.

We begin with a mixture of ZMAP model to generate a dataset with 4 clusters with 500 customers in each, 5 endogenous variables and 4 exogenous variables. The order, p , of CZMAP is set to 2 for all consumers. We generate different datasets spanning $T=100, 200$ and 500 time periods. For each cluster, we randomly set the parameters $B_k, G_k, \text{ and } \Sigma_k$. For each customer we randomly generate the dataset for the first p time periods. Based on B_i, G_i and covariance matrix Σ_k and the cluster membership of each consumer, we generate the data for each consumer for the rest of the time periods. We then compare the proposed CZMAP model to three alternative time-series clustering approaches: 1) two feature-based clustering approaches similar to the one in Baragona (2011) where we first estimate time series parameters separately for each customer under the ZMAP model followed by clustering the consumers based on these parameters using K-means and k-Medoid algorithms, and 2) a model-based approach where a VAR Mixture model is fit to the data (Fong et al. 2007). We employ purity of the clusters, measured as the *fraction of the consumers assigned to the cluster containing majority of their original group members*, to measure the clustering accuracy (Manning et al. 2008). We ran the experiment five times for each case and averaged the purities. The results are shown in Table 11.

	T=100	T=200	T=500
K-Medoid	0.748	0.822	0.952
K-Means	0.743	0.832	0.959
VAR Mixture (Fong et al. 2007)	0.822	0.841	0.848
CZMAP	0.967	0.982	0.989

Table 11. Clustering performance comparisons on simulation data.

We find that K-Means and K-Medoid do not perform well when $T=100$ or 200 . The estimation of individual-level ZMAP coefficients is unreliable due to limited amount of individual-level data and in situations with extreme sparseness, we are unable to estimate the ZMAP using data from individual consumers. When the length of time-series is longer, the performance of the algorithm improves. In contrast, CZMAP performs well even with shorter time-series data. Since the CZMAP uses weighted data of all consumers that belong to a cluster rather than data of individual customers in estimating

coefficients, it does not suffer from the sparsity issue. The performance of VAR Mixture model does not change much when we increase the span from 100 to 500 because, like CZMAP, the VAR Mixture also uses weighted data of all consumers to estimate the model. Therefore, sparsity is not a major concern. However, the purity of VAR Mixture is much lower than that of CZMAP because the Normal distribution assumption of the VAR model is a poor fit to the sparse count data that are generated in the simulation. Overall, the simulation highlights the need to use the multivariate time series model with the most appropriate distributional form for the dataset at hand.

Appendix C: Details of the Copula and Results

Appendix C.1: Continued Extension of Discrete Variables for Copula

The copula model works under the assumption of all marginal distributions being continuous. But when dealing with the with count data, the assumption does not hold. To overcome the problem, we use the continued extension of a discrete variable by adding a continuous variable U to each discrete variable. Based on Heinen and Rengifo (2007), we add a random variable $U_{i,j}^t \sim Uniform(0,1)$ to the discrete variable $x_{i,j}^t$ to create its continued extension:

$$x_{i,j}^{t*} = x_{i,j}^t + (U_{i,j}^t - 1) \quad (22)$$

Then, we calculate $z_{i,j}^t$, the CDF of generating $x_{i,j}^{t*}$ based on ZIP, by the following formula:

$$z_{i,j}^t = F_{zip}(x_{i,j}^{t*}) = F_{zip}(x_{i,j}^t - 1) + f_{zip}(x_{i,j}^t) \times U_{i,j}^t \quad (23)$$

Then, we calculate $u_{i,j}^t$, the Normal quantiles of the probability integral transforms of $X_{i,j}^t$ under the marginal densities ZIP from $z_{i,j}^t$:

$$u_{i,j}^t = \Phi^{-1}(z_{i,j}^t) \quad (24)$$

where Φ^{-1} is the inverse standard Normal distribution function.

Appendix C.2: Contemporaneous Correlations

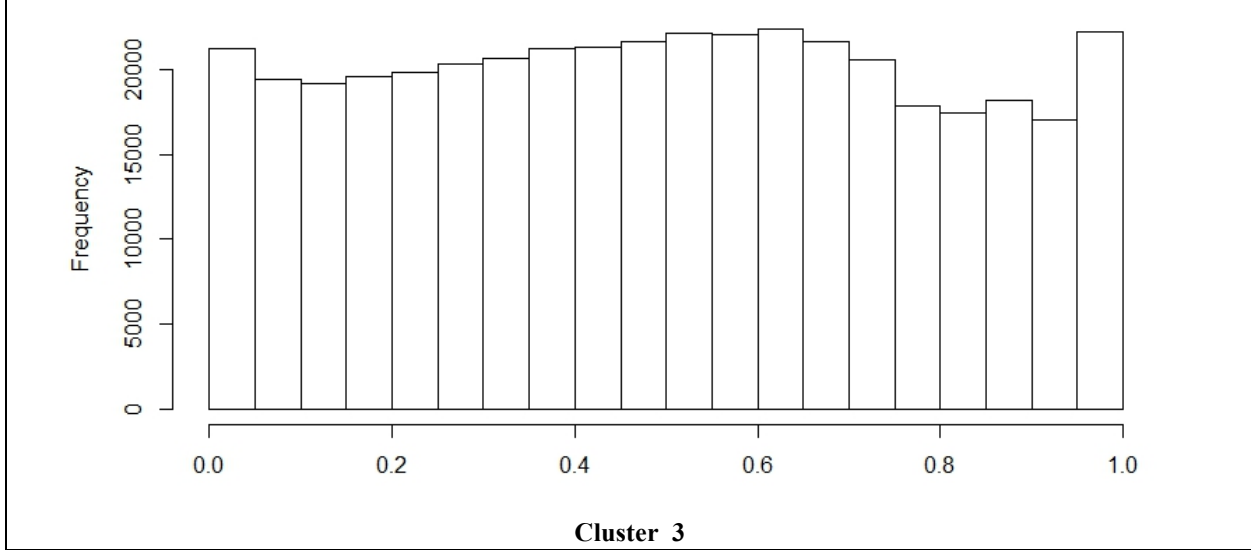
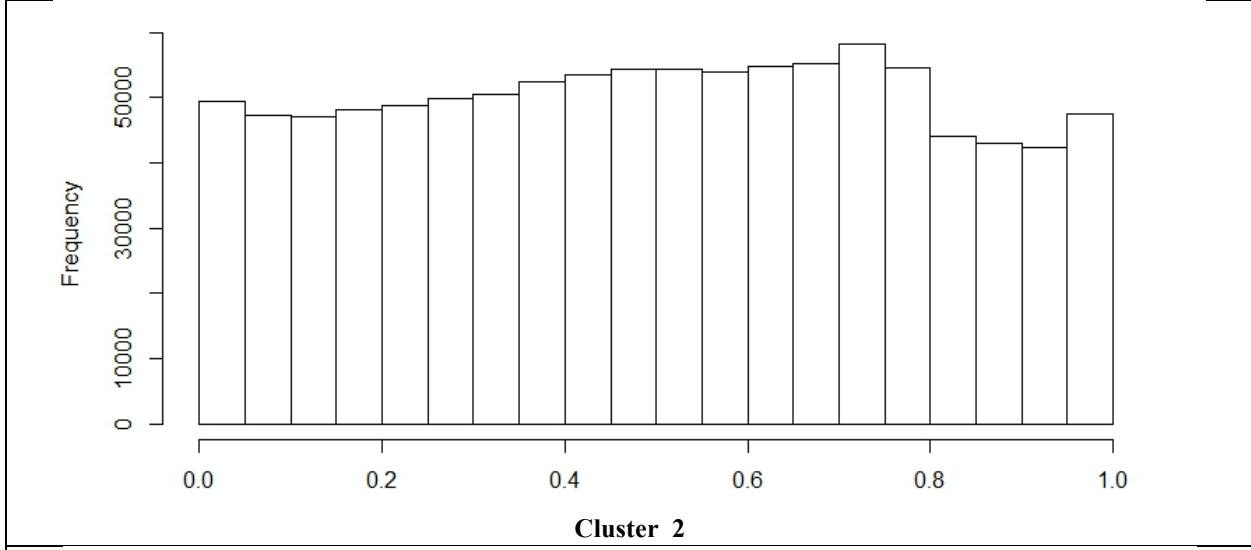
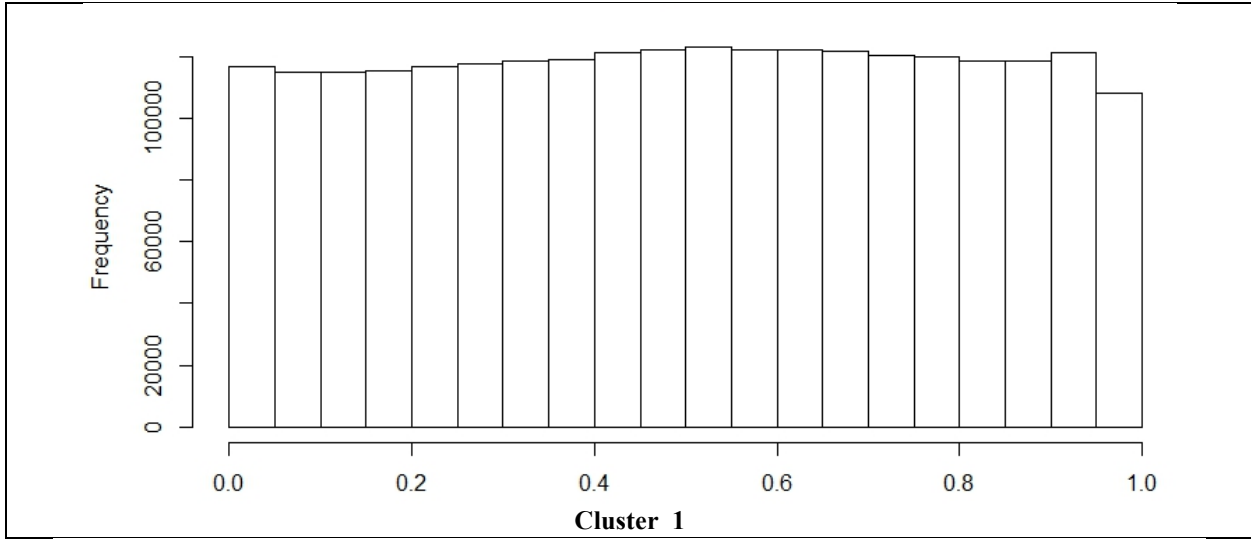
Correlation matrixes for Normal quantiles of the PITs for the five clusters are listed in Table 12. Most of the correlations between endogenous variables, although significant due to large sample size, are small in magnitude which indicates that the autoregression part explains endogenous variables well.

	Browse	Search	Online Purchase	Offline Purchase	Email
Cluster 1					
Browse	1.000				
Search	0.008***	1.000			
Online Purchase	0.044***	0.002	1.000		
Offline Purchase	0.011***	0.004	0.009***	1.000	
Email	0.025***	0.003	0.023***	0.014***	1.000
Cluster 2					
Browse	1.000				
Search	0.024***	1.000			
Online Purchase	0.166***	0.010***	1.000		
Offline Purchase	0.049***	0.007**	0.017***	1.000	
Email	0.066***	0.006**	0.070***	0.026***	1.000
Cluster 3					
Browse	1.000				
Search	0.021***	1.000			
Online Purchase	0.215***	0.015***	1.000		
Offline Purchase	0.069***	0.006	0.041***	1.000	
Email	0.089***	0.008*	0.114***	0.034***	1.000
Cluster 4					
Browse	1.000				
Search	0.013**	1.000			
Online Purchase	0.141***	0.004	1.000		
Offline Purchase	0.024***	0.002	0.033***	1.000	
Email	0.056***	0.003	0.080***	0.030***	1.000
Cluster 5					
Browse	1.000				
Search	0.010***	1.000			
Online Purchase	0.042***	0.002	1.000		
Offline Purchase	0.010***	0.000	0.016***	1.000	
Email	0.026***	0.000	0.022***	0.017***	1.000

Significance codes are 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

Table 12. Copula correlation matrices for the five clusters.

Following Heinen and Rengifo (2007) we examine the distribution of the probability integral transforms (PITs) of the endogenous variables. If the distributional assumptions for the components of the time series are correct, then the PITs would approximate a Uniform distribution. The plots are shown in Table 13. We can see that the distributions of the PITs do not show any major deviation from uniform distribution, suggesting that the models for the components of the time series are well specified.



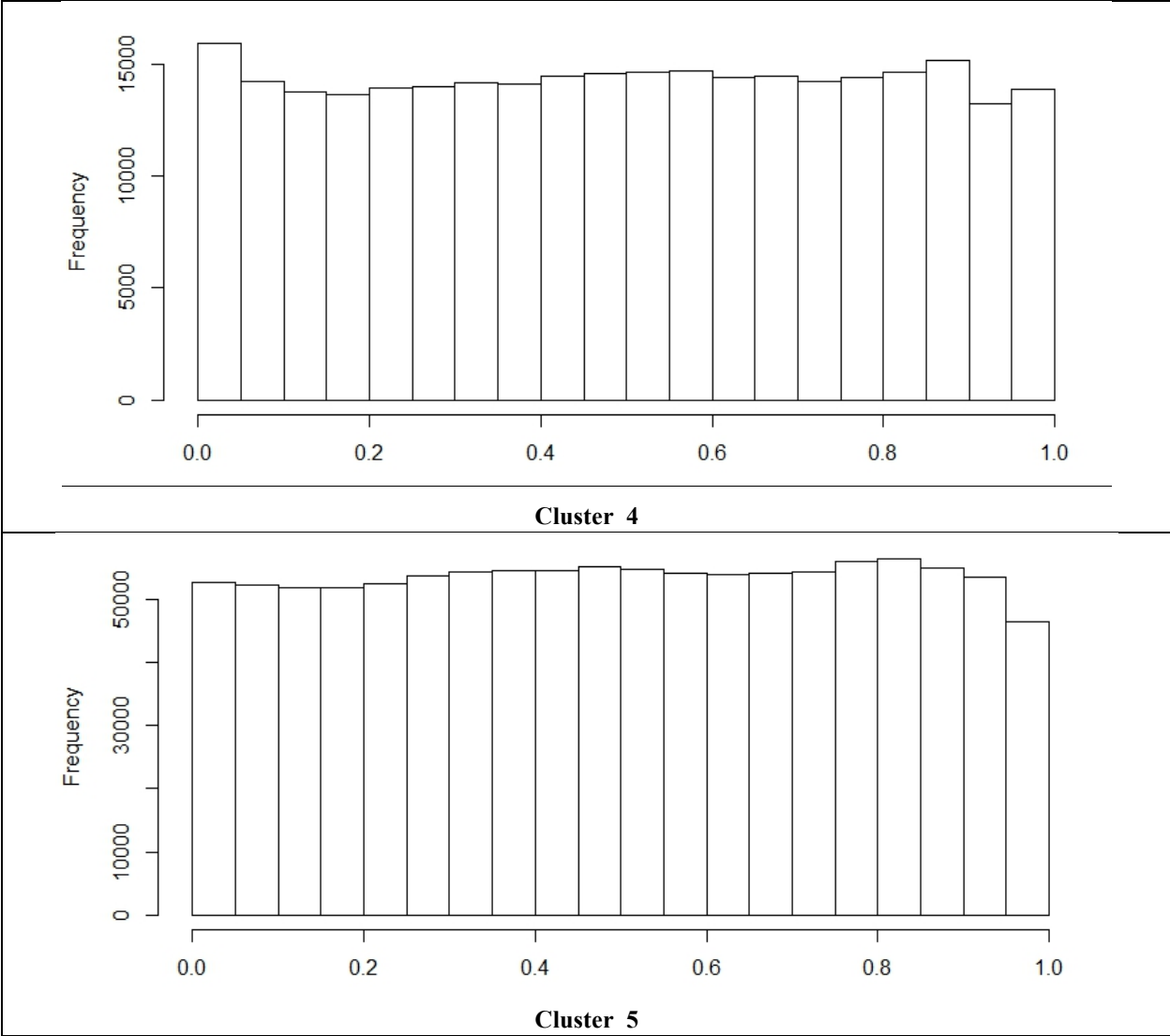


Table 13. Histograms of the Probability Integral Transforms show approximately uniform distribution, indicating that the marginal distribution for the components of the time series are well specified.

Appendix C.3: Copula Correlated Endogenous Variables and the Mean Function

Given the estimated copula correlation Σ a vector of endogenous variables with contemporaneous

correlation can be drawn from the marginal ZIP distributions according to the following process (Barbiero and Ferrari 2015).

1. $\mathbf{x}^t \sim MVN(0, \Sigma)$
2. $\mathbf{p}^t \sim CDF_{Normal(0,1)}(\mathbf{x}^t)$
3. $\mathbf{d}^t \sim CDF_{ZIP}^{-1}(\mathbf{p}^t; \mathcal{M}_{mean}^t, \Theta)$

where, \mathbf{d}^t is a draw from the ZIP marginal distributions with parameters Θ and contemporaneous correlation captured by the copula covariance matrix Σ , when the lagged values of the endogenous variables are \mathcal{M}_{mean}^t . Therefore, the mean of \mathbf{d}^t is given by the following function:

$$\boldsymbol{\mu}^t = m(\mathcal{M}_{mean}^t) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \varphi(\mathbf{x}^t; 0, \Sigma) CDF_{ZIP}^{-1}(CDF_{Normal}(\mathbf{x}^t); \mathcal{M}_{mean}^t, \Theta) dx_1^t \dots dx_j^t$$

where $\varphi(\mathbf{x}^t; 0, \Sigma)$ is the probability of generating \mathbf{x}^t based on the multivariable Normal distribution, whose mean is 0 and variance is Σ . For the j th element of $\boldsymbol{\mu}^t$ we get the following expression:

$$\begin{aligned} \mu_j^t &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \varphi(X^t; 0, \Sigma) CDF_{ZIP}^{-1}(CDF_{Normal}(x_j^t); \mathcal{M}_{mean}^t, \Theta_j) dx_1^t \dots dx_j^t \\ &= \int_{-\infty}^{\infty} CDF_{ZIP}^{-1}(CDF_{Normal}(x_j^t); \mathcal{M}_{mean}^t, \Theta_j) dx_j^t \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \varphi(X^t; 0, \Sigma) dx_1^t \dots dx_j^t \\ &= \int_{-\infty}^{\infty} CDF_{ZIP}^{-1}(CDF_{Normal}(x_j^t); \mathcal{M}_{mean}^t, \Theta_j) \varphi(x_j^t; 0, \Sigma_{jj}) dx_j^t \end{aligned}$$

where CDF_{Normal} is the standard normal CDF. Based on the above derivative, the off-diagonal elements of the copula covariance matrix are not involved in the mean calculation.

If inverse of marginal CDF was continuous, we can get the differential of this mean function directly. But since CDF_{ZIP}^{-1} is discrete, we need to transform the above equation in the following manner to calculate the differential. Set

$$z_i = z_i(\mathcal{M}_{mean}^t; \Theta) = CDF_{N(0,1)}^{-1}(CDF_{ZIP}(i; \mathcal{M}_{mean}^t, \Theta))$$

Note $z_i(\mathcal{M}_{mean}^t; \Theta)$ is continuous and differentiable with respect to the continuous variable \mathcal{M}_{mean}^t . Using

the result

$$\lim_{N \rightarrow \infty} z_N = \infty, \text{ because } \lim_{N \rightarrow \infty} CDF_{ZIP}(N; \mathcal{M}_{mean}^t, \theta) = 1$$

μ_j^t can be written as

$$\begin{aligned} \mu_j^t &= \int_{-\infty}^{\infty} CDF_{ZIP}^{-1}(CDF_{Normal}(x_j^t); \mathcal{M}_{mean}^t, \theta_j) \varphi(x_j^t; 0, \Sigma_{jj}) dx_j^t \\ &= \lim_{N \rightarrow \infty} \sum_{i=0}^N i \int_{z_{i-1}}^{z_i} \varphi(x_j^t; 0, \Sigma_{jj}) dx_j^t \\ &= \lim_{N \rightarrow \infty} \sum_{i=0}^N i \times (\Phi(z_i; 0, \Sigma_{jj}) - \Phi(z_{i-1}; 0, \Sigma_{jj})) \\ &= \lim_{N \rightarrow \infty} \sum_{i=0}^{N-1} (\Phi(z_N; 0, \Sigma_{jj}) - \Phi(z_i; 0, \Sigma_{jj})) = \lim_{N \rightarrow \infty} \sum_{i=0}^{N-1} (1 - \Phi(z_i; 0, \Sigma_{jj})) \end{aligned}$$

Last step uses the result $\lim_{N \rightarrow \infty} z_N = \infty \Rightarrow \lim_{N \rightarrow \infty} \Phi(z_N; 0, \Sigma_{jj}) = 1$.

In our empirical observation $\mu_j^t = \lim_{N \rightarrow \infty} \sum_{i=0}^{N-1} (1 - \Phi(z_i(\mathcal{M}_{mean}^t; \theta); 0, \Sigma_{jj}))$ converges quickly as N increases and in our application context above $N = 20$ mean doesn't increase significantly. Each term in the sum is differentiable with respect to \mathcal{M}_{mean}^t . Using this formulation differential of μ_j^t can be computed even for discrete marginal distributions such as Zero Inflated Poisson distribution.

Appendix D: Deriving the optimal marketing mix

Algorithm: Backward Recursion

- 1) $t = t_{end}$
 - 2) $V^{end}(x_{online}^{end+1}, x_{offline}^{end+1}) = 0$
 - 3) **While** $t \geq t_{start}$
 - a. $V^t(x_{online}^t, x_{offline}^t) = \min_{\text{OMD}^t} \{L(x_{online}^t, x_{offline}^t) + \theta V^{t+1}(x_{online}^{t+1}, x_{offline}^{t+1})\}$
 - b. $\text{OMD}^t = \underset{\text{OMD}^t}{\text{argmin}} \{L(x_{online}^t, x_{offline}^t) + \theta V^{t+1}(x_{online}^{t+1}, x_{offline}^{t+1})\}$
 - c. $t = t - 1$
 - 4) **End While**
 - 5) **Return** $V^{t_{start}}(x_{online}^{t_{start}}, x_{offline}^{t_{start}})$
-

Table 14. Backward recursion to derive optimal marketing policies.