

Detecting Customer Trends for Optimal Promotion Targeting

Lennart Baardman

Ross School of Business, University of Michigan, Ann Arbor, MI 48109, baardman@umich.edu

Setareh Borjian Boroujeni

Oracle RGPU, Burlington, MA 01803, setareh.borjian.boroujeni@oracle.com

Tamar Cohen-Hillel

Operations Research Center, MIT, Cambridge, MA 02139, tcohen@mit.edu

Kiran Panchamgam

Oracle RGPU, Burlington, MA 01803, kiran.panchamgam@oracle.com

Georgia Perakis

Sloan School of Management, MIT, Cambridge, MA 02139, georgiap@mit.edu

Problem Definition: Retailers have become increasingly interested in personalizing their products and services such as promotions. For this, we need new personalized demand models. Unfortunately, social data is not available to many retailers for cost and/or privacy issues. Thus, we focus on the problem of detecting customer relationships from transactional data, and using them to target promotions to the right customers.

Academic / Practical Relevance: From an academic point of view, this paper solves the novel problem of jointly detecting customer trends and using them for optimal promotion targeting. Notably, we estimate the causal customer-to-customer trend effect solely from transactional data, and target promotions for multiple items and time periods. In practice, we provide a new tool for Oracle Retail clients that personalizes promotions.

Methodology: We develop a novel probabilistic demand model distinguishing between a base purchase probability, capturing factors such as price and seasonality, and a customer trend probability, capturing customer-to-customer trend effects. The estimation procedure is based on regularized bounded variables least squares and instrumental variable methods. The resulting customer trend estimates feed into the dynamic promotion targeting optimization problem, formulated as a non-linear mixed-integer optimization model. Though it is NP-hard, we propose an adaptive greedy algorithm.

Results: We prove our customer-to-customer trend estimates are statistically consistent, and the adaptive greedy algorithm is provably good. Having access to Oracle Retail fashion client data, we show that our demand model reduces the WMAPE by 11% on average. Also, we provide evidence of the causality of our estimates. Finally, we demonstrate that the optimal policy increases profits by 3-11%.

Managerial Implications: The demand model with customer trend and the optimization model for targeted promotions form a decision support tool for promotion planning. Next to general planning, it also helps to find important customers and target them to generate additional sales.

Key words: Retail Operations, Demand Modeling, Instrumental Variables, Promotion Optimization, Promotion Targeting, Approximation Algorithms

1. Introduction

Trends play an important role in a variety of industries such as fashion clothing and consumer electronics. Over time, the products that customers are attracted to can change. At any point in time, this temporal interest is characterized as the current trend. In general, the trend has often been set by mass marketing and celebrity culture. Increasingly, customers are not only keeping up with this general trend, but also with trends in their own social circles. Social interactions have always been important in creating a trend, for example, due to the word-of-mouth effect. However, its importance has surged due to the rise of social media such as Facebook, Twitter, and Instagram. Nowadays, every person is able to create a social trend resulting in a change of the purchase behavior of their social connections.

Working together with the Oracle Retail group, we notice that their retail clients have become interested in personalizing their services such as pricing and promotions. At the same time, research related to personalized pricing has received increasing attention in the academic literature. Most purchases are able to create externalities whereby the sale to one customer affects other customers. We need to measure these customer-to-customer trend effects before we can personalize prices and target promotions effectively. In this work, our goal is to detect these customer-to-customer trends and use them to improve demand estimation and devise promotion targeting policies.

The process by which customers affect each other is complex. We are specifically interested in estimating the causal effect that one customer's purchase has on another customer's purchase decision. This is a difficult problem, because a purchase not only depends on trends, but can also be caused by factors such as an item's pricing or a time period's seasonality. Thus, for causal estimates of the customer trend, we need to disentangle the different factors. In this regard, social media data would be extremely useful, because it helps determine the important customer relationships. Unfortunately, privacy and/or cost issues make it hard for retailers to acquire detailed social data on their customers. To overcome this issue, we propose to detect customer relationships and estimate customer trends solely through transaction data that is readily available to the retailer.

Having estimated a demand model that accounts for customer trends, optimal dynamic promotion targeting can still be a complex problem. In previous literature, either complex graph models are estimated to describe social relationships without optimizing over the graph, or pricing is optimized to maximize revenue or influence spread over a given graph. In this work, we aim to combine both the estimation of the customer trend graph and the optimization of targeted promotions over this graph. For this purpose, we need to ensure that the customer trend can be optimized over. Therefore, we construct the customer trend demand model that can be represented by an interpretable graph model describing by how much one customer's purchase increases the probability of another customer's purchase.

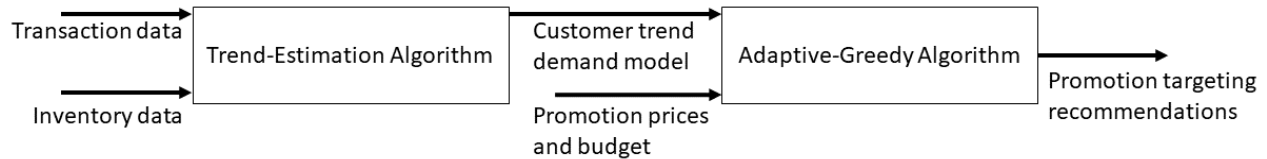


Figure 1 Workflow of the promotion recommendation tool

The above approach is depicted in Figure 1. In this paper, we develop the Trend-Estimation algorithm that estimates the customer trend demand model based on solely transaction data and inventory data, as well as the Adaptive-Greedy algorithm that optimizes the targeting of promotions given an estimated customer trend demand model and business parameters such as the available prices and budget.

1.1. Contributions

The main contributions of this paper are threefold. First, we develop a customer trend demand model that accounts for the fact that customers affect each other’s purchase probability. Second, we devise a model that solves the Dynamic Promotion Targeting Optimization Problem that also incorporates the customer trend effects. Finally, on real world instances from an Oracle Retail fashion client, we show that our models are effective in improving forecasting accuracy and increasing sales profit. The following gives a more detailed list of our contributions:

1. *Novel demand model accounting for customer trends.* We formulate an interpretable demand model estimating how the probability that a customer purchases an item at a time is driven by several factors as well as by other customers. The base of our model is formed by a classical demand model that captures traditional demand features such as price, brand, and seasonality. Added to this is the customer trend model that describes the change in purchase probability when other customers purchased an item earlier. We visualize the customer trend model as an unknown directed network where customers (or groups of similar customers) form the vertices, and edges describe the conditional probability of a customer to purchase given that another customer has purchased before them. Section 2.1 formulates the demand model and describes the required data for estimation.
 - (a) *Causal estimation procedure using only transaction data.* Due to the difficulty in acquiring social media data, we use purchase transaction data to detect customer trend effects. This means that we have to both detect customer relationships and estimate their strength. In contrast, when social data is available, we only need to estimate the strength of customer relationships. Nonetheless, we develop the *Trend-Estimation algorithm* that works for any general structure on the base purchase model and discovers causal customer trend effects. To deal with estimation complexity, we use a two-stage approach where first the base purchase model is estimated, and afterwards the customer trend model is fit. To assert causal

relationships, we adapt the instrumental variables (IV) method to fit our model, where we use lagged variables as instruments. Section 2.2 describes this estimation procedure and details how it addresses the aforementioned issues.

- (b) *Consistent estimation and finite-sample guarantee on estimation error.* Because we utilize the customer trend when targeting promotions, we want to guarantee that the estimates of the customer trend model are consistent. We prove this by showing that, under mild conditions, the estimated customer-to-customer-trend probabilities converge to the underlying ‘true’ probabilities. Under stronger assumptions, we extend this result to a probabilistic finite-sample guarantee on the estimation error between the estimated and ‘true’ customer-to-customer trend probabilities. Section 2.2.1 presents these results in more detail.

2. *Dynamic optimization model for targeting promotions.* We formulate the Dynamic Promotion Targeting Optimization Problem that maximizes the total expected revenue of multiple items over multiple time periods while satisfying business rules such as promotion limits and inventory shipping. A targeted promotion policy determines which customer should receive a special promotion for which item and at what time. In this model, we specifically account for the customer-to-customer trend effect. Not only are we able to target those customers who respond unusually strongly to promotions, but also those customers who generate notably more additional purchases through their network influence. Given its formulation, we show that the Dynamic Promotion Targeting Optimization Problem is NP-hard. Section 3.1 formulates the promotion targeting model and demonstrates its complexity.

- (a) *Tractable Adaptive-Greedy algorithm to approximate solution.* Though the dynamic promotion targeting optimization model is hard to solve, we develop a tractable *Adaptive-Greedy algorithm* that finds an approximate optimal solution. This algorithm can find the optimal solution exactly in special cases that we characterize. Section 3.2 describes the Adaptive-Greedy algorithm.
- (b) *Analytical guarantee through C -submodularity.* Invoking the concept of submodularity and the newly defined concept of *C -submodularity*, we are able to show analytical guarantees on more general cases. In the case where items are complementary, we prove that the revenue function is submodular in the promotion policy. As a result, we show that the Adaptive-Greedy algorithm receives at least a $(1 - \frac{1}{e})$ fraction of the optimal revenue. In the general case where some items can be complementary and others substitutable, we introduce the new concept of *C -submodularity*, which is a measure of how non-submodular a function is. In this part, we show a parametric guarantee on the revenue, as a function of the non-submodularity. Section 3.2.1 discusses and proves these analytical guarantees.

3. *Practical contribution.* Working together with the Oracle Retail group, we have access to data from a large fashion retail client. Using their transaction data, we are able to test the forecasting accuracy of our demand model, and we are able to evaluate the effectiveness of our targeted promotion policies. Section 4.1 introduces the available transaction data.
 - (a) *Strong forecasting accuracy on real world data.* We test the accuracy of our demand model relative to classical demand models that ignore customer trends. This comparison shows that including the customer trend model improves the Weighted Mean Absolute Percentage Error (WMAPE) by 5 percentage points or 11% on average. Also, we test and find that the instrumental variables are valid. Sections 4.2 and 4.2.1 present the improved forecasting metrics and the causality analysis.
 - (b) *Improved revenue on realistic instances.* We examine how our targeted promotion policies increase revenue compared to the retailer's existing practice. We observe that the retailer's revenue is increased by 3% up to 11%. The average revenue improvement is on the order of 7%. This increase is exclusively due to an improved targeted promotion policy, and hence significant, due to the thin margins in the retail industry. Section 4.3 presents the details of this comparison.

1.2. Literature Review

In the face of demand uncertainty, lack of historical data, and seasonal trends, demand prediction for the fashion industry is a significant challenge for retailers. This is why it is not surprising to see a large body of literature on demand prediction in the fashion industry (Nenni et al. 2013, Liu et al. 2013, Beheshti-Kashi et al. 2015, Mohr 2013). According to Beheshti-Kashi et al. (2015), alongside classic models, more and more research has been oriented towards understanding the impact of social media on demand and personalized demand estimation.

Nonetheless, the advantages of understanding the impact of social media are not limited to mere prediction. In Alves et al. (2016), the authors provide a wide review on the impact of well-planned Social Media Marketing (SMM). The authors mention that social media sentiment has a stronger impact on firm stock performance than conventional media.

The literature on SMM can be split into two main streams of literature. The first stream of literature deals with the question of how to correctly plan an effective SMM policy. This stream usually assumes prior partial knowledge of the social network and accordingly, find the central influential people that should be targeted for an SMM campaign (Pradiptarini 2011, Rathore et al. 2016, Atazky and Barone 2015). The second stream of literature deals with the question of how to measure the failure/success of an SMM campaign. Most of the common measures, such as the number of clicks or shares, provide little to no information regarding the impact of the campaign in terms of dollar return. Therefore, a large body of research has been focused on understanding

the return on investment (ROI) of SMM campaigns (Kumar and Mirchandani 2012). This research contributes to both streams on the literature on SMM.

To battle the question of finding influential customers, platforms (such as *Stack Influence*, *Tap-Influence*, *Influence Network*, *ApexDrop Influence Marketing*, etc.), were established in order to link influential people with fashion retailers. In such platforms, customers can register as influencers and are required to provide evidence for their impact on their environment (such as connections on social media, or daily views in their fashion blog). Retailers can then use this platform to locate influencers in specific areas. Considering only the self-declared influencers, without a measure of impact, can lead to suboptimal marketing strategies. In order for retailers to be able to plan an optimal strategy, they need to be able to view the entire social network.

Unfortunately, in many of the cases, the social network is unknown and must be inferred. Significant literature has focused on using field experiments to detect the strength of social relationships (Aral and Walker 2011, Bakshy et al. 2012, Bapna and Umyarov 2015, Eckles et al. 2017). Without social data available, we need to not only estimate the strengths, but also the underlying network.

Several papers analyze this problem through influence models proposing that diffusion happens with a certain probability (Goyal et al. 2010, Myers and Leskovec 2010, Gomez Rodriguez et al. 2010, 2011). These papers assume that a node can only be infected once. Whereas in our retail setting, multiple purchases can occur at different times, especially when we estimate the model at a level where a node represents a group of customers.

When learning a network for the purpose of promotion targeting optimization, it is important to be able to distinguish between spurious correlation and causal effect. While the literature above is not concerned with showing causal effects, Bramoullé et al. (2009) estimate causal effects in networks. They consider individuals as parts of a group and their outcome is described by a model that is linear in the average outcome of their group. The endogenous peer effect is estimated through the instrumental variables (IV) method; an adapted version of the IV method is also used in this work. For a detailed review on IV methods, we refer to Cameron and Trivedi (2005) and Imbens (2014).

Once the network is known, retailers remain with the question of how to find an optimal marketing strategy that utilizes the social connections in the network. One example of a marketing tool that utilizes the social network structure is viral marketing. While mass marketing indiscriminately promotes products to all customers. Direct marketing personalizes these strategies by promoting to those customers that are most likely to purchase. Viral marketing advances one step further by promoting to those customers that also have a large social influence, for example through the word-of-mouth effect.

In order to understand the impact of each customer on the social network, in Domingos and Richardson (2001) the authors determine the network value of customers in addition to their intrinsic

value. They propose a Markov random field model to describe the purchase decisions of customers, and with this model, they are able to compute the expected global lift in profit if a certain set of customers receive a promotion. Kempe et al. (2003, 2015) furthered this by considering the influence maximization problem: selecting the initial set of nodes to activate such that the final number of active nodes is maximized. Their main result is that the influence maximization problem can be efficiently approximated if the nodes become active according to a linear threshold or independent cascade model. Further literature improved the speed of the initial algorithms, some with guarantees (Leskovec et al. 2007, Borgs et al. 2014) and some without (Chen et al. 2009, 2010). A major difference with this literature is that we consider the more general case of offering promotion prices to maximize revenue instead of determining influencers to maximize adoption.

The problem of influence maximization is highly related to the well-studied vertex-cover and max-cover problems. In the case of max-cover, given several sets of elements and a number k , we must select at most k sets such that the maximum number of elements are covered. The objective in the case of the max-cover problem is to maximize the number of covered elements and hence, the problem has a submodular structure. In Nemhauser et al. (1978), the authors show that for maximization problems with monotone submodular objective functions, a greedy algorithm can find a solution that is at most $(1 - \frac{1}{e})$ of the optimal solution. This result was since extended to the online settings, as well as adaptive (interactive) optimization problems. For a review on Submodular Function Maximization (Krause and Golovin 2014). Randomized algorithms have also been proposed to battle the problem of Submodular Function Maximization (Buchbinder et al. 2014). In this work we aim to maximize the impact of targeted treatment for different products. The cross product effect can lead to a non-submodular objective, which implies that the existing literature may not be of help.

While the problem of Submodular Function Maximization has been well studied, the problem of maximizing over partially submodular functions remains unsolved. In this work, we extend the results in Nemhauser et al. (1978), and propose a bound on the optimality gap achieved by the greedy algorithm for the case of non-submodular functions and subject to cardinality constraints. We illustrate this bound with respect to the promotion targeting problem.

Among other applications for the influence maximization problem are network epidemic elimination (Drakopoulos et al. 2014, 2016, Aral et al. 2009), spread of computer viruses (Garetto et al. 2003), or diffusion of innovations (Sahin 2006). See Nowzari et al. (2016) for a survey on spreading processes on complex networks. One of the most commonly used applications for influence maximization is in promotion pricing. While the problem of influence maximization concerns with the questions of “who” to pick, the question of pricing over networks adds the dimension of the price. This problem is a generalization of the influence maximization problem, as it allows us to give different treatments to different customers.

Research of pricing over networks was initiated by Hartline et al. (2008). They assume customers have a valuation for the product, which is drawn from a known distribution. Different influence structures result in different valuation distributions, particularly, a customer’s valuation distribution changes as other customers buy the product (higher previous purchases implies a higher probability of a large valuation). Given this framework, Hartline et al. (2008) proposes the influence-and-exploit marketing strategy. Herein, selected customers are influenced through a free product, after which the general population is exploited by offering the product at an optimal myopic price. In contrast, Akhlaghpour et al. (2010) consider the problem without price discrimination. They show that even in the case of a single price, the problem is highly intractable. Alternatively, Arthur et al. (2009) consider the same problem in a “cashback” setting instead of a “free product” setting.

Subsequently, other papers have extended the problem of pricing over networks. Candogan et al. (2012), Bloch and Qu  rou (2013), Cohen and Harsha (2013), Lu and Lakshmanan (2012) consider a monopolist offering individualized prices to each customer in more general known networks, and determine the optimal individualized prices dependent on network externalities. Accounting for network externalities is shown to be able to lead to a significant increase in profit. For a review on pricing and promotion planning see Cohen-Hillel et al. (2019).

2. Customer Trend Demand Model

In what follows, we introduce and formulate our demand model that incorporates the customer trend model. For this purpose, the available transaction data needs to describe the purchases that n unique customers made of m distinct items, each item being available in the different stores for T_i time periods. More details about the available Oracle Retail client data are given in Section 4.

2.1. Model Formulation

As described, we consider a probabilistic model of demand that uses a variety of factors to describe the probability that a customer c purchases item i at time t . We are interested in a model that describes the probability of the binary random variable B_{cit} that equals 1 if customer c purchases item i at time t and equals 0 otherwise. The factors that explain the purchase probability can be split into two types: trend-independent effects (such as price, style, seasonality, and promotion) and trend-dependent effects (namely, customers following the customer-to-customer trend). In what follows, we will introduce a set of random variables that indicate what caused customer c to purchase item i at time t . In particular, in the case where the purchase decision is based on trend-independent factors, we let $F_{cit}^{c'it'} = 0$ denote that the purchase decision of customer c for item i at time t is not based on following the other customer c' for item i at previous time t' . On the other hand, in the case that the purchase decision is based on trend-dependent factors, we let $F_{cit}^{c'it'} = 1$ denote that the purchase decision of customer c for item i at time t is based on following what another customer c'

did with the same item i at a previous time t' . In this latter event, we assume that customers will only follow the decision of others between time $t - M$ and $t - 1$. Beyond the memory of M time periods ago, the customer-to-customer trend tapers off. For example, because the earlier customer might not be exerting effort to portray the product as fashionable anymore. Later in Section 4.2, we describe how the memory M can be tuned from the data. These events are collectively exhaustive, but we assume they are also mutually exclusive, meaning that $\sum_{c'=1}^n \sum_{t'=t-M}^{t-1} F_{cit'}^{c'it'} \leq 1$. This assumption of following exactly one other customer is natural in the network pricing literature (Candogan et al. 2012, Kempe et al. 2003). Given these options, we can use countable additivity to define the purchase probability as follows,

$$\mathbb{P}(B_{cit} = 1) = \mathbb{P}\left(B_{cit} = 1, F_{cit}^{c'it'} = 0 \forall c', t'\right) + \sum_{c'=1}^n \sum_{t'=t-M}^{t-1} \mathbb{P}\left(B_{cit} = 1, F_{cit}^{c'it'} = 1\right). \quad (1)$$

The first term of equation (1) is the base purchase model that describes the case where the purchase decision is based on trend-independent effects. This term can be viewed as a traditional demand model, for the reason that traditional demand forecasting features, such as an item's price and style or a time period's seasonality and promotion status, can be subsumed under the category of trend-independent factors. It is important to note is that we capture cross-item effects in this part of our demand model. We consider the set of prices that are presented to customer c for all other items at time t as features in the base model. If we denote the customer-item-time specific features by the vector x_{cit} , we use $q(x_{cit})$ to represent these traditional demand models. Examples of $q(\cdot)$ are linear regression, logistic regression, regression trees, or random forests among others. The application in Section 4 uses a logistic regression model for $q(\cdot)$.

The second term of equation (1) incorporates the customer trend model that captures the case where the purchase decision is based on trend-dependent effects. This term proposes that if the purchase decision is based on the customer trend, then this is because customer c decided to follow the decision of one specific customer c' at one specific time period t' . Now, if the purchase decision is based on following the behavior of customer c' for item i at time t' ($F_{cit'}^{c'it'} = 1$), it still needs to be true that an actual purchase was made. Thus, the probability that customer c purchases item i at time t based on following customer c' at time t' can be conditioned on the fact whether customer c' actually purchased item i at a previous time t' ,

$$\begin{aligned} \mathbb{P}\left(B_{cit} = 1, F_{cit}^{c'it'} = 1\right) &= \mathbb{P}\left(B_{cit} = 1, F_{cit}^{c'it'} = 1 | B_{c'it'} = 1\right) \mathbb{P}(B_{c'it'} = 1) \\ &\quad + \mathbb{P}\left(B_{cit} = 1, F_{cit}^{c'it'} = 1 | B_{c'it'} = 0\right) \mathbb{P}(B_{c'it'} = 0), \end{aligned}$$

where we note that $\mathbb{P}(B_{c'it'} = 0) = 1 - \mathbb{P}(B_{c'it'} = 1)$, to obtain the following,

$$\begin{aligned} \mathbb{P}(B_{cit} = 1, F_{cit}^{c'it'} = 1) &= P(B_{cit} = 1, F_{cit}^{c'it'} = 1 | B_{c'it'} = 0) \\ &\quad + \left(P(B_{cit} = 1, F_{cit}^{c'it'} = 1 | B_{c'it'} = 1) - P(B_{cit} = 1, F_{cit}^{c'it'} = 1 | B_{c'it'} = 0)\right) P(B_{c'it'} = 1). \quad (2) \end{aligned}$$

At time t , we know what happened at earlier times, and hence the actual probability $\mathbb{P}(B_{c'it'} = 1)$ will be known to be either 0 or 1. In equation (2), the customer-to-customer trend is characterized by the conditional purchase probability. When customer c' purchases item i at time t' it has the opportunity to create a trend for item i among the other customers in future periods. In our model, we assume that this customer-to-customer trend probability is independent of the specific item and time under consideration. Without this assumption, there would not be enough data to estimate any customer-to-customer trend probabilities from. The assumption that the customer trend is independent of the specific item is justified by the idea that the customer trend behaves similarly for items in the same category (e.g., women's dresses, women's tops, men's tops). Hence, when we fit the model, we ensure that all items in the dataset come from the same category. For the independence between the customer trend and the specific time period, we argue that the memory M captures that the customer trend only exists between close time periods.

Under this assumption, we can define the customer-to-customer trend probability in a less complicated manner as $p_{c'c} = P(B_{cit} = 1, F_{cit}^{c'it'} = 1 | B_{c'it'} = 1) - P(B_{cit} = 1, F_{cit}^{c'it'} = 1 | B_{c'it'} = 0)$. We can interpret the customer-to-customer trend probability as the increase or decrease in the purchase probability of customer c for item i at time t when following the decision of customer c' to buy or not buy item i at time $t' < t$. In this paper, we assume that customers are not likely to buy an item for the sole reason that another customer did not buy it, meaning that $P(B_{cit} = 1, F_{cit}^{c'it'} = 1 | B_{c'it'} = 0) = 0$. However, under the definition above, we can also capture the case where customers could have both positive and negative effects on each other.

Additionally, this representation of the customer-to-customer trend probability allows us to think of the customer trend as evolving over a network. Namely, the customer trend is defined by a weighted graph where each vertex is a customer c , each directed edge indicates whether customer c' affects customer c , and the weight on each edge is $p_{c'c}$. Figure 2 gives examples of such a network and the corresponding customer-to-customer trend probabilities.

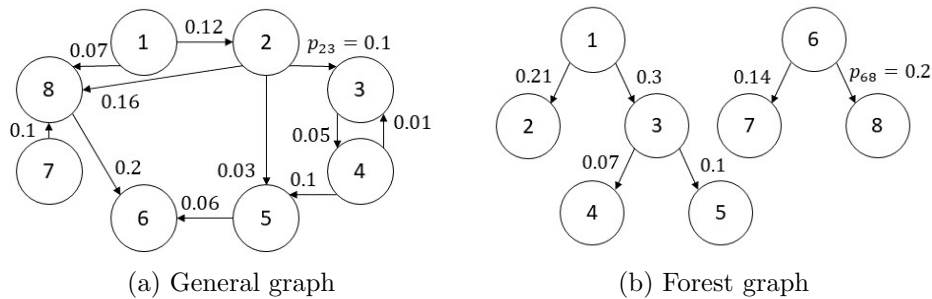


Figure 2 Examples of the customer trend portrayed as a graph with customers as the vertices, edges as the customer trends, and weights as the customer-to-customer trend probabilities

In addition, we simplify the notation of the purchase probability as $b_{cit} = \mathbb{P}(B_{cit} = 1)$ and the notation of the base purchase probability as $q_{cit} = q(x_{cit}) = \mathbb{P}(B_{cit} = 1, F_{cit}^{c'it'} = 0 \forall c', t') + \sum_{c'=1}^n \sum_{t'=t-M}^{t-1} P(B_{cit} = 1, F_{cit}^{c'it'} = 1 | B_{c'it'} = 0)$. Combining these ideas together, we propose the following customer trend demand model:

$$b_{cit} = q_{cit} + \sum_{c'=1}^n \sum_{t'=t-M}^{t-1} p_{c'c} b_{c'it'} \quad \text{or} \quad b_{cit} = q_{cit} + \sum_{c'=1}^n p_{c'c} \sum_{t'=t-M}^{t-1} b_{c'it'}. \quad (3)$$

In what follows, our goal is to estimate this model, assess its validity, and test its performance. This entails estimating the base model, producing $\hat{q}(\cdot)$, as well as the customer trend probabilities, generating $\hat{p}_{c'c}$. Equipped with these estimates, the customer trend demand model can be used for predicting purchases.

The model in equation (3) describes the structural equation underlying the purchase probability b_{cit} . However, the data that we observe is not generated exactly from this model for two reasons. Firstly, model misspecification such as certain variables affecting the purchase probability being omitted from the model, means that the observations of b_{cit} are different from the value given by equation (3). Secondly, the available transaction data can not directly tell us what the purchase probability is, we can only observe whether a transaction resulted in a purchase.

Let y_{cit} represent the outcome observed in the data. In other words, $y_{cit} = 1$ if the transaction data shows that customer c purchased item i at time t and equals 0 otherwise. For the most granular level of data (individual customer, single item, particular time), the value of y_{cit} is our best approximation to the purchase probability b_{cit} . However, when individual customers are aggregated into customer groups we are able to approximate the purchase probability more granularly using the empirical probability. For example, if \mathcal{C} is a set of customers, then we can use the following empirical probability instead (where $|\cdot|$ denotes the cardinality of a set),

$$y_{\mathcal{C}it} = \frac{1}{|\mathcal{C}|} \sum_{c \in \mathcal{C}} y_{cit}. \quad (4)$$

Here, the approximation of the purchase probability can be interpreted as the average probability that a customer in group \mathcal{C} buys item i at time t . In the rest of this paper, we use the notation y_{cit} regardless of whether the indices indicate individuals or groups.

Given the available data, we estimate $\hat{q}(\cdot)$ and $\hat{p}_{c'c}$ from the customer trend demand model:

$$y_{cit} = q_{cit} + \sum_{c'=1}^n p_{c'c} \sum_{t'=t-M}^{t-1} y_{c'it'} + \epsilon_{cit}, \quad (5)$$

where ϵ_{cit} is the error that captures the uncertainty to possible model misspecification and using y_{cit} as an approximation of b_{cit} .

2.2. Trend-Estimation Algorithm

Having described the model, we are interested in estimating the base purchase model q and the customer-to-customer trend probabilities $p_{c'c}$. Analyzing the customer trend demand model (5), we observe that it closely resembles a linear regression model for a given c . The model error ϵ_{cit} is additive to the structural linear part of the model. The base model q_{cit} can be seen as the constant of the linear regression model, the customer trend probabilities $p_{c'c}$ form the linear regression parameters, and the summations of past purchase probabilities $\sum_{t'=t-M}^{t-1} y_{c'it'}$ form the independent variables of the linear regression model. Our estimation procedure will use the fact that the model is close to a linear regression model. However, there are three complications that need to be accounted for when estimating this model: 1) estimation complexity: the model is hard to estimate, 2) parsimonious model: the customer-to-customer trend network is preferred to be sparse, 3) causal relationships: the estimates of the customer-to-customer trend need to indicate causal relationships.

First, the model is complex to estimate because the ‘constant’ of the linear regression model is actually formed by a possibly non-constant function $q(\cdot)$ of the customer-item-time specific features x_{cit} . For a general q this means it is hard to estimate the model directly. Nonetheless, we devise an estimation procedure that uses ideas from the Box-Jenkins method. The Box-Jenkins method is a multi-stage modeling approach that is often used to fit time series models. The first stage determines whether the time series is stationary, and if this is not the case, then the time series data is standardized by identifying and removing the factors that make the time series non-stationary. The second stage then fits a time series model to the standardized time series data.

Our estimation algorithm takes a similar multi-stage approach. First, we fit the purchase probability solely on the base purchase model to find $\hat{q}(\cdot)$. Then, we standardize the purchase probability by subtracting the predicted base purchase model to obtain the standardized demand model with customer trend:

$$y_{cit} - \hat{q}_{cit} = \sum_{c'=1}^n p_{c'c} \sum_{t'=t-M}^{t-1} y_{c'it'} + \epsilon_{cit}. \quad (6)$$

Finally, we fit the linear regression in equation (6) to find $\hat{p}_{c'c}$. This procedure significantly diminishes the estimation complexity, as it reduces to estimating a standard model to identify $\hat{q}(\cdot)$ and a linear regression to identify $\hat{p}_{c'c}$.

Secondly, there are two aspects that warrant the use of regularization in our estimation procedure. One aspect is that a large number of parameters need to be estimated, namely n^2 customer-to-customer trend probabilities, but realistically the model is sparse, as in that many of these parameters are equal to 0. Each parameter indicates the probability that one customer affects another, and realistically not every customer affects all other customers. Regularization can aid in estimating a parsimonious model by creating a sparse set of the most explanatory relationships. Another aspect

is that the purchase probabilities are measured with error. As explained before, the data y_{cit} is not exactly equal to the true purchase probability b_{cit} . This measurement error can also be dealt with by including regularization in the estimation algorithm (see for example Bertsimas and Copenhaver (2018)). When estimating the customer-to-customer trend probabilities, the objective of the estimation model includes a LASSO regularization penalty.

Lastly, because the goal is to optimize the targeting of promotions, we need to ensure that the estimated effect of a promotion on the likelihood of a purchase is causal. Without addressing this concern, an estimation method might capture spurious correlations, which would prevent us from finding the causal effects of the customer-to-customer trend. In the linear model described by equation (6), we can find a spurious non-causal effect if the variable of interest (here, recent purchase decisions by other customers) is correlated with other factors affecting the variable that we predict (here, current purchase decisions) that we do not account for in our model.

To address this possibility, we can take two measures: 1) include important control variables for the variables that we can account for, 2) use the instrumental variables method for variables that we cannot account for. We include all important control variables in our base purchase model. In our application, we include confounding factors such as trend and seasonality (to account for general fashion trends), pricing (to account for large promotion campaigns or loyalty discount programs), and inventory (to account for certain stores just receiving inventory earlier than others). Afterwards, we estimate the customer trend model using the instrumental variables method (for more details on instrumental variables we refer to Cameron and Trivedi (2005)). If we select valid instruments, this ensures that we account for all other factors beyond the control variables. Altogether, having accounted for a host of possible confounding factors and under the validity of our instruments, we argue that the model finds causal effects and not spurious relationships.

Putting our model in the context of instrumental variables, we can view the standardized demand model with customer trend (6) for a fixed customer c as a linear regression model: $y = X\beta + \epsilon$. We can view the vector of independent variables y as containing the standardized purchase probabilities $y_{cit} - \hat{q}_{cit}$ (observations are indexed by pairs of c , i and t). The matrix of dependent variables X consists of the past purchase probabilities of each customer $\sum_{t'=t-M}^{t-1} y_{c'it'}$ (observations are indexed by pairs of c , i and t , and variables are indexed by c' and c). The vector of linear regression parameters β contains the customer-to-customer trend probabilities $p_{c'c}$ (variables are indexed by c' and c). As shown above, we need at least one instrumental variable for each endogenous variable contained in X , i.e., for each column of X . This means an instrument is needed for $\sum_{t'=t-M}^{t-1} y_{c'it'}$ for each c' , or in other words, each customer's added purchase probability in the past M periods.

Our proposed instrument is $y_{c'',i,t-M-1}$ for each c'' , which is the purchase probability $M+1$ periods ago for each different customer. To argue that this is a good instrument, we have to assess how it

satisfies the two IV conditions: 1) the instrument is uncorrelated with the error term, and 2) the instrument is strongly correlated with the endogenous variables. Firstly, the instruments $y_{c''i,t-M-1}$ have to be uncorrelated with the error terms ϵ_{cit} . By definition of the customer trend model, there is no effect of $y_{c''it'}$ on y_{cit} if $t' < t - M$, and hence $y_{c''i,t-M-1}$ and ϵ_{cit} are uncorrelated. Though, practically, this means that M still has to be selected large enough, because only then we can guarantee that there is no direct effect of the purchases in period $t - M - 1$ on current purchase decisions. Secondly, the instruments $y_{c''i,t-M-1}$ have to be strongly correlated with the endogenous variables $\sum_{t'=t-M}^{t-1} y_{c''it'}$. Under the same customer trend model definition, the purchase decisions in the periods $t - M$ to $t - 1$ were actually affected by the purchases made $t - M - 1$ periods ago. Yet, when implementing the model, we need to choose an M that is not too large, because otherwise the purchases in period $t - M - 1$ might not be strongly correlated with all the purchase decisions in periods $t - M$ through $t - 1$, especially the later periods.

Collectively, this shows that the memory parameter M needs to be selected carefully: large enough to ensure purchases beyond M periods ago do not affect the current purchase decision directly, but small enough to ensure that purchases beyond M periods ago still affect the recent purchase decisions. Later in Section 4.2.1, we test the strength of this instrument on the available Oracle Retail client data. Also in Section 2.2.1, we establish that our estimated customer-to-customer trend probabilities are consistent when using IV estimation.

Having explained the different estimation challenges and their remedies, we now describe the actual estimation procedure that is used to estimate the customer trend demand model (5). The Trend-Estimation algorithm runs as follows for a given structure on the base model $q(\cdot)$, memory parameter M , and regularization parameter λ :

ALGORITHM 1. (Trend-Estimation)

1. Fit the base model $q(x_{cit})$ to the purchase data y_{cit} by using an appropriate statistical or machine learning method for $q(\cdot)$. Let the estimated base model be denoted by \hat{q}_{cit} .
2. Fit the customer trend model to $y_{cit} - \hat{q}_{cit}$, the residuals of the base purchase model, by using the IV estimation method:
 - a. Fit the first stage model by solving the OLS regression problem:

$$\min_{0 \leq w_{c''c'} \leq M} \sum_{c'=1}^n \sum_{i=1}^m \sum_{t=M+1}^{T_i} \left(\sum_{t'=t-M}^{t-1} y_{c''it'} - \sum_{c''=1}^n w_{c''c'} y_{c''i,t-M-1} \right)^2. \quad (7)$$

Let the estimated customer-to-customer trend probabilities be denoted by $\hat{w}_{c''c'}$, and let the prediction of $\sum_{t'=t-M}^{t-1} y_{c''it'}$ be denoted by the following: $\bar{y}_{c'it} = \sum_{c''=1}^n \hat{w}_{c''c'} y_{c''i,t-M-1}$.

- b. Fit the final stage model by solving the LASSO-regularized bounded variables least squares regression problem:

$$\min_{0 \leq p_{c'c} \leq 1} \sum_{c=1}^n \sum_{i=1}^m \sum_{t=M+2}^{T_i} \left(y_{cit} - \hat{q}_{cit} - \sum_{c'=1}^n p_{c'c} \bar{y}_{c'it} \right)^2 + \lambda \sum_{c'=1}^n \sum_{c=1}^n p_{c'c}. \quad (8)$$

Let the estimated customer trend probabilities be denoted by $\hat{p}_{c'c}$.

The different steps of the Trend-Estimation algorithm follow the previous discussion. The complexity of estimating is addressed by distinguishing between step 1 where the base purchase model is estimated (purchase probability is standardized) and step 2 where the customer trend model is estimated. In step 2, the customer-to-customer trend probabilities are estimated by using the two-stage method used for IV models. Whereas step 2a simply mimics the first stage of IV estimation, step 2b differs from regular IV estimation. In particular, we introduce LASSO-regularization to the objective, and we constrain the parameter estimates to guarantee that the estimated probabilities lie between 0 and 1.

Before we estimate the model, we also need to determine the structure of the base model q , the memory parameter M , and the regularization parameter λ . To determine these parameters we use cross-validation, by splitting the dataset into three parts: the training, validation, and testing sets. We fit our model for numerous options of q , M , and λ on the training set, and compute the measure of fit on the validation set. Afterwards, we select the model with the q , M , and λ associated with the best fit on the validation set, and assess its fit on the testing set.

In Section 4, we assess the practical performance of a demand model accounting for customer trends by applying the Trend-Estimation algorithm to a real dataset. In the following Section 2.2.1, we analyze the theoretical estimation performance of the Trend-Estimation algorithm described in Algorithm 1.

2.2.1. Statistical Consistency. We establish, under mild assumptions, that the customer-to-customer trend probabilities are estimated consistently, i.e., the estimates $\hat{p}_{c'c}$ converge to the true values $p_{c'c}$ as the number of observations increases. For this result, we need to assume that the defined instrument $y_{c'',i,t-M-1}$ is a valid instrument. Additionally, we need to assume that the base purchase model is estimated consistently, i.e., the estimate \hat{q}_{cit} converge to the true values q_{cit} as the number of observations increases. In our application, maximum likelihood estimation of a logistic regression results in consistent estimates (Amemiya 1985).

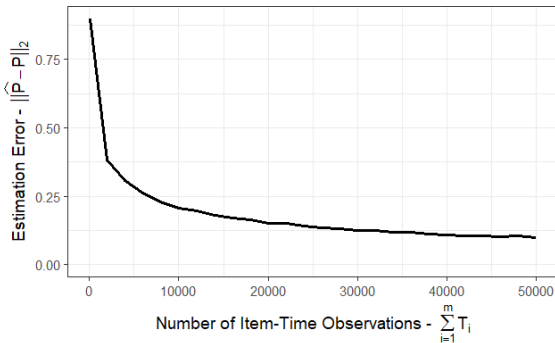
Under our assumptions, Theorem 1 below shows that the Trend-Estimation algorithm yields consistent estimates of the customer-to-customer trend probabilities. This result demonstrates that, for growing datasets, our instrumental variables based estimation procedure is able to capture the causal effect that one customer's past purchases have on another customer's purchase decision.

THEOREM 1. *Assume that $y_{c'',i,t-M-1}$ is a valid instrument and that the base purchase probability is estimated consistently. Let $\lambda = 0$ and consider the 'true' model (5), then the Trend-Estimation algorithm estimates the customer-to-customer trend probabilities consistently as*

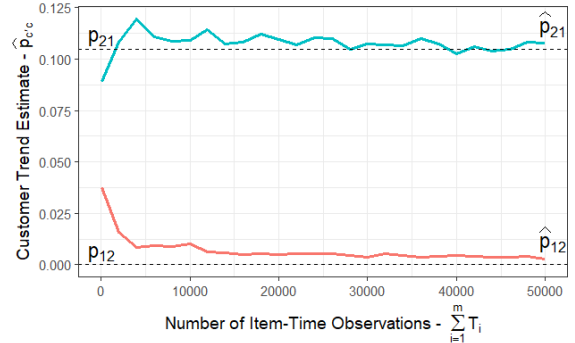
$$\hat{p}_{c'c} \rightarrow p_{c'c} \text{ as } \sum_{i=1}^m T_i \rightarrow \infty.$$

This theorem is proven in Appendix A.2.

In Figure 3 we present computational examples of the consistency of the estimates. These examples consider 8 customers and vary the number of observations per customer. We simulate a transaction dataset for each number of observations. With this dataset we estimate the customer-to-customer trend probabilities that are depicted in the figures. Figure 3a shows that the estimation error, measured using the Euclidean distance between the estimates and ‘true’ customer-to-customer trend probabilities $\|\hat{P} - P\|_2 = \sqrt{\sum_{c'=1}^n \sum_{c=1}^n (\hat{p}_{c'c} - p_{c'c})^2}$, decreases as the number of observations increases. Notably, the estimation error shrinks rapidly for a small number of observations. This indicates that even a small dataset can generate good estimates. Figure 3b presents two examples of customer trend estimates for differing dataset sizes. The red (lower) curve shows the estimated customer-to-customer trend probability from customer 1 to customer 2, while the blue (upper) curve shows the estimate from customer 2 to customer 1. The black dotted curves show the ‘true’ value of p_{12} and p_{21} . Confirming the result of Theorem 1, the estimates converge towards their respective ‘true’ values.



(a) Estimation error of customer trend estimates



(b) Examples of customer trend estimates

Figure 3 Figures showing the convergence of the customer-to-customer trend probability estimates to their respective ‘true’ values as the number of observations increases

3. Dynamic Promotion Targeting Optimization Problem

In this section, we illustrate how retailers can exploit the customer trend network to maximize their profit. In particular, we study the Dynamic Promotion Targeting Optimization Problem. A dynamic promotion targeting policy is a marketing tool, in which a select group of customers may receive a temporary price reduction for some items. To motivate the Dynamic Promotion Targeting Optimization Problem, we consider the following setting. Let us assume that at a given time t , a customer c receives a temporary price reduction for item i . The retailer expects to see that this temporary price reduction will increase the chances that customer c will buy item i at time t . Due

to the trend effect, the retailer also expects to see a contagion effect, where in the periods following time t all of the customers who are linked to customer c , will exhibit a higher tendency to buy item i . The result is an increase in profit from sales. Thus, we wish to answer the question of which customers should receive a temporary price reduction, for which item, and when?

3.1. Problem Description

We consider a retailer who sells m products over T time periods to a group of n known customers, that are embedded in a customer trend network. Here we assume that the network is known in advance, or was estimated using the procedure described in Section 2. In this problem, the retailer may target a specific customer, or group of customers, for a temporary promotion, in order to maximize the retailer's revenue.

When we consider customers who are embedded in a trend network, the impact of a temporary promotion may have a ripple effect. In other words, when the retailer deliberates whether to give a targeted promotion to a customer, it is important to assess not only the impact of this promotion on the specific customer, but also the impact that this customer's purchase decision may have on other customers in the future.

Throughout the rest of this section we use the following notation. We use ϕ to denote a temporary promotion. A promotion $\phi = (c_\phi, i_\phi, t_\phi)$ indicates that customer c_ϕ was offered a promotion on item i_ϕ at time t_ϕ . A promotion policy Φ is a set of promotions. The discounted price that is offered to a customer c for item i at time t , according to policy Φ , is set to be the discounted price of the item, denoted by d_i , if there exists a promotion $\phi \in \Phi$ such that $c_\phi = c, i_\phi = i$ and $t_\phi = t$. Otherwise, the price that is offered to a customer c for item i at time t , according to policy Φ , is set to be the full price of the item, denoted by d_i^0 .

With a slight abuse of notation, we refer to b_{cit}^Φ and q_{cit}^Φ as the purchase and base purchase probabilities respectively, given a promotion policy Φ . In a similar way, r_{cit}^Φ denotes the price that was offered to customer c for item i at time t , according to the promotion policy Φ .

The retailer's objective is to determine a promotion policy Φ that maximizes the total revenue

$$R(\Phi) = \sum_{i=1}^m \sum_{c=1}^n \sum_{t=1}^T r_{cit}^\Phi N_c b_{cit}^\Phi \quad (9)$$

In the case where each customer c represents a set of consumers, N_c denotes the cardinality of the set c . Otherwise, if each customer c represents a single customer, we set N_c to be 1.

When planning a promotion targeting policy, retailers must also consider existing inventory levels. In the case where a given item is in demand but out of inventory, retailers may have to backorder the item. For example, retailers may offer the customer a free shipping of the item to the customer's house. In this case, the retailer will incur the shipping cost. We refer to these costs as backorder penalties.

In order to capture the backorder penalty cost, we assume that for each item i , each store location l has an initial inventory of I_{il} units. Each customer c is associated with a preferable store $\ell(c)$. If a customer arrives to store l , and faces a stock-out, the store can order the item, and ship it to the customer at a cost of s_{il} . This means that the cost of backorder given a promotion policy Φ is given by

$$\sum_{i=1}^m \sum_{l=1}^{NL} \left(\sum_{c:\ell(c)=l} \sum_{t=1}^T N_c b_{cit}^{\Phi} - I_{il} \right)^+ \cdot s_{il}$$

where NL is the number of store locations.

Finally, incorporating this cost into the revenue of equation (9), the retailer's objective becomes,

$$R(\Phi) = \sum_{i=1}^m \sum_{c=1}^n \sum_{t=1}^T r_{cit}^{\Phi} N_c b_{cit}^{\Phi} - \sum_{i=1}^m \sum_{l=1}^{NL} \left(\sum_{c:\ell(c)=l} \sum_{t=1}^T N_c b_{cit}^{\Phi} - I_{il} \right)^+ \cdot s_{il}$$

In practice, a promotion plan is subject to limitations that are dictated by the retailer. One such example is a restriction on the number of promotions, in which for a given time horizon, the retailer limits the number of promotions. In other words, the retailer commits to a maximum of L promotions during the planning horizon (see constraint (10d) in the formulation below).

As a result, the optimal policy for the Dynamic Promotion Targeting Optimization Problem can be found by solving the following Non-Linear Mixed-Integer Programming formulation

$$\max_{\vec{r}, \vec{b}, \vec{\gamma}} \sum_{i=1}^m \sum_{c=1}^n \sum_{t=1}^T (r_{cit} N_c b_{cit}) - \sum_{i=1}^m \sum_{l=1}^{NL} \left(\sum_{c:\ell(c)=l} \sum_{t=1}^T N_c b_{cit} - I_{il} \right)^+ \cdot s_{il} \quad (10a)$$

$$\text{s.t. } r_{cit} = d_i^0 (1 - \gamma_{cit}) + d_i \gamma_{cit}, \quad \forall c, i, t \quad (10b)$$

$$b_{cit} = q_{cit}^{\Phi(\vec{r})} + \sum_{c'=1}^n \sum_{t'=t-M}^{t-1} b_{c'it'} p_{c',c}, \quad \forall c, i, t \quad (10c)$$

$$\sum_{i=1}^m \sum_{c=1}^n \sum_{t=1}^T \gamma_{c,i,t} \leq L \quad (10d)$$

$$\gamma_{cit} \in \{0, 1\} \quad \forall c, i, t \quad (10e)$$

where $\Phi(\vec{r})$ is the promotion policy dictated by the price vector \vec{r} .

Unfortunately, the formulation in (10) describes an NP-hard problem. A proof of the complexity of the problem can be found in Appendix C.2. Nevertheless, the following section suggests an Adaptive-Greedy algorithm that can find a near-optimal solution to the problem formulation (10). We show an analytical bound on the optimality gap achieved by this algorithm, and show two special cases in which this bound can be further improved.

3.2. Adaptive-Greedy Algorithm

Having described the Dynamic Promotion Targeting Optimization Problem, we now propose a fast Adaptive-Greedy algorithm that can find a near-optimal promotion targeting policy. We also provide a novel approach for the analysis of greedy algorithms, that allows us to bound the worst case optimality gap. In general, greedy algorithms provide fast and intuitive solutions. As described in Nemhauser et al. (1978), for maximization problems of a non-negative monotone submodular function subject to cardinality constraints, a greedy algorithm can find a near-optimal solution. Namely, a solution that is within a $(1 - \frac{1}{e})$ factor of the optimal solution. Nevertheless, in the general case, the optimality gap of the greedy algorithm is unbounded. In this section, we provide an extension to the results in Nemhauser et al. (1978). We identify a wide family of functions, referred to as *C-Submodular* functions, for which the greedy algorithm can find a near-optimal solution when subject to a cardinality constraint.

This section is outlined as follows. First, we describe the suggested Adaptive-Greedy algorithm for the Dynamic Promotion Targeting Optimization Problem. Then, we introduce the concept of *C-Submodular* functions, and use it to provide an analytical bound on the optimality gap of the algorithm. In Section 3.2.1, we show that under mild conditions, the revenue function in (10), becomes submodular, and the bound on the optimality gap can be reduced to $(1 - \frac{1}{e})$.

We begin by introducing the suggested Adaptive-Greedy algorithm for the Dynamic Promotion Targeting Optimization Problem.

ALGORITHM 2. (Adaptive-Greedy)

1. Initialize an empty promotion plan $\Phi = \emptyset$;
2. Repeat L iterations,
 - (a) For every promotion $\phi = (c_\phi, i_\phi, t_\phi)$, calculate the potential increase in revenue from adding ϕ to the promotion plan Φ

$$\rho(\phi, \Phi) = R(\Phi \cup \phi) - R(\Phi)$$

- (b) Choose the promotion ϕ with maximum potential increase in revenue, and add it to the promotion plan Φ .

In Section 4, we compute and test the Adaptive-Greedy algorithm on real data. In the remainder of this section, we provide an analysis for the worse-case performance of the algorithm. In order to show an analytical bound on the optimality gap of the Adaptive-Greedy algorithm described in Algorithm 2, we introduce the following definition.

DEFINITION 1 (*C-Submodular* FUNCTION). A function f is called *C-Submodular*, if for every X, Y with $X \subset Y$ and every element $x \notin X$,

$$f(Y \cup \{x\}) - f(Y) \leq C + f(X \cup \{x\}) - f(X).$$

This definition shares some similarity with the definition of K -convex functions. In inventory control theory, K -convex functions are non-linear functions with bounded level of non-convexity. K -convex functions are a weaker version of a convex function. Similarly, C -Submodular functions provide a weaker version to the concept of a submodular function. A C -Submodular function f , is not submodular, however the level of non-submodularity is bounded by the parameter C . Specifically, $C = 0$ in the case where the function f is submodular.

Next, Theorem 2, presents a parametric analytical bound on the optimality gap of the adaptive greedy algorithm, that is, a function of the parameter C , and the cardinality parameter L .

THEOREM 2. *Let f be a monotone C -Submodular non-negative function. The adaptive greedy algorithm, which starts with S as the empty set and for L (where L is the limit on the cardinality of the set) iterations picks an element x which maximizes the marginal benefit $f(S \cup \{x\}) - f(S)$, provides a set S of size L that achieves an objective of at least $(1 - \frac{1}{e})f(S^*) - LC(1 - \frac{1}{e \cdot (L-1)})$, where S^* is the optimal set.*

The proof of Theorem 2 can be found in Appendix B.1.

In the case of the Dynamic Promotion Targeting Optimization Problem that was described in problem (10), the limit on the cardinality of the set, L , is the limit on promotions over the time horizon. The non-submodularity parameter C is characterized in Theorem 3 below.

δ is bound on the level of substitution between the different items. This value captures the loss in demand for a specific item i , given that another item \tilde{i} was given a promotion. Formally, δ can be calculated as follows.

$$\delta = \max_{c, i, t, \Phi, \tilde{\phi}: t_{\tilde{\phi}} = t} \left\{ b_{cit}^{\Phi} - b_{cit}^{\Phi \cup \{\tilde{\phi}\}}, 0 \right\},$$

$\pi^*(P)$ is a parameter of the trend network. For tree structured networks, this parameter is equal to the highest trend in the network. Otherwise, in the case of a cyclic network, this value grows in the number and length of the cycles in the network. This value is defined as follows.

$$\pi^*(P) = \max_{c, c', t} \left\{ \sum_{\pi \in P_t(c, c')} \prod_{j=1}^{|\pi|-1} p_{\pi_j, \pi_{j+1}} \right\}.$$

C_q captures the level of non-submodularity of the base demand function q . in general, C_q is defined to be the lowest non-negative value such that for each Ψ, ψ and $\tilde{\psi}$,

$$(q_{c_{\tilde{\psi}}it_{\tilde{\psi}}}^{\Phi \cup \{\phi, \tilde{\phi}\}} - q_{c_{\tilde{\psi}}it_{\tilde{\psi}}}^{\Phi \cup \{\phi\}} - q_{c_{\tilde{\psi}}it_{\tilde{\psi}}}^{\Phi \cup \{\tilde{\phi}\}} + q_{c_{\tilde{\psi}}it_{\tilde{\psi}}}^{\Phi}) \leq C_q$$

$N_{c_{max}}$ is the cardinality of the largest group of customers that are aggregated into a single group. In the case of promotion targeting to individual customers, $N_{c_{max}} = 1$. The main significance of this

parameter is in the case where there is a large variability in the sizes of the customer groups. If the customers are all of a similar size, this value will have a similar linear impact on the optimal revenue. Nonetheless, in the case in which the customer groups are of different sizes, the optimal revenue will not grow linearly in the size of the biggest customer group. This implies that in the case where there is a big variability in the sizes of the customer groups, the bound on the optimality gap will decrease.

THEOREM 3. *If the base purchase probability q is a C_q -submodular function, then the revenue function in (10) is C -Submodular, with*

$$C = L \left(2d_i^0 \delta N_{c_{max}} + m d_i^0 C_q N_{c_{max}} + m(T-1)\pi^*(P)C_q d_i^0 \sum_{c=1}^n N_c + \delta T \sum_{l=1}^{NL} s_{il} \sum_{c:\ell(c)=l} N_c \right),$$

The proof of Theorem 3 can be found in Appendix C.3, and is based on the analysis in Appendix B.2.

In order to illustrate the optimality bound in Theorem 3, we preform a synthetic experiment that show the impact the different parameters on the optimality gap. Table 1 summarizes the results of this experiment. The examples in Table 1 were built in order to illustrate the bound in Theorem 3. In the case in which $C = 0$, the lower bound on the optimality gap is equal to 63.21%. Hence, the values in Table 1 are upper bounded by 63.21%. In the worst case, in which the base probability model is not submodular, and in addition, the cross item effect is negative, that is, the items have substitution effect, the lower bound on the optimality gap is as low as 40%. Nonetheless, in all of the cases, even when the lower bound on the revenue was relatively low, the algorithm found a solution that captures at least 96% of the optimal revenue. It is worth mentioning that while in all the cases in Table 1, the algorithm finds a sub-optimal solution, in the vast majority of the cases that were tested, the greedy algorithm, in fact, found the optimal solution.

Table 1 **Optimality ratio on the algorithm's revenue and the lower bound's revenue for different settings of the customer-trend demand model**

Parameters				Optimality Ratio	
C_q	δ	π	C	Revenue of Algorithm	Revenue Lower Bound
0	0	0.75	0	0.9728	0.6321
0	0	0.75	0	0.9745	0.6321
0.05	0	0.1	1000	0.9835	0.4443
0.05	0	0.13	1150	0.9816	0.4291
0	0.61	0.1	2440	0.9690	0.5210
0	0.61	0.2	2440	0.9667	0.5232
0.1	1	0.1	4960	0.9803	0.4054
0.1	1	0.2	5320	0.9789	0.3937

Next, in Section 3.2.1, we show that for a special case, $C = 0$, and therefore, the function is simply submodular. This case arises, for example, when all products are complements. In these cases, the greedy algorithm finds a $(1 - \frac{1}{e})$ approximation.

3.2.1. Submodularity of the Revenue Function for Complementary Items. Many retailers use targeted promotions in order to increase store traffic, and as a result, increase revenue from sales. This practice is based on the notion that targeted promotions draw customers into the store in order to buy a particular item, however once these customers are in the store, they are very likely to buy other items as well. This type of cross-item effect is referred to as the halo effect, in which the demand for a specific item is increased when the price of another item is decreased. Items with such impact on one another are referred to as complementary. In this section we consider the special case of the Dynamic Promotion Targeting Optimization Problem with complementary items. We show that for this type of problem, under some minor conditions, the revenue function in problem (10) becomes submodular.

The following theorem characterizes a condition on the base demand function q , under which the revenue function in problem (10) is submodular.

THEOREM 4. *In the case where (a) the base purchase probability function is a submodular function of the promotion policy, and (b) the items are complementary, the revenue function in problem (10) is submodular.*

The proof of Theorem 4 can be found in Appendix B.2.

In Section 2, we suggested to use the logistic regression in order to model the base demand function. In the case of complementary items, in the domain $[0.5, 1]$, the logistic regression is indeed submodular. Therefore, for base probabilities that are above 0.5, the condition in Theorem 4 is satisfied.

The following proposition describes how the submodularity leads to an analytical guarantee on the optimality gap.

PROPOSITION 1. *In the case where (a) the base purchase probability function is a submodular function of the promotion policy, and (b) the items are complementary, the Adaptive-Greedy algorithm in Algorithm 2 finds a solution that is at least $(1 - \frac{1}{e})$ of the optimal solution.*

Proposition 1 follows directly from Theorem 2.

Another special case of the Dynamic Promotion Targeting Optimization Problem is one in which the retailer chooses a promotion period t a priori. We refer to this case as the Single-Period Dynamic Promotion Targeting Optimization Problem. In this case, the retailer requires that all of the promotions will be offered during the promotion period t . In any other period, the prices are set to be the regular price. This requirement can be formulated by adding the following constraints into the

problem formulation (10): $\sum_{i=1}^m \sum_{c=1}^n \gamma_{cit} \leq L$ and $\gamma_{ci\tau} = 0, \forall \tau \neq t$. In this case, we can identify a set of problems for which the revenue function in problem (10) becomes linear. Hence, for these cases, the Adaptive-Greedy algorithm in Algorithm 2 finds the optimal solution. This result is summarized in the Corollary below.

COROLLARY 1. *In the case of the Single-Period Dynamic Promotion Targeting Optimization Problem, with $m = 1$, and no shipping cost, the Adaptive-Greedy algorithm in Algorithm 2 can find the optimal solution.*

The proof of Corollary 1 follows directly from the submodularity proof in Appendix B.2.

4. Results from a Large Fashion Retailer

In Sections 2 and 3 we described the demand prediction and Dynamic Promotion Targeting Optimization models. By working with Oracle Retail, we have access to data from a large fashion retailer. Using this data, we test our estimation process and the effectiveness of our optimization approach in terms of revenue increase relative to the retailer’s existing practice.

In the first part of this section, we describe the dataset that is used in this case study. In the second part of this section, we illustrate the impact of the customer trend model described in Section 2 on prediction accuracy. Additionally, we test the validity of the instruments. This test shows that the instruments are strong and thereby, we can regard the estimated customer-trend graph to describe the causal relationship, which is necessary to target promotions correctly. In the last part of this section, we illustrate the potential increase in revenue from using the Adaptive-Greedy algorithm described in Section 3. We compare the revenue of our algorithm to that of the pricing policy that the retailer used in practice.

4.1. Data Description

In order to test the estimation and optimization models that were suggested in Sections 2 and 3, we use data from a large fashion retailer. This retailer operates 635 stores in the United States, and Puerto Rico, with total revenues of two billion dollars a year.

The dataset that we present in this paper includes data that was collected over a period of roughly two years from 20 stores in the state of Ohio. In addition, to analyze the robustness of our results, we also have access to a smaller dataset from 4 stores in the state of Utah. These transaction datasets contain customer linked purchases. Each purchase is associated with a customer identifier, store location, item identifier, item’s color, item’s style, date of purchase, regular price of the item, and the price that was paid for the item in this purchase. In order to prevent biases due to the sparsity of the data, we aggregate the customers on their demographic features as well as the location in which they made their purchases. For the same reason, we aggregate items on the style level, and

time periods into weeks. We refer to Martec International, Inc (2011) for a detailed description of the style level.

In many cases the life cycle of a style varies between the different stores. For example, consider a store that often receives inventory earlier than other stores. Many of the styles in this store will have a life cycle that starts earlier than the same style in other stores. In this case, the customers who visit the first store would be more likely to buy earlier. As a result, our model could incorrectly predict a customer trend effect emanating from these customers. In order to prevent biases due to these differing life cycles, we focus on a set of styles that arrived to all of the relevant locations at the same time. For this we use information about the available inventory of each style in each store.

In the following sections we present our results for two departments: Mens Knit Tops and Womens Dressy Tops. As an example, the size of each dataset for Ohio is as follows:

1. Mens Knit Tops: the data in this department includes around 183,000 purchases of 324 different styles that were made by at least 31,123 different customers.
2. Womens Dressy Tops: the data in this department includes around 185,000 purchases of 402 different styles that were made by at least 42,681 different customers.

4.2. Accuracy of the Customer Trend Demand Model

Here, we illustrate the impact of the customer trend demand model described in Section 2 on our dataset. As previously discussed, we assume that the base purchase probability is described by a logistic regression. Thus, the base purchase probability of (aggregated) customer c for an item i at time t , given a promotion policy Φ , can be described as follows,

$$q_{cit}^{\Phi} = \frac{1}{1 + \exp\left(-(\beta_0 + \beta_1 WEEK_t + \beta_2 LOCATION_c + \beta_3 r_{cit}^{\Phi} + \sum_{i' \neq i} \beta_4^{(i')} r_{ci't}^{\Phi})\right)},$$

where $WEEK_t$ is an indicator variable for the week number associated with the time period t and $LOCATION_c$ is an indicator variable for the store location associated with customer c .

The parameters in the model above capture the following effects. First, the intercept β_0 captures the baseline probability. The seasonality effect on the weekly level is captured by β_1 , a vector of length 52. The effect of the location of the store is captured by β_2 , a vector of length 20. The response of the purchase probability to the item's price is captured by β_3 . Last, the cross item effects are captured by the values of $\beta_4^{(i')}$. Including these variables in the base purchase probability model, allows our customer trend demand model to capture a variety of classical demand features.

Given the logistic base model, we estimate the customer-trend demand model by using the Trend-Estimation algorithm described in Algorithm 1. When fitting the model, we use cross-validation to determine the values of the memory factor M and the regularization factor λ . The customers and items in the validation set are selected randomly, and the results in this section reflect the mean of 10 random splits.

To evaluate the quality of the prediction model we consider the Weighted Mean Absolute Percentage Error (WMAPE). The WMAPE gives a weighted measure of the relative prediction error. The weight of each error depends on the volume of the demand. In particular, this measure gives a higher weight to observations where demand was higher. More formally, the WMAPE is given by

$$WMAPE = \frac{1}{\sum_{t=1}^n A_t} \sum_{t=1}^n A_t \cdot \left| \frac{A_t - F_t}{A_t} \right|,$$

where A_t denotes the actual demand, and F_t denotes the forecasted demand at time t .

Table 2 presents the out-of-sample WMAPE of the predictions from the base demand model and the customer trend demand model, for the two different departments and states. Table 2 shows that incorporating the customer trend in the demand model leads to an absolute error improvement between 2 to 7 percentage points (on average 5), which translates to a relative error improvement of 4% to 16% (on average 11%). These forecasting metrics are good when predicting fashion clothing sales, where often sales are slightly harder to predict due to the irregularity of sales.

Table 2 Comparison of out-of-sample WMAPE of the base demand model and the customer trend demand model for different departments in the states of Ohio and Utah

Department	State	Demand Model		
		Base	Customer Trend	Relative Improvement
Mens Knit Tops	Ohio	0.5213	0.4683	10.2%
	Utah	0.5024	0.4845	3.6%
Womens Dressy Tops	Ohio	0.4548	0.3951	13.1%
	Utah	0.4322	0.3636	15.9%

In Figure 4, we illustrate how the prediction accuracy (measured in out-of-sample WMAPE) changes as a function of the memory M and the regularization parameter λ . Figure 4a shows that as the memory M increases, the prediction error initially decreases, and then stabilizes. Note that the case of no memory, $M = 0$, describes the base demand model, as there is no memory in the base model. Hence, the drastic initial decrease emphasizes the importance of including the customer-to-customer trend effect in the model. Accounting for the customer trend contributes greatly to understanding the purchase behavior of customers. Figure 4b indicates that the relation between the regularization parameter λ and the prediction error is not immediately clear. We observe that a larger regularization parameter results in larger errors for the womens datasets, in contrast to the smaller errors for mens datasets. Though, mostly, the error does not change significantly if the regularization parameter is chosen wrongly. However, this still indicates that tuning the regularization parameter is important. We remark that the regularization parameter also relates to how sparse the estimated customer-to-customer trend network is: larger regularization results in sparser networks. Generally, these graphs indicate that the mens network is sparser, while more customer relationships in the womens network help with predicting.

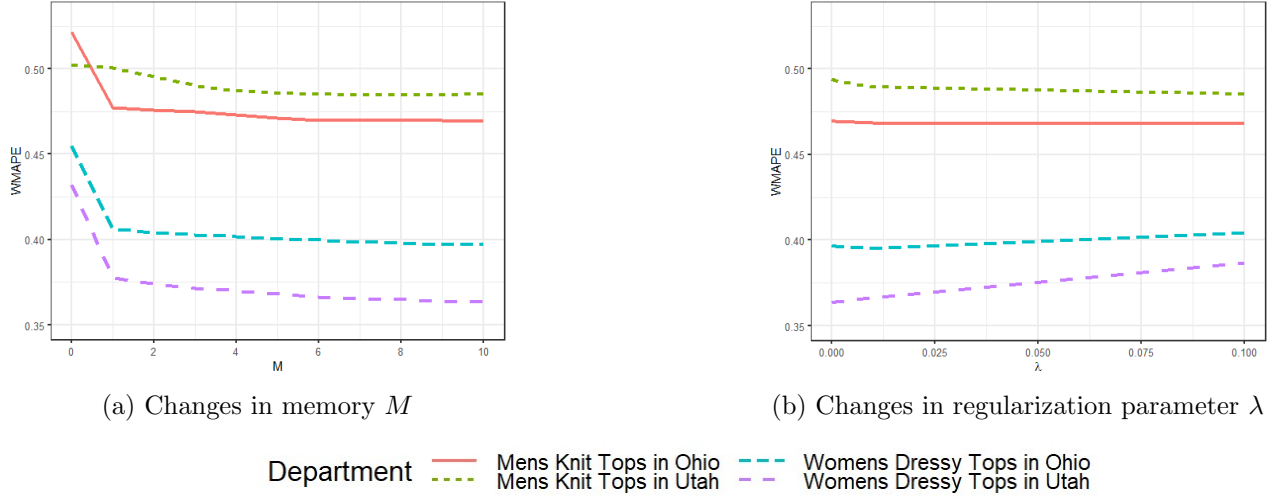
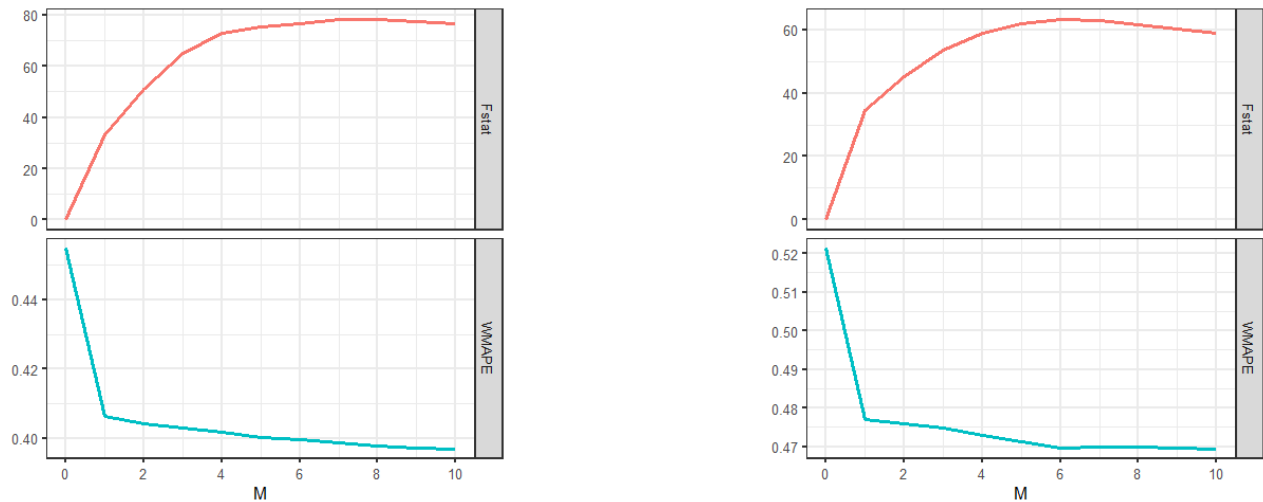


Figure 4 Figures showing the out-of-sample WMAPE for different departments in the states of Ohio and Utah as the memory and regularization parameter increase

4.2.1. Causal Analysis. Next, we examine the validity of the instruments proposed in Section 2, and hence, whether the estimates can be presumed to be causal. The validity of the instruments depends on whether they satisfy the two IV conditions. The first of these two conditions, that the instrument is uncorrelated with the error term, can not be tested in the standard linear regression model where the number of endogenous variables equals the number of instruments (Cameron and Trivedi 2005). However, as our model differs from the standard linear regression model, we can compute the sample correlation of the instruments and the error term. For our dataset, the largest sample correlation roughly equals 0.0001, which provides strong evidence for the assumption that the correlations between the instruments and errors are small.

The second condition, that the instruments are strongly correlated with the endogenous variables, can be tested through F-tests on the first-stage model (Cameron and Trivedi 2005). This entails fitting the linear regression presented in equation (7) and applying an F-test on whether the instruments have any explanatory power. In particular, the F-test considers the hypothesis whether $w_{c''c'} = 0$ for all c'' and c' , and essentially tests whether the respective estimates $\hat{w}_{c''c'}$ are too far from 0. Generally, if the F-statistics exceed 10, there is significant evidence that the instruments are strongly correlated with the endogenous variables. In Figure 5 we illustrate the F-statistics and the WMAPE resulting from our dataset for different levels of the memory parameter M .

Most importantly, we note that the F-statistic is considerably larger than 10 for any reasonable value of the memory parameter M (i.e., $M > 0$). This indicates that the instruments satisfy the condition that they need to be strongly correlated with the endogenous variables. In addition, we note that the F-statistic first increases as M increases and later slowly decreases. Especially, the initial rapid increase signifies that $y_{c'',i,t-M-1}$ becomes a stronger instrument for $\sum_{t'=t-M}^{t-1} y_{c'it'}$ as



(a) Statistics for Womens Dressy Tops

(b) Statistics for Mens Knit Tops

Figure 5 Figures showing the F-statistic of the first-stage regression and the out-of-sample WMAPE for two departments in the state of Ohio as the memory increases

M increases. Hence, purchases are not only good at explaining the next period’s purchases, but even more so at explaining purchases during the M periods ahead. Beyond this point the F-statistic stabilizes, indicating that these past purchases are not adding significant explanatory power further into the future. As we would expect, the explanatory power of purchases diminishes after 6 weeks. To select the right M , we see that the F-statistic is high (above 70 and 60 respectively) and the WMAPE is low (below 40% and 47% respectively) around 4 to 6 weeks. Hence, a good selection for the memory parameter would be between 4 to 6 periods.

4.3. Power of Dynamic Promotion Targeting Optimization

Next, we conduct an experiment with the goal of computing and testing the Adaptive-Greedy algorithm described in Algorithm 2 in Section 3. For a fair comparison, we compare the targeting policy that was found by the Adaptive-Greedy algorithm with the targeting policy that was used by the retailer based on *virtual profits*. These virtual profits are obtained by plugging the promotion policy (result of Adaptive-Greedy or current) into the customer trend demand model (5) to obtain the virtual demand, and then multiplying this virtual demand by the prices in the corresponding promotion policy. Essentially, this means that we assess the retailer’s policy on the same demand model as the policy of the Greedy-Adaptive algorithm.

To run the Adaptive-Greedy algorithm, we need to determine the parameters of the problem instances, and need to ensure they resemble the real-world instance. For each department and differing time horizons, we create an instance by selecting a set of representative items. From the dataset,

we know the number of customers in each group, the inventory of each item and location, the shipping cost, and the regular price of each item. Additionally, we compute the average promotion price that was used by the retailer for each item, and the number of promotions that were offered.

The results of these experiments are shown in Table 3. The items in the selected set all lasted for 20 weeks. To analyze our policy’s effectiveness over time, we compare the virtual profits at 10, 15, and 20 weeks. This table shows that, on average, the policy generated by our Adaptive-Greedy algorithm improves virtual profits by 3-11% relative to the policy that the retailer originally implemented. As in the accuracy of the demand model, the customer trend effect is more impactful in the womens dataset than the mens dataset. On average, at any time horizon, the profit improvement of the womens dataset is roughly 5% larger than the mens dataset.

Noticeably, the revenue improvement is monotonically increasing with respect to time for the mens dataset, because there is more time to optimize over. However, for the womens dataset there is a small decrease when optimizing over 20 weeks instead of 15 weeks. Likely, this is due to the fact that the Adaptive-Greedy algorithm assigned 13 promotions to the last 5 periods, compared to the 18 promotions that were offered by the retailer.

Finally, comparing the solution of each policy, we observe that the Adaptive-Greedy algorithm shifted a large number of promotions from one item (that the retailer promoted) to another item (that the Adaptive-Greedy algorithm preferred). The first item received 81 promotions from the retailer, it received 33 promotions from the Adaptive-Greedy algorithm. The second item received 77 promotions from the Adaptive-Greedy algorithm, and only 29 promotions from the retailer. This signifies that the Adaptive-Greedy algorithm detects the item that is more price-sensitive and for which the customer trend effect is more impactful.

Table 3 Comparison of virtual profits gained by Adaptive-Greedy algorithm and retailer’s policy for different departments and time horizons in Ohio

Department	Time Horizon	Virtual Profits		
		Adaptive-Greedy	Retailer’s Policy	Relative Improvement
Mens Knit Tops	10	1294.71	1248.83	3.7%
	15	1929.89	1848.94	4.4%
	20	2459.36	2333.13	5.4%
Womens Dressy Tops	10	334.17	309.10	8.1%
	15	501.68	453.56	10.6%
	20	665.33	604.95	10.0%

4.4. Insights and Discussion

In the following, we emphasize some of the insights from the previous discussion, and give some additional insight into the estimated customer trend and the resulting promotion policy. The estimation of the demand model validated the addition of the customer trend in two ways. Primarily,

the out-of-sample WMAPE improved significantly. This demonstrates that the demand model is more generalizable into the future with versus without the customer trend effect. In another test for causality, we showed that the proposed instruments satisfy the IV conditions. The F-test of the first stage regression strongly rejects the hypothesis that the instruments and the endogenous variables are uncorrelated. Furthermore, we note that there is a natural value for the right length of the memory. Both the WMAPE and the F-test support the notion that the memory lasts for around 4 to 6 weeks long. Interestingly, we observe that including the customer trend effect is more impactful in the women's fashion dataset than in the men's fashion dataset, signifying a trendier female customer base.

Next to significant improvements in prediction accuracy and promotional profits, the model also allows us to draw interesting insights into a customer base. The main reason being that the customer trend model can be visualized as a network. The graphical representation of the customer trend model shows how strongly customers affect each other. This visualization can help retail managers to understand the targeted promotion policy. In Figure 6, we present the customer trend network estimated from the Womens Dressy Tops dataset in Ohio. The vertices in this network indicate customer groups associated with a particular store location. Missing edges indicate no relationship between the two respective locations, while the thinner edges in the figure symbolize smaller customer-to-customer trend probabilities between 0 and 0.1, and thicker edges represent larger customer-to-customer trend probabilities between 0.1 and 0.5.

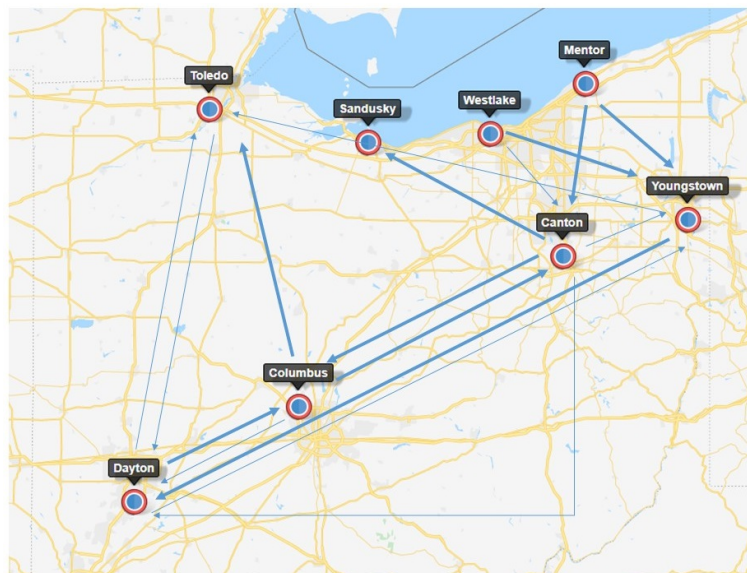


Figure 6 Customer trend network for the state of Ohio in which vertices are stores and the thickness of edges illustrates the strength of the customer-to-customer trend

From Figure 6, we observe that the network is roughly split into 2 segments. In the northeastern cluster, the customer trend largely originates from the Cleveland area (Westlake and Mentor), while Columbus and Dayton form the western cluster. Remarkably, the smaller city of Mentor has a large effect on the larger city Youngstown. This effect can be explained by the high social status of Mentor compared to the low social status of Youngstown, as Data USA (2019) depicts median incomes of \$70,625 and \$26,295, and poverty rates of 5.34% and 36.8% respectively. According to economics and sociology literature (Rege 2008), lower social classes wish to elevate their perceived social status as it presents more economic and social opportunities. This explains, for example, that the lower class in Youngstown mimics the fashion clothing of the higher social class in Mentor.

The Adaptive-Greedy algorithm also notices this as it mostly gives promotions to Canton, Columbus, Mentor, and Westlake. For the reason that Mentor and Westlake have a strong customer trend effect on Youngstown, the large city Youngstown does not receive a promotion, yet the smaller city Mentor does. This means that in order to get Mentor, Westlake, and Youngstown to purchase, we promote the less central locations that have an effect on a central location.

Altogether, the demand model's accuracy and the optimization model's profits, combined with the ease-of-understanding have intrigued retail managers. As a result, Oracle Retail executives work on productizing our models. The implementation of these models has a significant impact on Oracle Retail clients, and more generally on the retail industry.

References

- Akhlaghpour, Hessameddin, Mohammad Ghodsi, Nima Haghpanah, Vahab S Mirrokni, Hamid Mahini, Afshin Nikzad. 2010. Optimal iterative pricing over social networks. *International workshop on internet and network economics*. Springer, 415–423.
- Alves, Helena, Cristina Fernandes, Mário Raposo. 2016. Social media marketing: a literature review and implications. *Psychology & Marketing* **33**(12) 1029–1038.
- Amemiya, Takeshi. 1985. *Advanced econometrics*, chap. 9. Harvard university press, 270–274.
- Aral, Sinan, Lev Muchnik, Arun Sundararajan. 2009. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences* **106**(51) 21544–21549.
- Aral, Sinan, Dylan Walker. 2011. Creating social contagion through viral product design: A randomized trial of peer influence in networks. *Management science* **57**(9) 1623–1639.
- Arthur, David, Rajeev Motwani, Aneesh Sharma, Ying Xu. 2009. Pricing strategies for viral marketing on social networks. *International workshop on internet and network economics*. Springer, 101–112.
- Atazky, Ron, Ehud Barone. 2015. Advertising and incentives over a social network. US Patent App. 14/560,931.
- Bakshy, Eytan, Itamar Rosenn, Cameron Marlow, Lada Adamic. 2012. The role of social networks in information diffusion. *Proceedings of the 21st international conference on World Wide Web*. ACM, 519–528.
- Bapna, Ravi, Akhmed Umyarov. 2015. Do your online friends make you pay? a randomized field experiment on peer influence in online social networks. *Management Science* **61**(8) 1902–1920.

- Beheshti-Kashi, Samaneh, Hamid Reza Karimi, Klaus-Dieter Thoben, Michael Lütjen, Michael Teucke. 2015. A survey on retail sales forecasting and prediction in fashion markets. *Systems Science & Control Engineering* **3**(1) 154–161.
- Bertsimas, Dimitris, Martin S Copenhaver. 2018. Characterization of the equivalence of robustification and regularization in linear and matrix regression. *European Journal of Operational Research* **270**(3) 931–942.
- Bloch, Francis, Nicolas Quérou. 2013. Pricing in social networks. *Games and economic behavior* **80** 243–261.
- Borgs, Christian, Michael Brautbar, Jennifer Chayes, Brendan Lucier. 2014. Maximizing social influence in nearly optimal time. *Proceedings of the twenty-fifth annual ACM-SIAM symposium on Discrete algorithms*. SIAM, 946–957.
- Bramoullé, Yann, Habiba Djebbari, Bernard Fortin. 2009. Identification of peer effects through social networks. *Journal of econometrics* **150**(1) 41–55.
- Buchbinder, Niv, Moran Feldman, Joseph Seffi Naor, Roy Schwartz. 2014. Submodular maximization with cardinality constraints. *Proceedings of the twenty-fifth annual ACM-SIAM symposium on Discrete algorithms*. Society for Industrial and Applied Mathematics, 1433–1452.
- Cameron, A. Colin, Pravin K Trivedi. 2005. *Microeconometrics: methods and applications*. Cambridge university press.
- Candogan, Ozan, Kostas Bimpikis, Asuman Ozdaglar. 2012. Optimal pricing in networks with externalities. *Operations Research* **60**(4) 883–905.
- Chen, Wei, Chi Wang, Yajun Wang. 2010. Scalable influence maximization for prevalent viral marketing in large-scale social networks. *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1029–1038.
- Chen, Wei, Yajun Wang, Siyu Yang. 2009. Efficient influence maximization in social networks. *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 199–208.
- Cohen, Maxime, Pavithra Harsha. 2013. Designing price incentives in a network with social interactions. Available at SSRN: <https://ssrn.com/abstract=2376668>.
- Cohen-Hillel, Tamar, Kiran Panchamgam, Georgia Perakis. 2019. High-low promotion policies for peak-end demand models. Available at SSRN 3360680 .
- Data USA. 2019. <https://datausa.io/profile/geo/mentor-oh/?compare=youngstown-oh>. [Online; accessed 28-April-2019].
- Domingos, Pedro, Matt Richardson. 2001. Mining the network value of customers. *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 57–66.
- Drakopoulos, Kimon, Asuman Ozdaglar, John N Tsitsiklis. 2014. An efficient curing policy for epidemics on graphs. *IEEE Transactions on Network Science and Engineering* **1**(2) 67–75.
- Drakopoulos, Kimon, Asuman Ozdaglar, John N Tsitsiklis. 2016. When is a network epidemic hard to eliminate? *Mathematics of Operations Research* **42**(1) 1–14.
- Eckles, Dean, Brian Karrer, Johan Ugander. 2017. Design and analysis of experiments in networks: Reducing bias from interference. *Journal of Causal Inference* **5**(1).
- Garetto, Michele, Weibo Gong, Don Towsley. 2003. Modeling malware spreading dynamics. *IEEE INFOCOM 2003. Twenty-second Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE Cat. No. 03CH37428)*, vol. 3. IEEE, 1869–1879.

- Gomez Rodriguez, Manuel, David Balduzzi, Bernhard Schölkopf. 2011. Uncovering the temporal dynamics of diffusion networks. ArXiv preprint arXiv:1105.0697.
- Gomez Rodriguez, Manuel, Jure Leskovec, Andreas Krause. 2010. Inferring networks of diffusion and influence. *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1019–1028.
- Goyal, Amit, Francesco Bonchi, Laks V.S. Lakshmanan. 2010. Learning influence probabilities in social networks. *Proceedings of the third ACM international conference on Web search and data mining*. 241–250, ACM.
- Hartline, Jason, Vahab Mirrokni, Mukund Sundararajan. 2008. Optimal marketing strategies over social networks. *Proceedings of the 17th international conference on World Wide Web*. ACM, 189–198.
- Imbens, Guido W. 2014. Instrumental variables: An econometrician’s perspective. *Statistical Science* **29**(3) 323–358.
- Karp, Richard M. 1975. On the computational complexity of combinatorial problems. *Networks* **5**(1) 45–68.
- Kempe, David, Jon Kleinberg, Éva Tardos. 2003. Maximizing the spread of influence through a social network. *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 137–146.
- Kempe, David, Jon Kleinberg, Éva Tardos. 2015. Maximizing the spread of influence through a social network. *Theory of Computing* **11**(4) 105–147.
- Krause, Andreas, Daniel Golovin. 2014. *Tractability: Practical Approaches to Hard Problems*, chap. Chapter 3: Submodular Function Maximization. Cambridge University Press.
- Kumar, V, Rohan Mirchandani. 2012. Increasing the roi of social media marketing. *MIT sloan management review* **54**(1) 55.
- Leskovec, Jure, Andreas Krause, Carlos Guestrin, Christos Faloutsos, Jeanne VanBriesen, Natalie Glance. 2007. Cost-effective outbreak detection in networks. *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 420–429.
- Liu, Na, Shuyun Ren, Tsan-Ming Choi, Chi-Leung Hui, Sau-Fun Ng. 2013. Sales forecasting for fashion retailing service industry: a review. *Mathematical Problems in Engineering* **2013**.
- Lu, Wei, Laks V. S. Lakshmanan. 2012. Profit maximization over social networks. *2012 IEEE 12th International Conference on Data Mining* 479–488.
- Martec International, Inc. 2011. The merchandise hierarchy. URL <http://www.martec-international.com/demos/ur/pri06/03pri06.htm>.
- Mohr, Iris. 2013. The impact of social media on the fashion industry. *Journal of Applied Business and Economics* **15**(2) 17–22.
- Myers, Seth, Jure Leskovec. 2010. On the convexity of latent social network inference. *Advances in neural information processing systems*. 1741–1749.
- Nemhauser, George L, Laurence A Wolsey, Marshall L Fisher. 1978. An analysis of approximations for maximizing submodular set functions—i. *Mathematical programming* **14**(1) 265–294.
- Nenni, Maria Elena, Luca Giustiniano, Luca Pirolo. 2013. Demand forecasting in the fashion industry: a review. *International Journal of Engineering Business Management* **5** 37.

- Nowzari, Cameron, Victor M Preciado, George J Pappas. 2016. Analysis and control of epidemics: A survey of spreading processes on complex networks. *IEEE Control Systems Magazine* **36**(1) 26–46.
- Pradiptarini, Charity. 2011. Social media marketing: Measuring its effectiveness and identifying the target market. *UW-L Journal of Undergraduate Research XIV* 1–11.
- Rathore, Ashish K, P Vigneswara Ilavarasan, Yogesh K Dwivedi. 2016. Social media content and product co-creation: an emerging paradigm. *Journal of Enterprise Information Management* **29**(1) 7–18.
- Rege, Mari. 2008. Why do people care about social status? *Journal of Economic Behavior & Organization* **66**(2) 233–242.
- Sahin, Ismail. 2006. Detailed review of rogers' diffusion of innovations theory and educational technology-related studies based on rogers' theory. *Turkish Online Journal of Educational Technology-TOJET* **5**(2) 14–23.

Appendix A: Proofs of Section 2

A.1. Proof of Lemma 1

LEMMA 1. Assume that $y_{c',i,t-M-1}$ is a valid instrument. Let $\lambda = 0$ and let \bar{P} be the solution to the following system of linear equalities,

$$\sum_{c'=1}^n \sum_{i=1}^m \sum_{t=M+2}^{T_i} \sum_{t'=t-M}^{t-1} y_{c',i,t-M-1} y_{c'it'} \bar{p}_{c'c} = \sum_{i=1}^m \sum_{t=M+2}^{T_i} y_{c',i,t-M-1} (y_{cit} - \hat{q}_{cit}) \text{ for all } c' \text{ and } c.$$

If $0 \leq \bar{p}_{c'c} \leq 1$ for all c' and c , then the algorithm estimates $\hat{p}_{c'c} = \bar{p}_{c'c}$.

Proof. In this proof, we will use the model written in the following matrix notation. Let

$$\begin{aligned} Y &= (y_{cit})_{i=1,\dots,m,t=M+2,\dots,T_i}^{c=1,\dots,n}, & Q &= (q_{cit})_{i=1,\dots,m,t=M+2,\dots,T_i}^{c=1,\dots,n}, & X &= \left(\sum_{t'=t-M}^{t-1} y_{cit'} \right)_{i=1,\dots,m,t=M+2,\dots,T_i}^{c=1,\dots,n}, \\ Z &= (y_{c,i,t-M-1})_{i=1,\dots,m,t=M+2,\dots,T_i}^{c=1,\dots,n}, & P &= (p_{c'c})_{c'=1,\dots,n}^{c=1,\dots,n}, & W &= (w_{c'c})_{c'=1,\dots,n}^{c=1,\dots,n}. \end{aligned}$$

With this notation, step 2 of the Trend-Estimation algorithm runs as follows. The first stage estimate \widehat{W} of problem (7) in step 2a is found by solving

$$\min_{0 \leq W \leq M} \sum_{c=1}^n (X_c - ZW_c)^T (X_c - ZW_c), \quad (11)$$

then a first stage prediction is made $\bar{X} = Z\widehat{W}$, and afterwards the final stage estimate \widehat{P} of problem (8) in step 2b of the Trend-Estimation algorithm is found by solving

$$\min_{0 \leq P \leq 1} \sum_{c=1}^n \left((Y_c - \widehat{Q}_c - \bar{X}P_c)^T (Y_c - \widehat{Q}_c - \bar{X}P_c) + \lambda \|P_c\|_1 \right). \quad (12)$$

In the first stage, solving problem (11) results in the OLS estimator,

$$\widehat{W}_c = (Z^T Z)^{-1} Z^T X_c \text{ for all } c \Rightarrow \widehat{W} = (Z^T Z)^{-1} Z^T X.$$

This allows us to write the first stage prediction as $\bar{X} = Z\widehat{W} = Z(Z^T Z)^{-1} Z^T X$. In the final stage, given that $\lambda = 0$, solving the unconstrained version of problem (12) results in the OLS estimator,

$$\bar{P}_c = (\bar{X}^T \bar{X})^{-1} \bar{X}^T (Y_c - \widehat{Q}_c) \text{ for all } c \Rightarrow \bar{P} = (\bar{X}^T \bar{X})^{-1} \bar{X}^T (Y - \widehat{Q}).$$

Substituting $\bar{X} = Z(Z^T Z)^{-1} Z^T X$ yields

$$\begin{aligned} \bar{P} &= (X^T Z (Z^T Z)^{-1} Z^T Z (Z^T Z)^{-1} Z^T X)^{-1} X^T Z (Z^T Z)^{-1} Z^T (Y - \widehat{Q}) \\ &= (X^T Z (Z^T Z)^{-1} Z^T X)^{-1} X^T Z (Z^T Z)^{-1} Z^T (Y - \widehat{Q}) \\ &= (Z^T X)^{-1} (X^T Z (Z^T Z)^{-1})^{-1} X^T Z (Z^T Z)^{-1} Z^T (Y - \widehat{Q}) \\ &= (Z^T X)^{-1} Z^T (Y - \widehat{Q}). \end{aligned}$$

This implies that $Z^T X \bar{P} = Z^T (Y - \widehat{Q})$, which means that for each c' and c ,

$$\sum_{c'=1}^n \sum_{i=1}^m \sum_{t=M+2}^{T_i} \sum_{t'=t-M}^{t-1} y_{c',i,t-M-1} y_{c'it'} \bar{p}_{c'c} = \sum_{i=1}^m \sum_{t=M+2}^{T_i} y_{c',i,t-M-1} (y_{cit} - \hat{q}_{cit}).$$

If $0 \leq \bar{p}_{c'c} \leq 1$ for all c' and c , then the unconstrained estimator to problem (12) satisfies all the constraints, and is therefore also the constrained estimator. Hence, if $0 \leq \bar{p}_{c'c} \leq 1$ for all c' and c , then $\hat{p}_{c'c} = \bar{p}_{c'c}$ for all c' and c . \square

A.2. Proof of Theorem 1

Consider the unconstrained estimator $\bar{p}_{c'c}$ described in Lemma 1 that solves the following system of linear equations, for all c'' and c ,

$$\sum_{c'=1}^n \sum_{i=1}^m \sum_{t=M+2}^{T_i} \sum_{t'=t-M}^{t-1} y_{c'' , i, t-M-1} y_{c' it'} \bar{p}_{c'c} = \sum_{i=1}^m \sum_{t=M+2}^{T_i} y_{c'' , i, t-M-1} (y_{cit} - \hat{q}_{cit}).$$

Substituting $y_{cit} = q_{cit} + \sum_{c'=1}^n p_{c'c} \sum_{t'=t-M}^{t-1} y_{c' it'} + \epsilon_{cit}$ yields

$$\sum_{c'=1}^n \sum_{i=1}^m \sum_{t=M+2}^{T_i} \sum_{t'=t-M}^{t-1} y_{c'' , i, t-M-1} y_{c' it'} \bar{p}_{c'c} = \sum_{i=1}^m \sum_{t=M+2}^{T_i} y_{c'' , i, t-M-1} \left(q_{cit} + \sum_{c'=1}^n p_{c'c} \sum_{t'=t-M}^{t-1} y_{c' it'} + \epsilon_{cit} - \hat{q}_{cit} \right).$$

Consider the right-hand side of this equality,

$$\begin{aligned} & \sum_{i=1}^m \sum_{t=M+2}^{T_i} y_{c'' , i, t-M-1} \left(q_{cit} + \sum_{c'=1}^n p_{c'c} \sum_{t'=t-M}^{t-1} y_{c' it'} + \epsilon_{cit} - \hat{q}_{cit} \right) \\ &= \sum_{c'=1}^n \sum_{i=1}^m \sum_{t=M+2}^{T_i} \sum_{t'=t-M}^{t-1} y_{c'' , i, t-M-1} y_{c' it'} p_{c'c} + \sum_{i=1}^m \sum_{t=M+2}^{T_i} y_{c'' , i, t-M-1} \epsilon_{cit} + \sum_{i=1}^m \sum_{t=M+2}^{T_i} y_{c'' , i, t-M-1} (q_{cit} - \hat{q}_{cit}). \end{aligned}$$

Observe that, outside of $\bar{p}_{c'c}$ and $p_{c'c}$, the first term of this expression is identical to the left-hand side of the original equality. Thus, if we can show that the latter two terms of the above expression converge to 0 as $m \rightarrow \infty$, then we can prove that the unconstrained estimator $\bar{p}_{c'c}$ equals the ‘true’ parameter $p_{c'c}$. Consider the first of these two terms,

$$\sum_{i=1}^m \sum_{t=M+2}^{T_i} y_{c'' , i, t-M-1} \epsilon_{cit} = \sum_{i=1}^m (T_i - M - 1) \frac{1}{\sum_{i=1}^m (T_i - M - 1)} \sum_{i=1}^m \sum_{t=M+2}^{T_i} y_{c'' , i, t-M-1} \epsilon_{cit}.$$

By the law of large numbers and the instrumental variables condition that the instrumental variable and the error term are uncorrelated, $\mathbb{E}[y_{c'' , i, t-M-1} \epsilon_{cit}] = 0$ for any c, c'', i , and t , we establish that

$$\frac{1}{\sum_{i=1}^m (T_i - M - 1)} \sum_{i=1}^m \sum_{t=M+2}^{T_i} y_{c'' , i, t-M-1} \epsilon_{cit} \rightarrow \mathbb{E}[y_{c'' , i, t-M-1} \epsilon_{cit}] = 0 \text{ as } m \rightarrow \infty.$$

Consider the second of these two terms,

$$\sum_{i=1}^m \sum_{t=M+2}^{T_i} y_{c'' , i, t-M-1} (q_{cit} - \hat{q}_{cit}).$$

Under the assumption that $\hat{q}_{cit} \rightarrow q_{cit}$ as $m \rightarrow \infty$, we establish that

$$\sum_{i=1}^m \sum_{t=M+2}^{T_i} y_{c'' , i, t-M-1} (q_{cit} - \hat{q}_{cit}) \rightarrow \sum_{i=1}^m \sum_{t=M+2}^{T_i} y_{c'' , i, t-M-1} (q_{cit} - q_{cit}) = 0 \text{ as } m \rightarrow \infty.$$

Together, this means that as $m \rightarrow \infty$, the unconstrained estimator $\bar{p}_{c'c}$ solves the following system of linear equations,

$$\sum_{c'=1}^n \sum_{i=1}^m \sum_{t=M+2}^{T_i} \sum_{t'=t-M}^{t-1} y_{c'' , i, t-M-1} y_{c' it'} \bar{p}_{c'c} = \sum_{c'=1}^n \sum_{i=1}^m \sum_{t=M+2}^{T_i} \sum_{t'=t-M}^{t-1} y_{c'' , i, t-M-1} y_{c' it'} p_{c'c}.$$

Given that the instrument is valid, this means that $\bar{p}_{c'c} \rightarrow p_{c'c}$ as $m \rightarrow \infty$. Knowing that $0 \leq p_{c'c} \leq 1$, this implies that $0 \leq \bar{p}_{c'c} \leq 1$ as $m \rightarrow \infty$. Hence, applying the result of Lemma 1, we obtain that $\hat{p}_{c'c} \rightarrow p_{c'c}$ as $m \rightarrow \infty$.

Appendix B: Proofs of Section 3

B.1. Proof of Theorem 2

This proof share some similarities with the proof of Proposition 4.3 in Nemhauser et al. (1978).

Let us denote by S_i , the subset that was chosen by the greedy algorithm at step i . We denote by $S^* = \{x_1^*, x_2^*, \dots, x_L^*\}$, the optimal set of size L that maximizes the function f .

Due to monotonicity, for every i , $f(S^*) \leq f(S^* \cup S_i)$. Therefore,

$$\begin{aligned} f(S^*) \leq f(S^* \cup S_i) &= f(S^* \cup S_i) + \sum_{j=1}^L f(S_i \cup \{x_1^*, \dots, x_j^*\}) - f(S_i \cup \{x_1^*, \dots, x_j^*\}) \\ &= f(S_i) + \sum_{j=1}^L f(S_i \cup \{x_1^*, \dots, x_j^*\}) - f(S_i \cup \{x_1^*, \dots, x_{j-1}^*\}) \end{aligned}$$

Following C -Submodularity of the function f ,

$$f(S_i \cup \{x_1^*, \dots, x_j^*\}) - f(S_i \cup \{x_1^*, \dots, x_{j-1}^*\}) \leq C + f(S_i \cup \{x_j^*\}) - f(S_i)$$

and therefore,

$$f(S^*) \leq f(S_i) + \sum_{j=1}^L C + f(S_i \cup \{x_j^*\}) - f(S_i) = f(S_i) + LC + \sum_{j=1}^L f(S_i \cup \{x_j^*\}) - f(S_i)$$

At each step i , the greedy algorithm chooses the element that maximizes the potential increase, therefore,

$$f(S^*) \leq f(S_i) + LC + \sum_{j=1}^L f(S_i \cup \{x_j^*\}) - f(S_i) \leq f(S_i) + LC + L(f(S_{i+1}) - f(S_i))$$

This can lead us to the following

$$f(S^*) - f(S_{i+1}) \leq (1 - \frac{1}{L})(f(S^*) - f(S_i)) + C$$

Applying to the last step of the algorithm,

$$\begin{aligned} f(S^*) - f(S_L) &\leq (1 - \frac{1}{L})f(S^*) - f(S_{L-1}) + C \leq (1 - \frac{1}{L})^2((f(S^*) - f(S_{L-2})) + (1 + (1 - \frac{1}{L}))C) \\ &\leq \dots \leq (1 - \frac{1}{L})^L(f(S^*) - f(S_0)) + C \sum_{j=0}^{L-1} (1 - \frac{1}{L})^j = (1 - \frac{1}{L})^L f(S^*) + C \frac{L^2(1 - \frac{1}{L}) - L}{L - 1} \end{aligned}$$

Therefore,

$$f(S_L) \geq (1 - \frac{1}{L})f(S^*) - C \frac{L^2(1 - \frac{1}{L}) - L}{L - 1}$$

B.2. Proof of Theorem 4

The proof of Theorem 4 is outlined as follows. In the first part of the proof, we analyze and simplify the potential function $\rho(\phi, \Phi)$, that was introduced in Algorithm 2. Then, in the second part of the proof, we introduce the marginal potential function $\partial\rho(\phi, \tilde{\phi}, \Phi) = \rho(\phi, \Phi \cup \tilde{\phi}) - \rho(\phi, \Phi)$, which captures the marginal change in the potential function as a result of changes in the policy Φ . In the second part of the proof we consider a simple case in which the shipping cost is equal to 0 for all items and locations. We show that the marginal potential function in these special cases is always non-positive and therefore, the following shows that the revenue function is submodular,

$$\partial\rho \leq 0 \Leftrightarrow \rho(\phi, \Phi \cup \tilde{\phi}) - \rho(\phi, \Phi) \leq 0 \Leftrightarrow R(\Phi \cup \{\phi, \tilde{\phi}\}) - R(\Phi \cup \tilde{\phi}) - R(\Phi \cup \phi) + R(\Phi) \leq 0$$

Last, in the third part, we use the results of the second part in order to show submodularity for the general case.

Step.1 We begin by analyzing the potential function ρ .

In this proof we make use of the potential function ρ ,

$$\begin{aligned} \rho(\phi, \Phi) = R(\Phi \cup \phi) - R(\Phi) = & \quad (13) \\ & \sum_{i=1}^m \sum_{c=1}^n \sum_{t=1}^T r_{cit}^{\Phi \cup \{\phi\}} N_c b_{cit}^{\Phi \cup \{\phi\}} - \sum_{i=1}^m \sum_{l=1}^{NL} \left(\sum_{c:\ell(c)=l} \sum_{t=1}^T N_c b_{cit}^{\Phi \cup \{\phi\}} - I_{il} \right)^+ \cdot s_{il} - \\ & \sum_{i=1}^m \sum_{c=1}^n \sum_{t=1}^T r_{cit}^{\Phi} N_c b_{cit}^{\Phi} + \sum_{i=1}^m \sum_{l=1}^{NL} \left(\sum_{c:\ell(c)=l} \sum_{t=1}^T N_c b_{cit}^{\Phi} - I_{il} \right)^+ \cdot s_{il} \end{aligned}$$

The sum in (13), can be rewritten as a sum of 7 functions

$$\rho(\phi, \Phi) = R(\Phi \cup \phi) - R(\Phi) = \quad (14)$$

$$\underbrace{\sum_{i=1}^m \sum_{c=1}^n N_c \sum_{t=1}^{t_\phi-1} (r_{cit}^{\Phi \cup \{\phi\}} b_{cit}^{\Phi \cup \{\phi\}} - r_{cit}^{\Phi} b_{cit}^{\Phi})}_{f_{t < t_\phi}(\phi, \Phi)} + \underbrace{\sum_{i=1}^m \sum_{c \neq c_\phi} N_c (r_{cit}^{\Phi \cup \{\phi\}} b_{cit}^{\Phi \cup \{\phi\}} - r_{cit}^{\Phi} b_{cit}^{\Phi})}_{f_{c \neq c_\phi, t \leq t_\phi}(\phi, \Phi)} + \quad (15)$$

$$\underbrace{N_{c_\phi} (r_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\phi\}} b_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\phi\}} - r_{c_\phi i_\phi t_\phi}^{\Phi} b_{c_\phi i_\phi t_\phi}^{\Phi})}_{f_{c_\phi, i_\phi, t_\phi}(\phi, \Phi)} + \underbrace{\sum_{c=1}^n N_c \sum_{t=t_\phi+1}^T (r_{c i_\phi t}^{\Phi \cup \{\phi\}} b_{c i_\phi t}^{\Phi \cup \{\phi\}} - r_{c i_\phi t}^{\Phi} b_{c i_\phi t}^{\Phi})}_{f_{c i_\phi t > t_\phi}(\phi, \Phi)} + \quad (16)$$

$$\underbrace{\sum_{i \neq i_\phi} N_{c_\phi} (r_{c_\phi i t_\phi}^{\Phi \cup \{\phi\}} b_{c_\phi i t_\phi}^{\Phi \cup \{\phi\}} - r_{c_\phi i t_\phi}^{\Phi} b_{c_\phi i t_\phi}^{\Phi})}_{f_{c_\phi, i \neq i_\phi, t_\phi}(\phi, \Phi)} + \underbrace{\sum_{i \neq i_\phi} \sum_{c=1}^n N_c \sum_{t=t_\phi+1}^T (r_{cit}^{\Phi \cup \{\phi\}} b_{cit}^{\Phi \cup \{\phi\}} - r_{cit}^{\Phi} b_{cit}^{\Phi})}_{f_{c, i \neq i_\phi, t > t_\phi}(\phi, \Phi)} + \quad (17)$$

$$\underbrace{\sum_{i=1}^m \sum_{l=1}^{NL} - \left(\sum_{c:\ell(c)=l} \sum_{t=1}^T N_c b_{cit}^{\Phi \cup \{\phi\}} - I_{il} \right)^+ \cdot s_{il}}_{f_{shipping\ fees}(\phi, \Phi)} + \underbrace{\sum_{i=1}^m \sum_{l=1}^{NL} \left(\sum_{c:\ell(c)=l} \sum_{t=1}^T N_c b_{cit}^{\Phi} - I_{il} \right)^+ \cdot s_{il}}_{f_{shipping\ fees}(\phi, \Phi)} = \quad (18)$$

$$\underbrace{\left(r_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\phi\}} b_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\phi\}} - r_{c_\phi i_\phi t_\phi}^{\Phi} b_{c_\phi i_\phi t_\phi}^{\Phi} \right) N_{c_\phi}}_{f_{c_\phi, i_\phi, t_\phi}(\phi, \Phi)} + \underbrace{\sum_{c=1}^n \sum_{t=t_\phi+1}^T \left(b_{c i_\phi t}^{\Phi \cup \{\phi\}} - b_{c i_\phi t}^{\Phi} \right) r_{c i_\phi t}^{\Phi} N_c}_{f_{c i_\phi t > t_\phi}(\phi, \Phi)} + \quad (19)$$

$$\underbrace{\sum_{i \neq i_\phi} \left(b_{c_\phi i t_\phi}^{\Phi \cup \{\phi\}} - b_{c_\phi i t_\phi}^{\Phi} \right) r_{c_\phi i t_\phi}^{\Phi} N_{c_\phi}}_{f_{c_\phi, i \neq i_\phi, t_\phi}(\phi, \Phi)} + \underbrace{\sum_{i \neq i_\phi} \sum_{c=1}^n \sum_{t=t_\phi+1}^T \left(b_{cit}^{\Phi \cup \{\phi\}} - b_{cit}^{\Phi} \right) r_{cit}^{\Phi} N_c}_{f_{c, i \neq i_\phi, t > t_\phi}(\phi, \Phi)} - \quad (20)$$

$$\underbrace{\sum_{i=1}^m \sum_{l=1}^{NL} \left(\left(\sum_{c:\ell(c)=l} \sum_{t=1}^T N_c b_{cit}^{\Phi \cup \{\phi\}} - I_{il} \right)^+ - \left(\sum_{c:\ell(c)=l} \sum_{t=1}^T N_c b_{cit}^{\Phi} - I_{il} \right)^+ \right) \cdot s_{il}}_{f_{shipping\ fees}(\phi, \Phi)} \quad (21)$$

In what follows, we analyze the changes in each of the 5 functions when we make changes in the promotion plan Φ . In order to simplify the 5 functions, we use Lemma 2. Lemma 2 also make use of the following notations.

Given a trend graph p , we denote by $P_j(c, c')$, the set of all paths of length exactly j , that are starting at node c and ending at node c' . We use π to denote a path $\pi \in P_j(c, c')$. The length of path π is denoted by $|\pi|$, and the i^{th} node in path π is denoted by π_i .

The following lemma reveals, the changes in the purchase probability of a given customer (c) and for a given item (i) at a given time (t), when we make a small changes in the base probability of possibly another customer (c') at an earlier time ($t' < t$) for the same item (i). We then use this insight in order to simplify the 5 functions in (14).

LEMMA 2. *Without loss of generality, assume that the purchase probability b_{cit} is being decreased by ϵ . Let us denote by $b_{c'i't'}$, the purchase probability for customer c' and item i' at period t' , given that the purchase probability of customer c for item i at time t is $b_{cit}^\epsilon = b_{cit} - \epsilon$. Then for every $c', i' = i, t' > t$,*

$$b_{c'i't'} - b_{c'i't'}^\epsilon = (b_{cit} - b_{cit}^\epsilon) \sum_{\pi \in P_{t'-t}(c,c')} \prod_{j=1}^{|\pi|-1} p_{\pi_j, \pi_{j+1}} \quad (22)$$

Proof. In the simple case where $t' = t + 1$, the gap in (22) can be easily calculated.

$$b_{c'i't'} - b_{c'i't'}^\epsilon = (b_{cit} - b_{cit}^\epsilon) p_{c,c'} \quad (23)$$

On the other hand, if $t' > t + 1$, the gap in (22) can be calculated as follows.

$$b_{c'i't'} - b_{c'i't'}^\epsilon = \sum_{c''=1}^n (b_{c''i't'-1} - b_{c''i't'-1}^\epsilon) p_{c'',c'}$$

We observe that for any $\tau < t$, $b_{c''i\tau} - b_{c''i\tau}^\epsilon = 0 \quad \forall c''$. For $\tau = t$, $b_{c''i\tau} - b_{c''i\tau}^\epsilon = 0 \quad \forall c'' \neq c$. Using (23) as the base step, we can inductively find

$$b_{c'i't'} - b_{c'i't'}^\epsilon = (b_{cit} - b_{cit}^\epsilon) \sum_{\pi \in P_{t'-t}(c,c')} \prod_{j=1}^{|\pi|-1} p_{\pi_j, \pi_{j+1}}$$

Then, the potential function in (14) becomes

$$\rho(\phi, \Phi) = R(\Phi \cup \phi) - R(\Phi) =$$

$$\underbrace{\left(r_{c\phi i\phi t\phi}^{\Phi \cup \{\phi\}} b_{c\phi i\phi t\phi}^{\Phi \cup \{\phi\}} - r_{c\phi i\phi t\phi}^{\Phi} b_{c\phi i\phi t\phi}^{\Phi} \right) N_{c\phi}}_{f_{c,\phi i\phi t\phi}(\phi, \Phi)} + \underbrace{\sum_{c=1}^n \sum_{t=t_\phi+1}^T \left(\left(b_{c\phi i\phi t\phi}^{\Phi \cup \{\phi\}} - b_{c\phi i\phi t\phi}^{\Phi} \right) \sum_{\pi \in P_{t-t_\phi}(c\phi, c)} \prod_{j=1}^{|\pi|-1} p_{\pi_j, \pi_{j+1}} \right) r_{c\phi i\phi t}^{\Phi} N_c}_{f_{ci\phi t > t_\phi}(\phi, \Phi)} \quad (24)$$

$$\underbrace{\sum_{i \neq i_\phi} \left(b_{c\phi i\phi t\phi}^{\Phi \cup \{\phi\}} - b_{c\phi i\phi t\phi}^{\Phi} \right) r_{c\phi i\phi t\phi}^{\Phi} N_{c\phi}}_{f_{c,\phi i \neq i_\phi, t\phi}(\phi, \Phi)} + \underbrace{\sum_{i \neq i_\phi} \sum_{c=1}^n \sum_{t=t_\phi+1}^T \left(\left(b_{c\phi i\phi t\phi}^{\Phi \cup \{\phi\}} - b_{c\phi i\phi t\phi}^{\Phi} \right) \sum_{\pi \in P_{t-t_\phi}(c\phi, c)} \prod_{j=1}^{|\pi|-1} p_{\pi_j, \pi_{j+1}} \right) r_{c\phi i\phi t}^{\Phi} N_c}_{f_{c, i \neq i_\phi, t > t_\phi}(\phi, \Phi)} \quad (25)$$

$$\underbrace{- \sum_{i=1}^m \sum_{l=1}^{NL} \left(\left(\sum_{c:l(c)=l} \sum_{t=1}^T N_c b_{cit}^{\Phi \cup \{\phi\}} - I_{il} \right)^+ - \left(\sum_{c:l(c)=l} \sum_{t=1}^T N_c b_{cit}^{\Phi} - I_{il} \right)^+ \right)}_{f_{shipping \text{ fees}}(\phi, \Phi)} \cdot s_{il} \quad (26)$$

$$\underbrace{- \sum_{i=1}^m \sum_{l=1}^{NL} \left(\left(\sum_{c:l(c)=l} \sum_{t=1}^T N_c b_{cit}^{\Phi \cup \{\phi\}} - I_{il} \right)^+ - \left(\sum_{c:l(c)=l} \sum_{t=1}^T N_c b_{cit}^{\Phi} - I_{il} \right)^+ \right)}_{f_{shipping \text{ fees}}(\phi, \Phi)} \cdot s_{il} \quad (27)$$

Step.2 In the second part of this proof, we analyze how the potential function $\rho(\phi, \Phi)$ changes, with changes in the policy Φ . In this part we consider the special case in which there is no shipping cost.

Let us denote by $\partial\rho(\phi, \tilde{\phi}, \Phi)$, the marginal potential function,

$$\begin{aligned} \partial\rho(\phi, \tilde{\phi}, \Phi) &= \rho(\phi, \Phi \cup \{\tilde{\phi}\}) - \rho(\phi, \Phi) = \\ & f_{c_\phi i_\phi t_\phi}(\phi, \Phi \cup \{\tilde{\phi}\}) - f_{c_\phi i_\phi t_\phi}(\phi, \Phi) + f_{c i_\phi t > t_\phi}(\phi, \Phi \cup \{\tilde{\phi}\}) - f_{c i_\phi t > t_\phi}(\phi, \Phi) + \\ & f_{c_\phi, i \neq i_\phi, t_\phi}(\phi, \Phi \cup \{\tilde{\phi}\}) - f_{c_\phi, i \neq i_\phi, t_\phi}(\phi, \Phi) + f_{c, i \neq i_\phi, t > t_\phi}(\phi, \Phi \cup \{\tilde{\phi}\}) - f_{c, i \neq i_\phi, t > t_\phi}(\phi, \Phi) \end{aligned} \quad (28)$$

In order to show that the revenue function R is submodular, we show that the marginal potential function in (28) is non-positive. We show the non-positivity for each of the 4 components of the potential function separately.

(a) $f_{c_\phi i_\phi t_\phi}(\phi, \Phi \cup \{\tilde{\phi}\}) - f_{c_\phi i_\phi t_\phi}(\phi, \Phi)$. This component of the marginal potential function in (28) is equal to,

$$\left(r_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\phi, \tilde{\phi}\}} b_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\phi, \tilde{\phi}\}} - r_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\tilde{\phi}\}} b_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\tilde{\phi}\}} - r_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\phi\}} b_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\phi\}} + r_{c_\phi i_\phi t_\phi}^{\Phi} b_{c_\phi i_\phi t_\phi}^{\Phi} \right) N_{c_\phi}.$$

We first note that $r_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\phi, \tilde{\phi}\}} = r_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\phi\}}$ and similarly, $r_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\tilde{\phi}\}} = r_{c_\phi i_\phi t_\phi}^{\Phi}$. In addition, $r_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\phi, \tilde{\phi}\}} \leq r_{c_\phi i_\phi t_\phi}^{\Phi}$.

Therefore,

$$\left(r_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\phi, \tilde{\phi}\}} b_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\phi, \tilde{\phi}\}} - r_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\tilde{\phi}\}} b_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\tilde{\phi}\}} - r_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\phi\}} b_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\phi\}} + r_{c_\phi i_\phi t_\phi}^{\Phi} b_{c_\phi i_\phi t_\phi}^{\Phi} \right) N_{c_\phi} = \quad (29)$$

$$\left(r_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\phi, \tilde{\phi}\}} (b_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\phi, \tilde{\phi}\}} - b_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\phi\}}) - r_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\tilde{\phi}\}} (b_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\tilde{\phi}\}} - b_{c_\phi i_\phi t_\phi}^{\Phi}) \right) N_{c_\phi} \leq \quad (30)$$

$$r_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\phi, \tilde{\phi}\}} \left(b_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\phi, \tilde{\phi}\}} - b_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\phi\}} - b_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\tilde{\phi}\}} + b_{c_\phi i_\phi t_\phi}^{\Phi} \right) N_{c_\phi} \quad (31)$$

Where the last inequality is due to the complementary assumption and the fact that $r_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\phi, \tilde{\phi}\}} = r_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\tilde{\phi}\}}$. Last, the trend component in both purchase probabilities in the potential function $\rho(\phi, \Phi \cup \{\tilde{\phi}\})$ are equal. Same is true for $\rho(\phi, \Phi)$. Therefore,

$$\begin{aligned} & r_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\phi, \tilde{\phi}\}} \left(b_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\phi, \tilde{\phi}\}} - b_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\phi\}} - b_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\tilde{\phi}\}} + b_{c_\phi i_\phi t_\phi}^{\Phi} \right) N_{c_\phi} = \\ & r_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\phi, \tilde{\phi}\}} \left(q_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\phi, \tilde{\phi}\}} - q_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\phi\}} - q_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\tilde{\phi}\}} + q_{c_\phi i_\phi t_\phi}^{\Phi} \right) N_{c_\phi} \leq 0 \end{aligned}$$

The last inequality is due to the submodularity of the base probability function.

(b) $f_{c i_\phi t > t_\phi}(\phi, \Phi \cup \{\tilde{\phi}\}) - f_{c i_\phi t > t_\phi}(\phi, \Phi)$. This component of the marginal potential function in (28) is equal to,

$$\begin{aligned} & \sum_{c=1}^n \sum_{t=t_\phi+1}^T \left(\left(b_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\phi, \tilde{\phi}\}} - b_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\tilde{\phi}\}} \right) \sum_{\pi \in P_{t-t_\phi}(c_\phi, c)} \prod_{j=1}^{|\pi|-1} p_{\pi_j, \pi_{j+1}} \right) r_{c i_\phi t}^{\Phi \cup \{\tilde{\phi}\}} N_c - \\ & \sum_{c=1}^n \sum_{t=t_\phi+1}^T \left(\left(b_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\phi\}} - b_{c_\phi i_\phi t_\phi}^{\Phi} \right) \sum_{\pi \in P_{t-t_\phi}(c_\phi, c)} \prod_{j=1}^{|\pi|-1} p_{\pi_j, \pi_{j+1}} \right) r_{c i_\phi t}^{\Phi} N_c \end{aligned}$$

For every $(c, i_\phi, t) \neq (c_\phi, i_\phi, t_\phi) \neq (c_\phi, i_\phi, t_\phi)$, $r_{c i_\phi t}^{\Phi} = r_{c i_\phi t}^{\Phi \cup \{\tilde{\phi}\}}$. However, in the case where $(c, i_\phi, t) = (c_\phi, i_\phi, t_\phi)$, $r_{c i_\phi t}^{\Phi} > r_{c i_\phi t}^{\Phi \cup \{\tilde{\phi}\}}$. Since promotion $(c_\phi i_\phi t_\phi)$ can only increase the probability that customer c_ϕ will buy item i_ϕ at time t_ϕ ,

$$\sum_{c=1}^n \sum_{t=t_\phi+1}^T \left(\left(b_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\phi, \tilde{\phi}\}} - b_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\tilde{\phi}\}} \right) \sum_{\pi \in P_{t-t_\phi}(c_\phi, c)} \prod_{j=1}^{|\pi|-1} p_{\pi_j, \pi_{j+1}} \right) r_{c i_\phi t}^{\Phi \cup \{\tilde{\phi}\}} N_c -$$

$$\sum_{c=1}^n \sum_{t=t_\phi+1}^T \left(\left(b_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\phi\}} - b_{c_\phi i_\phi t_\phi}^\Phi \right) \sum_{\pi \in P_{t-t_\phi}(c_\phi, c)} \prod_{j=1}^{|\pi|-1} p_{\pi_j, \pi_{j+1}} \right) r_{ci_\phi t}^\Phi N_c \leq$$

$$\sum_{c=1}^n \sum_{t=t_\phi+1}^T \left(\left(b_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\phi, \tilde{\phi}\}} - b_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\tilde{\phi}\}} - b_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\phi\}} + b_{c_\phi i_\phi t_\phi}^\Phi \right) \sum_{\pi \in P_{t-t_\phi}(c_\phi, c)} \prod_{j=1}^{|\pi|-1} p_{\pi_j, \pi_{j+1}} \right) r_{ci_\phi t}^\Phi N_c$$

Since promotion ϕ can not impact the trend component in $b_{c_\phi i_\phi t_\phi}$, regardless of the promotion plan, we can simplify the function above as follows.

$$\sum_{c=1}^n \sum_{t=t_\phi+1}^T \left(\left(q_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\phi, \tilde{\phi}\}} - q_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\tilde{\phi}\}} - q_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\phi\}} + q_{c_\phi i_\phi t_\phi}^\Phi \right) \sum_{\pi \in P_{t-t_\phi}(c_\phi, c)} \prod_{j=1}^{|\pi|-1} p_{\pi_j, \pi_{j+1}} \right) r_{ci_\phi t}^\Phi N_c \quad (32)$$

Due to the submodularity of the base probability function q , the function in (32) is always negative.

- (c) $f_{c_\phi, i \neq i_\phi, t_\phi}(\phi, \Phi \cup \{\tilde{\phi}\}) - f_{c_\phi, i \neq i_\phi, t_\phi}(\phi, \Phi)$. We provide an analysis for each item i separately. The case where $c_{\tilde{\phi}} = c_\phi$, $i_{\tilde{\phi}} = i$ and $t_{\tilde{\phi}} = t_\phi$, this case is identical to case **(Step.2a)**. Otherwise, we note that $r_{c_\phi i t_\phi}^{\Phi \cup \{\phi, \tilde{\phi}\}} = r_{c_\phi i t_\phi}^{\Phi \cup \{\phi\}} = r_{c_\phi i t_\phi}^{\Phi \cup \{\tilde{\phi}\}} = r_{c_\phi i t_\phi}^\Phi$. In addition, the trend component in both purchase probabilities in the potential function $\rho(\phi, \Phi \cup \{\tilde{\phi}\})$ are equal. Same is true for $\rho(\phi, \Phi)$. Therefore,

$$\left(r_{c_\phi i t_\phi}^{\Phi \cup \{\phi, \tilde{\phi}\}} b_{c_\phi i t_\phi}^{\Phi \cup \{\phi, \tilde{\phi}\}} - r_{c_\phi i t_\phi}^{\Phi \cup \{\tilde{\phi}\}} b_{c_\phi i t_\phi}^{\Phi \cup \{\tilde{\phi}\}} - r_{c_\phi i t_\phi}^{\Phi \cup \{\phi\}} b_{c_\phi i t_\phi}^{\Phi \cup \{\phi\}} + r_{c_\phi i t_\phi}^\Phi b_{c_\phi i t_\phi}^\Phi \right) N_{c_\phi} =$$

$$\left(r_{c_\phi i t_\phi}^{\Phi \cup \{\phi, \tilde{\phi}\}} \left(q_{c_\phi i t_\phi}^{\Phi \cup \{\phi, \tilde{\phi}\}} - q_{c_\phi i t_\phi}^{\Phi \cup \{\phi\}} - q_{c_\phi i t_\phi}^{\Phi \cup \{\tilde{\phi}\}} + q_{c_\phi i t_\phi}^\Phi \right) \right) N_{c_\phi} \leq 0$$

Where the last inequality is due to the submodularity of q .

- (d) $f_{c, i \neq i_\phi, t > t_\phi}(\phi, \Phi \cup \{\tilde{\phi}\}) - f_{c, i \neq i_\phi, t > t_\phi}(\phi, \Phi)$. We provide an analysis for each item i separately. Similarly to the argument in **(Step.2b)**, this component of the marginal potential function in (28) is bounded by

$$\sum_{c=1}^n \sum_{t=t_\phi+1}^T \left(\left(b_{c_\phi i t_\phi}^{\Phi \cup \{\phi, \tilde{\phi}\}} - b_{c_\phi i t_\phi}^{\Phi \cup \{\tilde{\phi}\}} - b_{c_\phi i t_\phi}^{\Phi \cup \{\phi\}} + b_{c_\phi i t_\phi}^\Phi \right) \sum_{\pi \in P_{t-t_\phi}(c_\phi, c)} \prod_{j=1}^{|\pi|-1} p_{\pi_j, \pi_{j+1}} \right) r_{cit}^\Phi N_c$$

Since promotion ϕ can not impact the trend component in $b_{c_\phi i t_\phi}$ at time t_ϕ , regardless of the promotion plan, we can simplify the function above as follows.

$$\sum_{c=1}^n \sum_{t=t_\phi+1}^T \left(\left(q_{c_\phi i t_\phi}^{\Phi \cup \{\phi, \tilde{\phi}\}} - q_{c_\phi i t_\phi}^{\Phi \cup \{\tilde{\phi}\}} - q_{c_\phi i t_\phi}^{\Phi \cup \{\phi\}} + q_{c_\phi i t_\phi}^\Phi \right) \sum_{\pi \in P_{t-t_\phi}(c_\phi, c)} \prod_{j=1}^{|\pi|-1} p_{\pi_j, \pi_{j+1}} \right) r_{cit}^\Phi N_c \quad (33)$$

Due to the submodularity of the base probability function q , the function in (33) is always non-positive.

Step.3 Next we show that these results can be extended to the general case of positive shipping fees.

In the general case, we consider the marginal potential function

$$\partial \rho(\phi, \tilde{\phi}, \Phi) = \rho(\phi, \Phi \cup \{\tilde{\phi}\}) - \rho(\phi, \Phi) = \quad (34)$$

$$f_{c_\phi i_\phi t_\phi}(\phi, \Phi \cup \{\tilde{\phi}\}) - f_{c_\phi i_\phi t_\phi}(\phi, \Phi) + f_{c i_\phi t > t_\phi}(\phi, \Phi \cup \{\tilde{\phi}\}) - f_{c i_\phi t > t_\phi}(\phi, \Phi) +$$

$$f_{c_\phi, i \neq i_\phi, t_\phi}(\phi, \Phi \cup \{\tilde{\phi}\}) - f_{c_\phi, i \neq i_\phi, t_\phi}(\phi, \Phi) + f_{c, i \neq i_\phi, t > t_\phi}(\phi, \Phi \cup \{\tilde{\phi}\}) - f_{c, i \neq i_\phi, t > t_\phi}(\phi, \Phi) +$$

$$f_{\text{shipping fees}}(\phi, \Phi \cup \{\tilde{\phi}\}) - f_{\text{shipping fees}}(\phi, \Phi)$$

We now show that for every item i and location l , the component in (34) that corresponds with item i and location l is always non-positive. For the sake of brevity, we use AS_{il}^{Φ} to denote difference between the number of units of item i that were sold in location l according to promotion plan Φ , and the initial inventory for item i at location l . Formally,

$$AS_{il}^{\Phi \cup \{\phi, \tilde{\phi}\}} = \sum_{c:\ell(c)=l} \sum_{t=1}^T N_c b_{cit}^{\Phi \cup \{\phi, \tilde{\phi}\}} - I_{il} \quad , \quad AS_{il}^{\Phi \cup \{\tilde{\phi}\}} = \sum_{c:\ell(c)=l} \sum_{t=1}^T N_c b_{cit}^{\Phi \cup \{\tilde{\phi}\}} - I_{il}$$

$$AS_{il}^{\Phi \cup \{\phi\}} = \sum_{c:\ell(c)=l} \sum_{t=1}^T N_c b_{cit}^{\Phi \cup \{\phi\}} - I_{il} \quad , \quad AS_{il}^{\Phi} = \sum_{c:\ell(c)=l} \sum_{t=1}^T N_c b_{cit}^{\Phi} - I_{il}$$

For each combination of item i and location l , we consider the following distinct cases.

- (a) The first case is where $AS_{il}^{\Phi \cup \{\tilde{\phi}\}} > 0$. Here once again, we consider two cases.
- i. The first case is where $AS_{il}^{\Phi \cup \{\phi\}} > 0$. In this case,

$$f_{\text{shipping fees}}(\phi, \Phi \cup \{\tilde{\phi}\}) - f_{\text{shipping fees}}(\phi, \Phi) \leq \quad (35)$$

$$-AS_{il}^{\Phi \cup \{\phi, \tilde{\phi}\}} + AS_{il}^{\Phi \cup \{\tilde{\phi}\}} + AS_{il}^{\Phi \cup \{\phi\}} - AS_{il}^{\Phi}. \quad (36)$$

The expression in 35 is not necessarily non-positive. Nevertheless, since $r_{cit}^{\Phi} > s_{il}$, for item i at location l , the revenue function can be re-written as follows,

$$\sum_{c:\ell(c)=l} \sum_{t=1}^T N_c b_{cit}^{\Phi} (r_{cit}^{\Phi} - s_{il}) + I_{il} s_{il}$$

This allows us to use the results from part (**Step.2**) of this proof, and claim that the revenue function is non-positive for these cases.

- ii. In the second case, $AS_{il}^{\Phi \cup \{\phi\}} \leq 0$,

$$f_{\text{shipping fees}}(\phi, \Phi \cup \{\tilde{\phi}\}) - f_{\text{shipping fees}}(\phi, \Phi) \leq -AS_{il}^{\Phi \cup \{\phi, \tilde{\phi}\}} + AS_{il}^{\Phi \cup \{\tilde{\phi}\}} = \quad (37)$$

$$\left(- \sum_{c:\ell(c)=l} \sum_{t=1}^T N_c b_{cit}^{\Phi \cup \{\phi, \tilde{\phi}\}} + I_{il} + \sum_{c:\ell(c)=l} \sum_{t=1}^T N_c b_{cit}^{\Phi \cup \{\tilde{\phi}\}} - I_{il} \right) \cdot s_{il} \quad (38)$$

Due to the assumption of complementary items, promotions can never decrease the number of sold unites. Therefore, the function in (38) is always non-positive.

- (b) The second case is the case where $AS_{il}^{\Phi \cup \{\tilde{\phi}\}} < 0$. Here again we can consider two cases,

- i. The first case is where $AS_{il}^{\Phi \cup \{\phi\}} > 0$. This case is symmetrical to the case in **Step.3**(a)ii.
- ii. In the second case, $AS_{il}^{\Phi \cup \{\phi\}} \leq 0$. In this case,

$$f_{\text{shipping fees}}(\phi, \Phi \cup \{\tilde{\phi}\}) - f_{\text{shipping fees}}(\phi, \Phi) \leq 0 \quad (39)$$

Under the assumption of complementary, all other cases are not possible. Therefore, in all cases, the marginal potential function is non-positive.

Appendix C: Supplementary Material

C.1. Finite-Sample Guarantee on Estimation Error

Besides the fact that $y_{c'',i,t-M-1}$ needs to be a valid instrument, we need to make additional assumptions to prove a probabilistic finite-sample guarantee on the error of the estimator. First, the base purchase probability is independent of the item and only positive in the first period: $q_{cit} = q_c$ for $t = 1$ and $q_{cit} = 0$ for $t > 1$. Through this assumption, we are guaranteed that all purchases after the initial period are made because of the customer trend effect. This allows us in finding a closed-form solution for the estimated of the customer-to-customer trend probabilities.

Second, we assume that the customer trend network is known to be a forest. If a directed graph is a forest, this means that each vertex in the graph has at most one incoming edge. As shown in Figure 2b, the customer trend can be represented by such a directed graph. In the case that this network is a forest, every customer is affected by at most one other customer, i.e., for each customer c there is one customer c' such that $p_{c'c} \geq 0$ while $p_{c''c} = 0$ for all other customers c'' . Commonly, the customer c' that affects the purchase decision of customer c is called the parent of customer c , which we denote by $\pi(c) = c'$. As with the assumption on the base purchase model, this assumption on the customer trend model is needed for establishing the closed-form expression of the customer-to-customer trend estimates. Together these assumptions are reasonable for a number of scenarios. In the fashion setting, this would be the case if celebrities have many dedicated fashion followers, but are themselves not affected by their followers or other celebrities.

Furthermore, in line with the instrument being valid, we also assume that each customer purchased at least one item during its first time period. Naturally, customers are only present in the transaction dataset when they make a purchase, and likely they have made a purchase in the first time period as well.

Finally, the memory is assumed to last for one time period, meaning that $M = 1$. For example, if the time periods last a week, this means that a purchase from one week can only affect other customers' purchase decisions in the next week. In the case of celebrities this could make sense, because there is a small time window in which celebrities flaunt the fashion clothing.

Given these assumptions, we can prove Theorem 5 that establishes a finite-sample guarantee on the estimation error between the estimated and 'true' customer-to-customer trend probabilities with high probability. This result shows that the probability of an error is inversely proportional to the number of distinct items in the dataset. Hence, as expected, a larger error becomes less likely when the dataset grows in size.

THEOREM 5. *Assume that $y_{c'',i,t-M-1}$ is a valid instrument, that the base purchase probability q_{cit} is independent of the item and only positive in the first period, that the customer trend network is known to be a forest, that at least one purchase is made by each customer in the first period, and that the memory lasts one period. Let $\lambda = 0$ and consider the 'true' model (5), then the difference between the estimated and 'true' customer-to-customer trend probabilities is bounded with high probability by*

$$\mathbb{P}(|\hat{p}_{\pi(c),c} - p_{\pi(c),c}| > \epsilon) \leq \frac{1}{4\epsilon^2 m q_{\pi(c)}}.$$

Proof. Given that at least one purchase is made and that the initial base purchase probability is independent of the item, we note that

$$\sum_{i=1}^m y_{\pi(c),i,1} \sim 1 + \text{Bin}(m-1, q_{\pi(c)}) \quad \text{and} \quad \sum_{i=1}^m \sum_{t=3}^{T_i} y_{cit} \mid \sum_{i=1}^m \sum_{t=3}^{T_i} y_{\pi(c),i,t-1} = x \sim \text{Bin}(x, p_{\pi(c),c}).$$

To bound this tail probability, let us substitute in the estimator from Lemma 4, after which we condition on the value of the denominator in the estimator,

$$\begin{aligned} \mathbb{P}(|\widehat{p}_{\pi(c),c} - p_{\pi(c),c}| > \epsilon) &= \mathbb{P}\left(\left|\frac{\sum_{i=1}^m \sum_{t=3}^{T_i} y_{cit}}{\sum_{i=1}^m \sum_{t=3}^{T_i} y_{\pi(c),i,t-1}} - p_{\pi(c),c}\right| > \epsilon\right) \\ &= \sum_{x=1}^{\sum_{i=1}^m (T_i-2)} \mathbb{P}\left(\left|\frac{\sum_{i=1}^m \sum_{t=3}^{T_i} y_{cit}}{x} - p_{\pi(c),c}\right| > \epsilon \mid \sum_{i=1}^m \sum_{t=3}^{T_i} y_{\pi(c),i,t-1} = x\right) \mathbb{P}\left(\sum_{i=1}^m \sum_{t=3}^{T_i} y_{\pi(c),i,t-1} = x\right). \end{aligned}$$

Next, we can apply Chebyshev's inequality to the first probability term and plug in the variance of a binomial random variable,

$$\begin{aligned} \mathbb{P}(|\widehat{p}_{\pi(c),c} - p_{\pi(c),c}| > \epsilon) &\leq \sum_{x=1}^{\sum_{i=1}^m (T_i-2)} \frac{\text{Var}(\sum_{i=1}^m \sum_{t=3}^{T_i} y_{cit} \mid \sum_{i=1}^m \sum_{t=3}^{T_i} y_{\pi(c),i,t-1} = x)}{x^2 \epsilon^2} \mathbb{P}\left(\sum_{i=1}^m \sum_{t=3}^{T_i} y_{\pi(c),i,t-1} = x\right) \\ &= \sum_{x=1}^{\sum_{i=1}^m (T_i-2)} \frac{x p_{\pi(c),c} (1 - p_{\pi(c),c})}{x^2 \epsilon^2} \mathbb{P}\left(\sum_{i=1}^m \sum_{t=3}^{T_i} y_{\pi(c),i,t-1} = x\right) \\ &= \frac{p_{\pi(c),c} (1 - p_{\pi(c),c})}{\epsilon^2} \sum_{x=1}^{\sum_{i=1}^m (T_i-2)} \frac{1}{x} \mathbb{P}\left(\sum_{i=1}^m \sum_{t=3}^{T_i} y_{\pi(c),i,t-1} = x\right) \\ &= \frac{p_{\pi(c),c} (1 - p_{\pi(c),c})}{\epsilon^2} \mathbb{E}\left[\frac{1}{\sum_{i=1}^m \sum_{t=3}^{T_i} y_{\pi(c),i,t-1} + 1}\right] \\ &\leq \frac{p_{\pi(c),c} (1 - p_{\pi(c),c})}{\epsilon^2} \mathbb{E}\left[\frac{1}{\sum_{i=1}^m y_{\pi(c),i,1} + 1}\right]. \end{aligned}$$

We note that in the last inequality, the expectation is upper bounded by shrinking the denominator from $\sum_{i=3}^{T_i} y_{\pi(c),i,t-1}$ to $y_{\pi(c),i,1}$. Next, we use the result of Lemma 3 to bound the expectation.

LEMMA 3. *For $X \sim \text{Bin}(n, p)$ we have,*

$$\mathbb{E}\left[\frac{1}{X+1}\right] = \frac{1 - (1-p)^{n+1}}{(n+1)p}$$

Proof. Given that $X \sim \text{Bin}(n, p)$, the expectation is given by

$$\begin{aligned} \mathbb{E}\left[\frac{1}{X+1}\right] &= \sum_{x=0}^n \frac{1}{x+1} \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} = \sum_{x=0}^n \frac{n!}{(x+1)!(n-x)!} p^x (1-p)^{n-x} \\ &= \frac{1}{(n+1)p} \sum_{x=0}^n \frac{(n+1)!}{(x+1)!(n-x)!} p^{x+1} (1-p)^{n-x} = \frac{1}{(n+1)p} \sum_{x=1}^{n+1} \frac{(n+1)!}{x!(n-x+1)!} p^x (1-p)^{n-x+1} \\ &= \frac{1 - (1-p)^{n+1}}{(n+1)p}. \end{aligned}$$

□

Using Lemma 3, we obtain the following

$$\mathbb{P}(|\widehat{p}_{\pi(c),c} - p_{\pi(c),c}| > \epsilon) \leq \frac{p_{\pi(c),c} (1 - p_{\pi(c),c})}{\epsilon^2} \mathbb{E}\left[\frac{1}{\sum_{i=1}^m y_{\pi(c),i,1} + 1}\right] = \frac{p_{\pi(c),c} (1 - p_{\pi(c),c})}{\epsilon^2} \frac{1 - (1 - q_{\pi(c)})^m}{m q_{\pi(c)}}.$$

Finally, we can use the fact that $p_{\pi(c),c} (1 - p_{\pi(c),c}) \leq \frac{1}{4}$ and $1 - (1 - q_{\pi(c)})^m \leq 1$ for any $m > 0$ to obtain

$$\mathbb{P}(|\widehat{p}_{\pi(c),c} - p_{\pi(c),c}| > \epsilon) \leq \frac{1}{4\epsilon^2 m q_{\pi(c)}}.$$

□

In this proof, we use the following Lemma 4 that establishes a closed-form expression for our estimator given the assumptions.

LEMMA 4. *Assume that $y_{c'',i,t-M-1}$ is a valid instrument, that the base purchase probability q_{cit} is only positive in the first period, and that the customer trend network is known to be a forest. Let $\lambda = 0$, then the Trend-Estimation algorithm estimates the customer-to-customer trend probabilities as*

$$\widehat{p}_{c'c} = \begin{cases} \frac{\sum_{i=1}^m \sum_{t=M+2}^{T_i} y_{\pi^2(c),i,t-M-1} y_{cit}}{\sum_{i=1}^m \sum_{t=M+2}^{T_i} \sum_{t'=t-M}^{t-1} y_{\pi^2(c),i,t-M-1} y_{\pi(c),i,t'}} & \text{if } c' = \pi(c) \\ 0 & \text{if } c' \neq \pi(c) \end{cases}.$$

Specifically, if the memory lasts for only one period,

$$\widehat{p}_{c'c} = \begin{cases} \frac{\sum_{i=1}^m \sum_{t=3}^{T_i} y_{cit}}{\sum_{i=1}^m \sum_{t=3}^{T_i} y_{\pi(c),i,t-1}} & \text{if } c' = \pi(c) \\ 0 & \text{if } c' \neq \pi(c) \end{cases}.$$

Proof. Consider the unconstrained estimator $\bar{p}_{c'c}$ described in Lemma 1 that solves the following system of linear equations, for all c'' and c ,

$$\sum_{c'=1}^n \sum_{i=1}^m \sum_{t=M+2}^{T_i} \sum_{t'=t-M}^{t-1} y_{c'',i,t-M-1} y_{c'it'} \bar{p}_{c'c} = \sum_{i=1}^m \sum_{t=M+2}^{T_i} y_{c'',i,t-M-1} (y_{cit} - \widehat{q}_{cit}).$$

Given that the base purchase probability is 0 and that the customer trend network forms a known directed forest, we can write for all c ,

$$\sum_{i=1}^m \sum_{t=M+2}^{T_i} \sum_{t'=t-M}^{t-1} y_{\pi^2(c),i,t-M-1} y_{\pi(c),i,t'} \bar{p}_{\pi(c),c} = \sum_{i=1}^m \sum_{t=M+2}^{T_i} y_{\pi^2(c),i,t-M-1} y_{cit}.$$

In turn, given that the instrument is valid, this implies that

$$\bar{p}_{\pi(c),c} = \frac{\sum_{i=1}^m \sum_{t=M+2}^{T_i} y_{\pi^2(c),i,t-M-1} y_{cit}}{\sum_{i=1}^m \sum_{t=M+2}^{T_i} \sum_{t'=t-M}^{t-1} y_{\pi^2(c),i,t-M-1} y_{\pi(c),i,t'}}.$$

We note that $0 \leq \bar{p}_{\pi(c),c} \leq 1$, because both the numerator and denominator are positive as well as the fact that the denominator is always at least as large as the numerator. Hence, applying the result of Lemma 1 yields that estimation algorithm 1 results in $\widehat{p}_{c'c} = \bar{p}_{c'c}$. \square

C.2. NP Hardness of the Dynamic Promotion Targeting Optimization Problem

In what follows, we describe a simple polynomial time reduction from the Set-Cover problem to the Dynamic Promotion Targeting Optimization Problem. We show that if we can solve the Dynamic Promotion Targeting Optimization Problem in polynomial time, then we can solve the Set-Cover problem in polynomial time as well. This is in contradiction to the fact that the Set-Cover problem is NP-hard problem, as it is one of Karp's 21 NP-complete problems (Karp (1975)). The Set-Cover problem is defined as follows.

DEFINITION 2 (SET-COVER). Given a set of elements $\Omega = \{1, 2, \dots, n\}$, referred to as the universe and a collection \mathcal{S} of m sets whose union equals the universe, and a constant \mathcal{K} , the decision version of the set cover problem is to identify a sub-collection of \mathcal{S} , of size at most \mathcal{K} , whose union equals the universe.

Next we define a polynomial time reduction from the Set-Cover problem to the Dynamic Promotion Targeting Optimization Problem. This reduction defines a set of steps, such that given an instance of the Set-Cover problem, denoted by \mathcal{I}_{SC} , allows us to construct an instance of the Dynamic Promotion Targeting Optimization Problem, denoted by \mathcal{I}_{SPT} . Then, in Lemma 5, we show that given an optimal solution to the Dynamic

Promotion Targeting Optimization Problem, we can construct an optimal solution to the Set-Cover problem, and therefore, the Dynamic Promotion Targeting Optimization Problem is just as hard as the Set-Cover problem.

Given an instance of the Set-Cover problem, $\mathcal{I}_{SC} = \{\Omega, \mathcal{S}, \mathcal{K}\}$, we construct an instance of the Dynamic Promotion Targeting Optimization Problem, \mathcal{I}_{SPT} , as follows. We consider a problem with a single item $m = i$, with time horizon $T = 2$, and $n = |\mathcal{S}| + |\Omega|$ customers. We construct two types of customers. The first type, also referred to as *Set-Customers*, consists of $|\mathcal{S}|$ customers, where each customer corresponds to a set in \mathcal{S} . The second type, also referred to as *Element-Customers*, consists of $|\Omega|$ customers, where each customer corresponds to an element in Ω .

The price ladder is $d_1 = \{d_1 = 0, d_1^0 = 1\}$. Table 4 illustrates the base probabilities for the different price levels, different periods, and different types of customers.

Table 4 Caption

	Set Costumers		Element Costumers	
	$t = 1$	$t = 2$	$t = 1$	$t = 2$
$r_{c1t} = d_1^0$	0	0	0	0
$r_{c1t} = d_1$	1	0	0	0

Each store location is associated with a single customer, and the initial inventory is equal to 1 for all the different locations. The shipping cost s_{1l} is equal to ∞ for all locations.

We construct the trend graph as follows. There exists a trend from each set customer c_s to all of the element costumers c_i such that $\omega_i \in \mathcal{S}_s$ (where ω_i is the element that is associated with costumer c_i , and \mathcal{S}_s is the set that is associated with costumer c_s). The probability $p_{c,c'}$ for each such trend is equal to 1. For each element costumer c_i , we demote by S_i , the set of *Set-Customers* c_s such that there exists a trend from c_s to c_i . according to the network that was described above.

Last, we set the limit on the number of promotions, L to be equal to \mathcal{K} .

LEMMA 5. *There exists an optimal solution Φ^* to \mathcal{I}_{SPT} , such that $R(\Phi^*) = n$, if and only if there exist a sub-collection of \mathcal{S} , of size at most \mathcal{K} , whose union equals the universe.*

Proof. We first make the following observations:

1. The high shipping cost implies that under no circumstance an optimal solution will encore a backorder. Therefore, each customer can buy at most 1 unit of item 1.
2. Promotion price has no affect on the purchase probability of customers of type *Element-Customers*. Therefore, an optimal solution will never offer a promotion to customers of type *Element-Customers*. Similarly, an optimal solution will not offer a promotion to customers of type *Set-Customers* for the first period.
3. Since there are no incoming trends into customers of type *Set-Customers*, and the base probability is equal to 0 when there is no promotion price, a customer of type *Set-Customers* purchase if and only if he was offered a promotion. This can happen only in the first period.

4. A customer c of type *Element-Customers* can purchase only in the second period, if and only if, one of the *Set-Customers* that influences customer c , purchased.

Building on this set of observations, we can rewrite the revenue function achieved by Φ^* as follows:

$$R(\Phi^*) = \sum_{i \in \Omega: \exists \phi \in \Phi^*, c_\phi \in S_i} 1 \cdot 1 = |\{i \in \Omega : \exists \phi \in \Phi^*, c_\phi \in S_i\}| \quad (40)$$

We observe that the expression above is equal to the number of *Element-Customers* c_i such that at least one of the customers in the set S_i was offered a promotion according to Φ^* .

The optimal solution to \mathcal{I}_{SPT} that maximizes the revenue function in (40) is one that maximizes the number of *Element-Customers* c_i for which $b_{c_i}^{\Phi^*} = 1$.

Let us assume that there exists a promotion plan Φ of size at most L such that $R(\Phi^*) = n$. In other words, for each *Element Customer* c_i there exists a *Set-Customers* $c_s \in S_i$ that was offered a promotion according to Φ . Then, the sub-collection of \mathcal{S} that corresponds to the *Set-Customers* who was offered promotions, is of size at most K , and covers the universe.

Similarly, given a sub-collection S of \mathcal{S} that covers the universe, we can construct a promotion policy Φ , such that each customer that correspond to a set in S is offered a promotion in the first period. Then, the revenue function $R(\Phi) = n$.

C.3. Proof of Theorem 3

The proof of Theorem 3 is outlined as follows. In the first part, we consider the case of,

$$f(Y \cup \{x\}) - f(Y) \leq C + f(X \cup \{x\}) - f(X),$$

where $|Y \setminus X| = 1$. Then, in the second part, we extend the result to the general case in which $|Y \setminus X| \leq L$.

Step.1 We begin by analyzing the simple case in which $|Y \setminus X| = 1$. The proof of part follows a similar lines to the proof of Theorem 4. We calculate an upper bound for each component of the marginal potential function in (28).

- (a) $f_{c_\phi i_\phi t_\phi}(\phi, \Phi \cup \{\tilde{\phi}\}) - f_{c_\phi i_\phi t_\phi}(\phi, \Phi)$. This component of the marginal potential function in (28) is equal to the formulation in (30).

$$\begin{aligned} & \left(r_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\phi, \tilde{\phi}\}} (b_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\phi, \tilde{\phi}\}} - b_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\phi\}}) - r_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\tilde{\phi}\}} (b_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\tilde{\phi}\}} - b_{c_\phi i_\phi t_\phi}^{\Phi}) \right) N_{c_\phi} = \\ & \left(r_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\phi, \tilde{\phi}\}} (b_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\phi, \tilde{\phi}\}} - b_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\phi\}} - b_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\tilde{\phi}\}} + b_{c_\phi i_\phi t_\phi}^{\Phi}) + (r_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\phi, \tilde{\phi}\}} - r_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\tilde{\phi}\}}) (b_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\tilde{\phi}\}} - b_{c_\phi i_\phi t_\phi}^{\Phi}) \right) N_{c_\phi} = \\ & \left(d_i (b_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\phi, \tilde{\phi}\}} - b_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\phi\}} - b_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\tilde{\phi}\}} + b_{c_\phi i_\phi t_\phi}^{\Phi}) + (d_i - d_i^0) (b_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\tilde{\phi}\}} - b_{c_\phi i_\phi t_\phi}^{\Phi}) \right) N_{c_\phi} \leq \\ & (d_i C_q + (d_i^0 - d_i) \delta) N_{c_\phi} \end{aligned}$$

- (b) $f_{c_i \phi t > t_\phi}(\phi, \Phi \cup \{\tilde{\phi}\}) - f_{c_i \phi t > t_\phi}(\phi, \Phi)$. This component of the marginal potential function in (28) is upper bounded by the formulation in 32,

$$\sum_{c=1}^n \sum_{t=t_\phi+1}^T \left(\left(q_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\phi, \tilde{\phi}\}} - q_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\tilde{\phi}\}} - q_{c_\phi i_\phi t_\phi}^{\Phi \cup \{\phi\}} + q_{c_\phi i_\phi t_\phi}^{\Phi} \right) \sum_{\pi \in P_{t-t_\phi}(c_\phi, c)} \prod_{j=1}^{|\pi|-1} p_{\pi_j, \pi_{j+1}} \right) r_{c_i \phi t}^{\Phi} N_c \leq \quad (41)$$

$$\sum_{c=1}^n \sum_{t=t_\phi+1}^T \left(\sum_{\pi \in P_{t-t_\phi}(c_\phi, c)} \prod_{j=1}^{|\pi|-1} p_{\pi_j, \pi_{j+1}} \right) C_q r_{c_i \phi t}^{\Phi} N_c \leq \quad (42)$$

$$\sum_{c=1}^n \sum_{t=t_\phi+1}^T \left(\sum_{\pi \in P_{t-t_\phi}(c_\phi, c)} \prod_{j=1}^{|\pi|-1} p_{\pi_j, \pi_{j+1}} \right) C_q d_i^0 N_c \leq (T-1) \pi^*(P) C_q d_i^0 \sum_{c=1}^n N_c \quad (43)$$

Where the second to last inequality is due to the C_q -submodularity of the base probability function q .

- (c) $f_{c_\phi, i \neq i_\phi, t_\phi}(\phi, \Phi \cup \{\tilde{\phi}\}) - f_{c_\phi, i \neq i_\phi, t_\phi}(\phi, \Phi)$. We provide an analysis for each item i separately. The case where $c_{\tilde{\phi}} = c_\phi$, $i_{\tilde{\phi}} = i$ and $t_{\tilde{\phi}} = t_\phi$, this case is identical to case **(Step.1a)**.

$$\left(r_{c_\phi i t_\phi}^{\Phi \cup \{\phi, \tilde{\phi}\}} b_{c_\phi i t_\phi}^{\Phi \cup \{\phi, \tilde{\phi}\}} - r_{c_\phi i t_\phi}^{\Phi \cup \{\tilde{\phi}\}} b_{c_\phi i t_\phi}^{\Phi \cup \{\tilde{\phi}\}} - r_{c_\phi i t_\phi}^{\Phi \cup \{\phi\}} b_{c_\phi i t_\phi}^{\Phi \cup \{\phi\}} + r_{c_\phi i t_\phi}^{\Phi} b_{c_\phi i t_\phi}^{\Phi} \right) N_{c_\phi} \leq (d_i C_q + (d_i^0 - d_i) \delta) N_{c_\phi}$$

Otherwise, we note that $r_{c_\phi i t_\phi}^{\Phi \cup \{\phi, \tilde{\phi}\}} = r_{c_\phi i t_\phi}^{\Phi \cup \{\phi\}} = r_{c_\phi i t_\phi}^{\Phi \cup \{\tilde{\phi}\}} = r_{c_\phi i t_\phi}^{\Phi}$. In addition, the trend component in both purchase probabilities in the potential function $\rho(\phi, \Phi \cup \{\tilde{\phi}\})$ are equal. Same is true for $\rho(\phi, \Phi)$. Therefore,

$$\begin{aligned} & \left(r_{c_\phi i t_\phi}^{\Phi \cup \{\phi, \tilde{\phi}\}} b_{c_\phi i t_\phi}^{\Phi \cup \{\phi, \tilde{\phi}\}} - r_{c_\phi i t_\phi}^{\Phi \cup \{\tilde{\phi}\}} b_{c_\phi i t_\phi}^{\Phi \cup \{\tilde{\phi}\}} - r_{c_\phi i t_\phi}^{\Phi \cup \{\phi\}} b_{c_\phi i t_\phi}^{\Phi \cup \{\phi\}} + r_{c_\phi i t_\phi}^{\Phi} b_{c_\phi i t_\phi}^{\Phi} \right) N_{c_\phi} = \\ & \left(r_{c_\phi i t_\phi}^{\Phi \cup \{\phi, \tilde{\phi}\}} (q_{c_\phi i t_\phi}^{\Phi \cup \{\phi, \tilde{\phi}\}} - q_{c_\phi i t_\phi}^{\Phi \cup \{\phi\}} - q_{c_\phi i t_\phi}^{\Phi \cup \{\tilde{\phi}\}} + q_{c_\phi i t_\phi}^{\Phi}) \right) N_{c_\phi} \leq d_i^0 C_q N_{c_\phi} \end{aligned}$$

Where the last inequality is due to the C_q -submodularity of q .

- (d) $f_{c, i \neq i_\phi, t > t_\phi}(\phi, \Phi \cup \{\tilde{\phi}\}) - f_{c, i \neq i_\phi, t > t_\phi}(\phi, \Phi)$. We provide an analysis for each item i separately. Similarly to the argument in **(Step.1b)**, this component of the marginal potential function in (28) is bounded by the formulation in 33,

$$\sum_{c=1}^n \sum_{t=t_\phi+1}^T \left(\left(q_{c_\phi i t_\phi}^{\Phi \cup \{\phi, \tilde{\phi}\}} - q_{c_\phi i t_\phi}^{\Phi \cup \{\tilde{\phi}\}} - q_{c_\phi i t_\phi}^{\Phi \cup \{\phi\}} + q_{c_\phi i t_\phi}^{\Phi} \right) \sum_{\pi \in P_{t-t_\phi}(c_\phi, c)} \prod_{j=1}^{|\pi|-1} p_{\pi_j, \pi_{j+1}} \right) r_{cit}^\Phi N_c \leq \quad (44)$$

$$\sum_{c=1}^n \sum_{t=t_\phi+1}^T \left(\sum_{\pi \in P_{t-t_\phi}(c_\phi, c)} \prod_{j=1}^{|\pi|-1} p_{\pi_j, \pi_{j+1}} \right) C_q r_{cit}^\Phi N_c \leq \quad (45)$$

$$\sum_{c=1}^n \sum_{t=t_\phi+1}^T \left(\sum_{\pi \in P_{t-t_\phi}(c_\phi, c)} \prod_{j=1}^{|\pi|-1} p_{\pi_j, \pi_{j+1}} \right) C_q d_i^0 N_c \leq (T-1) \pi^*(P) C_q d_i^0 \sum_{c=1}^n N_c \quad (46)$$

Where the second to last inequality is due to the C_q -submodularity of the base probability function q .

- (e) $f_{\text{shipping fees}}(\phi, \Phi \cup \{\tilde{\phi}\}) - f_{\text{shipping fees}}(\phi, \Phi)$. Out of the 4 cases that we considered in this part of the proof of Theorem 4, case **(Step.3(a)ii)** is the only case in which we used the assumption in Theorem 4. This is the case in which $AS_{il}^{\Phi \cup \{\phi\}} > 0$ and $AS_{il}^{\Phi \cup \{\tilde{\phi}\}} \leq 0$. Then,

$$\begin{aligned} & f_{\text{shipping fees}}(\phi, \Phi \cup \{\tilde{\phi}\}) - f_{\text{shipping fees}}(\phi, \Phi) \leq -AS_{il}^{\Phi \cup \{\phi, \tilde{\phi}\}} + AS_{il}^{\Phi \cup \{\tilde{\phi}\}} = \\ & \left(\sum_{c: \ell(c)=l} \sum_{t=1}^T N_c (b_{cit}^{\Phi \cup \{\tilde{\phi}\}} - b_{cit}^{\Phi \cup \{\phi, \tilde{\phi}\}}) \right) \cdot s_{il} \leq \delta T s_{il} \sum_{c: \ell(c)=l} N_c \end{aligned}$$

Summing up all the bounds for the different components of the marginal potential, we find the bound in Theorem 3,

$$\begin{aligned} & (d_i C_q + (d_i^0 - d_i) \delta) N_{c_\phi} + (T-1) \pi^*(P) C_q d_i^0 \sum_{c=1}^n N_c + (d_i C_q + (d_i^0 - d_i) \delta) N_{c_\phi} \\ & + (m-2) d_i^0 C_q N_{c_\phi} + (m-1) (T-1) \pi^*(P) C_q d_i^0 \sum_{c=1}^n N_c + \delta T \sum_{l=1}^{NL} s_{il} \sum_{c: \ell(c)=l} N_c \leq \\ & 2d_i^0 \delta N_{c_{max}} + m d_i^0 C_q N_{c_{max}} + m (T-1) \pi^*(P) C_q d_i^0 \sum_{c=1}^n N_c + \delta T \sum_{l=1}^{NL} s_{il} \sum_{c: \ell(c)=l} N_c \end{aligned}$$

Step.2 We now analyze the general case in which $|Y \setminus X| \leq L$. Having **Step.1**, we assume that for any X, Y such that $X \subset Y$ and $|Y \setminus X| = 1$.

$$f(Y \cup \{x\}) - f(Y) \leq C + f(X \cup \{x\}) - f(X).$$

Let us denote by x_1, \dots, x_j , the set of items in $\setminus X$ in an arbitrary order. Then, based on the results of **Step.1**,

$$\begin{aligned} f(Y \cup \{x\}) - f(Y) &\leq \\ f(Y \setminus \{x_1\} \cup \{x\}) - f(Y \setminus \{x_1\}) + C &\leq \dots \leq f(Y \setminus \{x_1, \dots, x_j\} \cup \{x\}) - f(Y \setminus \{x_1, \dots, x_j\}) + j \cdot C \\ &\leq LC + f(X \cup \{x\}) - f(X). \end{aligned}$$