Marketing in the Internet Age

by

Feihong Xia

B.S in English & International Business, Shanghai Institute of Foreign Trade, 2005

M.B.A in Marketing & Supply Chain, Pennsylvania State University, 2011

Submitted to the Graduate Faculty of

Katz Graduate School of Business in partial fulfillment

of the requirements for the degree of

PhD in Marketing

University of Pittsburgh

2016

UNIVERSITY OF PITTSBURGH

KATZ GRADUATE SCHOOL OF BUSINESS

This dissertation was presented

by

Feihong Xia

It was defended on

April 28, 2016

and approved by

J. Jeffrey Inman, Professor, University of Pittsburgh

Jerrold H. May, Professor, University of Pittsburgh

Min Ding, Professor, Pennsylvania State University

Dissertation Advisors:

Rabikar Chatterjee, Professor, University of Pittsburgh

R. Venkatesh, Professor, University of Pittsburgh

Marketing in the Internet Age

Feihong Xia, PhD

University of Pittsburgh, 2016

The Internet has changed the landscape of marketing, and presented us with new marketing phenomena and challenges. In this dissertation, I focus on two emerging research topics at the frontier of marketing research –capturing consumer purchase patterns from big data and online groupbuy.

Marketers have recognized that the probability of a consumer's (or household's) purchase in a particular product category may be influenced by past purchases in the same category, and also by purchases in other, related categories. Past studies of cross-category effects have focused on a limited number of product categories, and have often ignored intertemporal effects in their analyses. The availability of such enormous consumer shopping data sets, and the value of analyzing the complex relationships across categories and over time (for example, for personalized promotions) suggest the need for computationally efficient modeling and estimation methods. We explore the nature of intertemporal cross-product patterns in an enormous consumer purchase data set, using a model that adopts the structure of conditional restricted Boltzmann machines (CRBM). Our empirical results demonstrate that our proposed approach, employing the efficient estimation algorithm embodied in the CRBM, enables us to process very large data sets, and to capture the consumer decision patterns, for both predictive and descriptive purposes, that might not otherwise be apparent.

Online group buying, under which the seller offers discounts based on the size of the pool of buyers, is rapidly growing in popularity in both developed and emerging markets. Our study categorizes the diverse mechanisms across group buying sites into three main types – volume strategy (e.g., Groupon), collective buying (e.g., Mercata), and referral reward (e.g., LivingSocial) – and associated subtypes. Our seller faces a market comprising four segments of consumers who are heterogeneous in their product knowledge or intrinsic valuation or both. Informed consumers may inform and raise the valuation of their less informed peers. Consideration of a broader strategy space combined with a richer consumer behavior model, relative to the extant literature, provides new insights on when and how specific strategies or subtypes are optimal. Collective buying emerges with the largest domain of optimality, followed by referral reward strategy and then the volume strategy. Within collective buying, its subtypes favoring limited penetration are often optimal.

TABLE OF CONTENTS

LIST of TABLES

LIST OF FIGURES

PREFACE

I would like to thank my advisors, Rabikar Chatterjee and R. Venkatesh, for their great mentoring.

Thanks also to Jeff Inman, Jerry May and Min Ding for lots of inspirations and comments.

Finally, lots of thanks are due to the PhD office, and the entire marketing faculty at University of Pittsburgh, for helping me with my PhD study and my dissertation.

1.0 CHAPTER ONE. CAPTURING CROSS-PRODUCT CATEGORY AND INTERTEMPORAL DECISION PATTERNS FROM LARGE CONSUMER SHOPPING DATASETS

1.1 INTRODUCTION

The emergence of big data and of mobile marketing has changed the landscape of marketing. Now, nearly all retailers keep records of consumer shopping histories, and seek to gain useful insights from that data, to improve the effectiveness of marketing efforts. Personalized promotion has become increasingly popular, thanks to the advancement of mobile technology. Retailers such Target and Walgreens have launched mobile apps that send consumers personalized coupons and advertisements. Other intermediaries, such as Coupons.com and Alibaba, now offer mobile coupon services to all types of retailers. In order to target consumers effectively, and to improve personalized marketing practices, such as mobile couponing, it is critical that marketing managers should “get the timing right” and “rely on data to make offers relevant” (Miles 2013). Personalized coupons should be based on consumers’ purchasing patterns, captured from shopping data, and the coupons should be sent at the right time, when they are most effective with consumers. For personalized marketing, marketers need a model that captures and predicts intertemporal shopping patterns efficiently from large sets of consumer data, to assist in forecasting consumers' future shopping baskets. To meet such needs, we propose a model that captures intertemporal, multi-category purchasing patterns from large consumer data sets.

Consumer purchasing decisions are related across categories and over time. In order to estimate the intertemporal cross-category purchasing patterns accurately, we need a model that potentially includes all such effects, while also being able to efficiently estimate its parameters, and to present its results in a way that supports insightful post-estimation analysis. Our aim is to capture purchasing patterns from massive consumer data efficiently, without having to resort to supercomputing.

Our ability to model intertemporal cross-product effects on a major scale has been influenced by two factors. First, we now have access to massive amounts of consumer- (household-) level purchase data, recorded at the point-of-sale by individual retailers, and compiled by marketing research companies, such as A.C. Nielsen. Second, motivated by the availability of rich and massive data sets, efficient methods have been developed that permit the analysis of such data. However, most current approaches in marketing models follow the econometrics tradition, emphasizing unbiased parameter estimation, but not algorithm efficiency (Mehta 2007). Machine learning models, such as neural networks, are able to handle large datasets, but researchers often find it difficult to interpret the results of such approaches. In order to model large-scale, intertemporal, cross-category effects using the massive datasets available today, we need a model whose parameters may be estimated with sufficient efficiency, but that also has a tractable and clear model structure that permits meaningful interpretation of the results.

Previous research on multi-category consumer purchasing decisions focused primarily on a limited number of product categories. Further, discussion of the intertemporal aspects of multi-category buying decision is sparse (Manchanda et al. 1999, Seetharaman et al. 2005). The typical approach is to preselect a few categories, which are expected be either complements or substitutes, and then employ multivariate probit (MVP) or multivariate logit (MVL) methods to model the cross-effects. Product pairs such as spaghetti and sauce have often been chosen for analysis (Manchanda et al. 1999, Duvvuri et al. 2007, Russell and Peterson 2000). The limitation of focusing on a few preselected categories is that the analysis is likely to miss significant intertemporal cross-category effects. Furthermore, for business practitioners, all significant associations between product categories should be considered for designing promotions and other marketing tactics, with overall profit maximization in mind (Shocker et al. 2004).

A more recent stream of literature in marketing focuses on empirical methods for handling high-dimensional data. An intuitive approach is to first reduce the dimensionality to a manageable level, and then to conduct the analysis on the lower dimensional data set (Lawrence 2004). Latent linear models, such as factor analysis and principal component analysis (PCA), are the most popular dimensional reduction methods in marketing research. However, the benefits of applying factor analysis or PCA in multi-category analysis are not clear. Duvvuri (2010) proposed a Bayesian multi-level factor model for consumer price sensitivity analysis across product categories, which captures the cross-category price sensitivities by latent factors, instead of by a covariance matrix. While the factor model does indeed reduce the number of parameters needed to model the underlying associations, the estimation procedure is not efficient enough to handle high-dimensional data. Usually, Markov Chain Monte Carlo (MCMC) sampling methods for model inference can be very time consuming as the number of parameters increases.

A model that utilizes a sparse representation of the associations between product purchasing decisions, along with an efficient inference algorithm, is needed for high-dimensional, multi-category purchasing decision analysis. The restricted Boltzmann machine (RBM) possesses such advantages, and it is the basic building block for our proposed model. Hruschka (2014) employed the RBM model for multi-category analysis, and showed the efficiency of the basic RBM model. Our model is a significant extension of RBM that captures intertemporal effects as well as contemporaneous multi-category cross-effects, and that allows for additional variables, such as demographics, to influence the purchasing patterns. To the best of our knowledge, this is the first time such a model has been adapted for application to a marketing problem.

In summary, this paper:

- Proposes an efficient model for the analysis of intertemporal cross-category consumer purchasing patterns on a very large scale, and compares the performance of that approach with current popular models, and
- Interprets the intertemporal cross-category effects induced by the model, and motivates how those findings may be applied to improve marketing decisions in practice, for example, in personalizing mobile couponing.

The rest of the paper is organized as follows. Section 2 presents the proposed model and the estimation algorithm. Section 3 describes the data used in our empirical analysis. Section 4 discusses and interprets the results, including a comparative evaluation of the model's performance. Section 5 examines the managerial implications of our findings, and Section 6 concludes the paper with a summary and ideas for future research.

1.2 MODEL DESCRIPTION AND ESTIMATION

We first describe the basic structure of the restricted Boltzmann machine (RBM) as it applies to modeling large cross-category purchase data (Section 2.1). Next, we extend RBM, using a Conditional RBM structure to model the contemporaneous *and* intertemporal cross-category effects efficiently, while also allowing for the inclusion of other variables, such as demographics, that might be expected to impact consumer behavior (Section 2.2). Section 2.3 describes the estimation procedure (Section 2.3).

1.2.1 Restricted Boltzmann Machine

RBM is a random Markov field (RMF) with one visible layer and one hidden layer, with connections only between the visible layer and the hidden layer (Hinton 2002; Pearl 2014). The imposition of a bipartite structure improves efficiency, without compromising the model's capability for pattern recognition (Burnap et al 2014). In our model setting, the visible layer consists of binary variables representing observed purchasing decisions for a single customer or household. Let I denote the number of product categories and let t denote a time period, typically a week. The variable x_{it} in the visible layer takes on the value 1 if the customer purchased at least one unit of one item from category i in period t, and 0 otherwise. The probability that a consumer purchases an item from product category i in week t is denoted by $\boldsymbol{P(x_{it}=1)}$, $i = 1\ldots I$. For multi-category analysis, our task is to model the joint distribution of $P(x_{1t}, x_{2t} ... x_{It})$.

In the multivariate probit (MVP) model, the symmetric covariance matrix captures the cross-category associations explicitly. The number of entries in a covariance matrix grows at an

O(I^2) rate, where I denotes the number of product categories, so that the MVP approach becomes rapidly inefficient (computationally) as the number of product categories increases. In contrast, the RBM adopts an approach to capturing relationships among the variables at time period t, using fewer parameters than are necessary for MVP. Specifically, by using hidden variables to capture relationships among the variables in the visible layer, the complexity of the RBM model grows at a rate less than O(I). For each product category i at week t, a product purchase decision can be modeled as a binary outcome that is influenced by the hidden variables.

$$P(x_{it} = 1 | H_t) = \frac{1}{1 + \exp(-\beta_i - \sum_J \beta_{ji} h_{jt})}, \tag{1}$$

where H_t is a J dimensional hidden vector, $H_t = (h_{1t}, h_{2t}, \cdots, h_{(J-1)t}, h_{Jt})$, and β_i is a constant term for each product category.

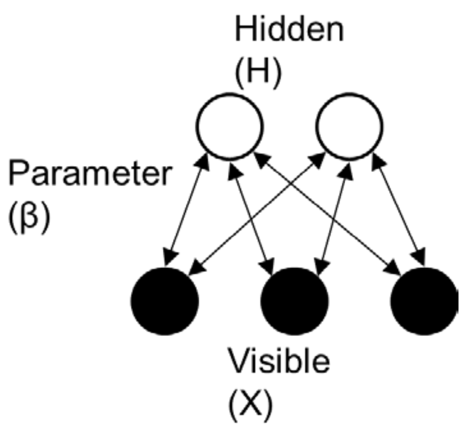
The inclusion of hidden variables in the RBM model is consistent with the intuition that purchasing decisions are influenced by unobservable consumer attributes. For instance, some consumers may be more likely to buy ramen noodles, frozen pizza and beers than others, and unobservable consumer attributes, such as laziness, might lead to such idiosyncratic shopping behaviors. Instead of modeling the decisions explicitly, we can employ unobservable consumer attributes to explain the relationships among purchasing decisions. Being unobservable, we model them as binary hidden variables, and infer their values based upon the observed purchasing decisions. In the RBM model, if a hidden variable is turned on, certain associations among category purchases will be activated. Because of the complexity of purchasing behaviors, it is usually necessary to use several hidden variables h_j (j = 1,…, J) to describe a consumer's buying patterns.

Latent state models, such as the Hidden Markov Model (HMM), also use hidden variables, but the RBM model makes more efficient use of them than does the HMM. The J hidden variables in the RBM are able to represent up to 2^J hidden states. The HMM assumes that the observed variables can only belong to one hidden state, and, as a result, there are only J hidden states in total, given J hidden variables (Salakhutdinov et al. 2007). The RBM has a symmetric bipartite structure; hence, conditioning on all the purchasing decisions, the hidden variables are also independent of each other.

$$P(h_{jt} = 1 | X_t) = \frac{1}{1 + \exp(-\beta_j - \sum_I \beta_{ji} x_{it})}, \qquad (2)$$

which implies that if we observed the purchasing outcomes, we should be able to infer the hidden labels in reverse, as well. The structure of the RBM is represented graphically in Figure 1.

Figure 1 Graphical Representation of the Restricted Boltzmann Machine



The joint distribution of both hidden and visible variables is defined as:

$$P(X_t, H_t) = \frac{\exp(\sum_J \sum_I \beta_{ij} x_{it} h_{jt} + \sum_I \beta_i x_{it} + \sum_J \beta_j h_{jt})}{Z}, \qquad (3)$$

where the exponentiation ensures a positive probability, and z is the normalizing constant. We can sum over H_t to get the marginal distribution of the vector X_t, where each unit in the vector represents a category purchasing decision at time t.

$$P(X_t) = \frac{\sum_H \exp(\sum_J \sum_I \beta_{ij} x_{it} h_{jt} + \sum_I \beta_i x_{it} + \sum_J \beta_j h_{jt})}{z}. \tag{4}$$

The simple 2-layer structure of Figure 1 is advantageous for large-scale data analysis, because it reduces the number of parameters needed for modeling the relationships among the variables, and it makes inference efficient. That efficiency is a critical feature that enables researchers to greatly scale up the model, while ensuring computation feasibility of parameter estimation.

1.2.2 Conditional RBM

The basic RBM model only captures contemporaneous cross-category effects. We propose to use its extension, the Conditional RBM (CRBM), to model both contemporaneous and intertemporal effects (Taylor and Hinton 2009).

A consumer's past purchasing decisions have been shown to influence future purchasing decisions, and marketing models employ past purchasing decisions as predictors (Fader et al. 2005; Keenan 1982). However, such models focus on a product category's own effect, without incorporating intertemporal, cross-category effects. Our model captures intertemporal cross-category effects to present a significantly more complete picture of consumer purchasing patterns.

Diversity in consumers' demographic backgrounds might account for some of the heterogeneity in purchasing patterns. We let demographic variables have direct effects on each product purchasing decision and on the hidden variables. They could also moderate the intertemporal effects and the relationships among decisions at time *t*. The addition of past purchasing decisions and demographic variables does not change the bipartite structures of the RBM. The hidden variables are still independent of each other, conditioned on the visible variables, and vice versa, so that the model inference procedure remains efficient. The extended model is given by Equations (5) - (7):

$$P(X_t, H_t \mid D, X_{<t}) = \frac{\exp(-\sum_I \beta_{it} x_{it} - \sum_J \beta_{jt} h_{jt} - \sum_I \sum_J \beta_{ij} x_{it} h_{it} - \sum_I \sum_J \sum_P \beta_{ijp} x_{it} h_{jt} d_p)}{z}, \quad (5)$$

$$\beta_{it} = \beta_i + \sum_{I'} \sum_M \sum_P \beta_{ii'p}^{t-m} x_{i'}^{t-m} d_p \text{, and} \quad (6)$$

$$\beta_{jt} = \beta_j + \sum_{I'} \sum_M \sum_P \beta_{ji'p}^{t-m} x_{i'}^{t-m} d_p , \quad (7)$$

where *M* is the number of time lags. *P* is the number of demographic variables, and β_{it} and β_{jt} are dynamic terms that affect the visible variable x_{it} and the hidden variable h_{jt} respectively. Both dynamic terms represent the effects from past purchasing decisions, which may be moderated by demographics. In the full form of the CRBM model, demographics and past purchasing decisions may also have direct effects on the hidden or visible variables. The effects can be added to the dynamic terms as follows:

$$\beta_{it} = \beta_i + \sum_{I'} \sum_M \sum_P \beta_{ii'p}^{t-m} x_{i'}^{t-m} d_p + \sum_P \beta_{ip} d_p + \sum_{I'} \sum_M \beta_{ii'}^{t-m} x_{i'}^{t-m} \text{, and} \quad (8)$$

$$\beta_{jt} = \beta_j + \sum_{I'} \sum_M \sum_P \beta_{ji'p}^{t-m} x_{i'}^{t-m} d_p + \sum_P \beta_{jp} d_p + \sum_{I'} \sum_M \beta_{ji'}^{t-m} x_{i'}^{t-m} , \quad (9)$$

where β_{ip} is the direct effect of demographic attribute *p* on purchasing decision *i* at time *t*, and $\beta_{ii'}^{t-m}$ is intertemporal effect on *i* of a purchase from category *i'* at time *t-m*. By the same logic, β_{jp} and $\beta_{ji'}^{t-m}$ represent similar effects on the hidden variables.

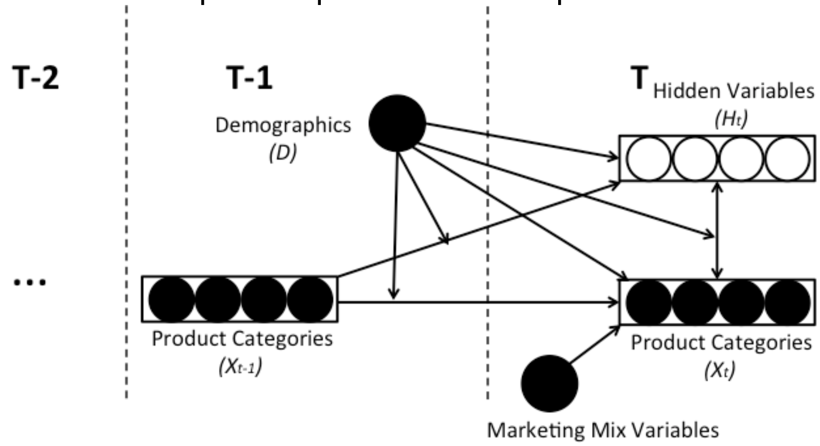
Heterogeneity among consumers may be further captured by changing the constant term β_i to β_{ni}, which represents the baseline probability of purchasing a product for each individual *n*. The dynamic term β_{it} then changes to

$$\beta_{njt} = \beta_{nj} + \sum_{I'} \sum_{M} \sum_{P} \beta_{ii'p}^{t-m} x_{i'}^{t-m} d_p + \sum_{P} \beta_{jp} d_p + \sum_{I'} \sum_{M} \beta_{ii'}^{t-m} x_{i'}^{t-m} . \qquad (10)$$

We can also integrate a hierarchical Bayesian model with CRBM to form a "Hierarchical Deep Model" (HD), representing the heterogeneity among consumers (Salakhutdinov et al. 2013). However, modeling heterogeneity in either manner did not reduce the testing error, and therefore neither is included in the final model. The Hierarchical Bayesian extension of the model is described in Appendix 2. For more information on HD models, please refer to Salakhutdinov et al. (2013).

For the empirical analysis in this paper, we focus on modeling heterogeneity in terms of demographics. We can also incorporate price and promotion variables to model their cross and own effects on product purchasing decisions. Figure 2 provides a graphical representation of the full model, constructed to capture all the effects and interactions discussed above.

Figure 2 Graphical Representation of Proposed CRBM Model



In order to keep such a complex model sparse and efficient, we factorized the parameter matrices (Taylor and Hinton 2009). For example, we assumed that demographics would moderate the associations between visible variables and hidden variables, which is a three-way interaction modeled by the β_{ijp} terms. Explicitly modeling all possible three-way interactions requires $I \times J \times P$ parameters. If, instead, we break down the three-way interactions into pairwise interactions, the interaction term reduces to:

$$\sum_I \sum_J \sum_P \beta_{ijp} x_{it} h_{jt} d_p \approx \sum_F \sum_I \sum_J \sum_P \beta_{if} x_{it} \beta_{jf} h_{jt} \beta_{pf} d_p \,. \tag{11}$$

That manipulation reduces the number of parameters to $F \times I + F \times J + F \times P$, and we can adjust the number of factors F to control the total number of parameters. Similar procedures were applied to other three-way interaction terms, to reduce the growth rate from $O(N^3)$ to $O(N^2)$.

Unlike the intertemporal effects, the cotemporaneous effects are captured implicitly by the hidden variables. We can infer the cross-effects from the partial derivative $\frac{\partial x_{it}}{\partial x_{jt}}$. By the chain rule, $\frac{\partial x_{it}}{\partial x_{jt}} = \frac{\partial x_{it}}{\partial H_t} \frac{\partial H_t}{\partial x_{jt}}$, and the partial derivative $\frac{\partial x_{it}}{\partial x_{jt}}$ changes with the values of x_{it} and x_{jt}. We replace x_{it} and H_t by their expected values (Hruschka 2014), which can be estimated efficiently by variational inference (Appendix 3).

1.2.3 Estimation

The parameters of the CRBM model are estimated numerically, by searching for the values that maximize the log likelihood of the training data. A first-order algorithmic approach to that maximization problem would be to calculate the gradient of the log-likelihood function with respect to the parameters, and then to perform a search using that gradient. Thanks to the bipartite structure of the RBM, each component of the gradient has the form

$$\frac{\partial \log P(x_{it})}{\partial \beta_{ij}} = E(x_{it} h_{jt})_{data} - E(\hat{x}_{it} \hat{h}_{jt})_{\mathrm{model}} \qquad (12)$$

where β_{ij} is the parameter capturing the interaction between the purchase decision variable for category i and hidden variable j. The first term in Equation (12) is the expected value of the product of x_{it} and h_{jt} over households and time periods in the training data, and the second term is the expected value over the distribution of the same product assumed by the model. The second term could be estimated by brute force by conducting Gibbs sampling to generate samples from the model, with random initial values for $\hat{\boldsymbol{x}}_{it}$ and $\hat{\boldsymbol{h}}_{it}$, and then calculating the average of the product of $\hat{\boldsymbol{x}}_{it}$ and $\hat{\boldsymbol{h}}_{it}$ for each product category. However, such a naïve sampling approach would require a lot of training time, which is not ideal for large data analysis.

Various methods have been proposed for making the inference procedure $E(x_{it} h_j)_{data}$ efficient. Contrastive Divergence (Hinton 2002) estimates the data-dependent statistics in

Equation (12) by a one-step sampling process. The hidden variables *H* are drawn, conditional on the original training data. For the data-independent term, $E(\hat{x}_{it}\hat{h}_j)_{\text{model}}$, we first sample hidden units from the original data, and then sample simulated data from the sample hidden units. The average value of the product of the simulated data and the hidden variable sample is then used as the data-independent term in the gradient. While one-step Gibbs sampling cannot yield an accurate estimate of $E(\hat{x}_{it}\hat{h}_j)_{\text{model}}$, the gradient based on the sampling method has been shown to work well in practice (Hinton 2002).

We shortened the time to convergence in parameter estimation by using the momentum method and batch gradient descent. The momentum method is used to speed up the learning rate when the direction of the gradient stays the same on consecutive iterations, and to slow it down when the gradient changes direction. The intuition behind the momentum method is illustrated by the simple analogy of a ball rolling down the side of a steep valley, gaining momentum until it reaches the bottom, but losing momentum as it rolls past the minimum up the other side of the valley. The ball eventually settles at the bottom (minimum). The update equation is as follows (Hinton 2010):

$$\Delta\beta_{ij}(t) = \alpha\Delta\beta_{ij}(t-1) + \varepsilon \frac{\partial \log P(x_{it})}{\partial \beta_{ij}}(t), \tag{13}$$

where *t* and *t*-1 indicates the current and previous iterations, $\Delta\beta_{ij}(t)$ is the parameter increment at *t*-1, α is the momentum, and ε is the learning rate. Log $P(x_{it})$ is the log likelihood function that we are trying to maximize.

Dividing the training data into batches, and updating the parameters after each batch, can also speed up the model training process, because a group of smaller matrices can typically be processed faster than can a single, very large matrix. Both gradient descent and stochastic gradient descent can be viewed as variations of batch gradient descent. If we use each single data point as a batch, it becomes stochastic gradient descent. If we use the whole training data as a batch, then it is traditional gradient descent. In our empirical application, we used each (individual) household's shopping data as a batch.

In order to prevent overfitting, in addition to conventional regularization, we used the dropout method (Hinton 2012), and found that it prevented overfitting and improved the model's predictive performance. The dropout method requires a change to the implementation of the training algorithm. During each batch training phase, some hidden variables are randomly dropped from the model, with probability p. After the model training is complete, the parameters are multiplied by p, and all hidden variables are retained in the final model. The dropout procedure prevents hidden variables from trying to fit the idiosyncratic characteristics of the training data, and it may be viewed as a form of model averaging.

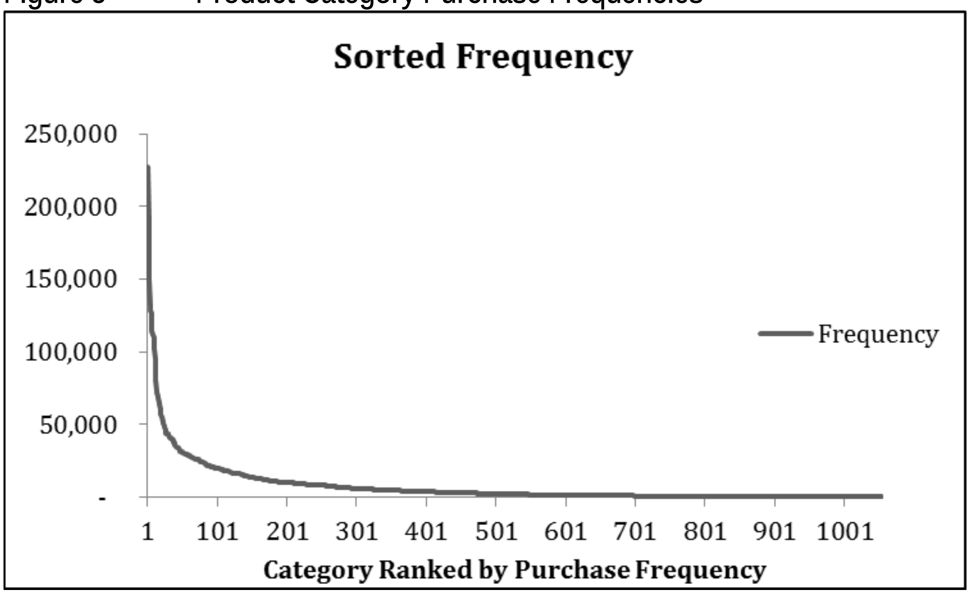
The complete training algorithm for the final model is described in Appendix 4. A detailed discussion of efficient training algorithms for the RBM can be found in Salakhutdinov and Hinton (2012).

1.3 THE DATA

The consumer panel data set we analyzed was provided by A.C. Nielsen for the sole purpose of academic research. The data set contains households' shopping histories, as well as their demographic background information. There are, in total, 6,695,833 observations in the data set. Each observation is a purchase transaction made by a household between 2004 and 2008. In the data set, the three stores most frequently visited by consumers are Kroger, Walmart, and Meijer.

There are 4000 households' purchasing histories over 211 weeks and across 1055 product categories (defined by A.C. Nielsen). Each household in the data set made, on average, 2.74 shopping trips per week. Detailed demographic information for each household is also available. We selected 2086 households with complete purchasing history and demographic information for our analysis. The 1055 product categories vary significantly in the frequency of their purchase. Milk is the most frequently purchased category; other categories, such as pain relief medication, are purchased far less frequently. A plot of the category purchase frequencies, with the categories sorted in decreasing order of frequency, is shown in Figure 3.

Figure 3 Product Category Purchase Frequencies



The frequency plot shows a "long tail" of infrequently purchased product categories. 100 products were purchased more than 20,000 times by households in the dataset. For our illustrative empirical application, we selected the 100 most purchased categories for our analysis, including products from dairy milk to nutritional supplements.

We converted each household's purchasing history into a binary matrix. The rows represent the 211 weeks recorded in the dataset and the columns represent the 100 product categories. The product categories are ordered in decreasing overall purchase frequency. For example, if a household purchased at least one unit of one item from the dairy milk category in Week 1, then the value in Row 1, Column 1 is 1. If the household did not make such a purchase, then the entry is zero. There are 2086 such matrices in all, one for each selected household. We randomly selected 1500 matrices for the training set. The remaining 586 matrices were used for testing the model.

We conducted cross validation on the training data to select the optimal level of complexity (Friedman et al. 2001). Only demographic variables which have shown significant effects on purchasing decisions in previous research (e.g., the presence of children, household income, and household size) were initially included in the full model (Manchanda et al. 1999; Hansen et al. 2006). The results indicated that only the inclusion of household size would further reduce the estimated prediction error.

We also trained and tested models with promotion and price indices at the product category level. The price and promotion indices were constructed following approaches in previous multi-category research (Manchanda et al. 1999, Duvvuri et al 2007, Duvvuri and Gupta 2010). For the price index, if a household purchased a product, the unit price the household paid was used. If the household did not make a purchase in a product category that week, a weighted average based on the household's purchasing history across all brands in that category was constructed as the price index. The promotion index was constructed in the same fashion. If there was a promotion, the index was one; otherwise it was a weighted average.

As discussed in marketing literature, the operationalization of promotion and price at the product category level may not be accurate (Manchanda et al. 1999, Duvvuri et al 2007, Hruschka 2014). Especially in our case, there is a large amount of price and promotion information missing, and many households seldom purchase products outside of the top 40 most purchased product categories. As for the promotion index, the operationalization may manufacture strong correlation with the purchasing decision. A promotion is only recorded when a household made a purchase; therefore, if the promotion index is one, it is a prefect predictor that the household has made the purchase. Because coupon or sales promotion would result in reduced priced paid, price and promotion indices are also highly correlated in the dataset. Possibly due to the inaccurate measure, some multicategory studies that have used similar operationalization of price and promotion have found the price effects to be either marginal or insignificant (Russell and Peterson 2000, Boztug and Hildebrandt 2008, Hruschka 2014). By including price index in the model, we found that the model performance decreased in terms of testing error. The incorporation of promotion index reduced testing error, however, as aforementioned, it might just be the result of imperfect operationalization of promotion. We compared the model with and without promotion index, and have found no significant difference in the analysis results of the cross-category associations.

The initial values for the learning rate and for the regularization penalty term were set at 0.01 and 0.0002, respectively. A momentum term of 0.9 was applied to the training process after five epochs. We used a recommended dropout rate of 0.5 for training (Srivastava et al. 2014). Those values are typical for CRBM training (Taylor and Hinton 2009), and we did not find a significant difference in results for different values (consistent with Hinton 2010). Each household’s purchase history, over the 211 week period, was used as a mini-batch during training.

1.4 RESULTS

1.4.1 Model Comparison

We first compared the model's performance with that of arguably the most popular "traditional" multi-category model employed for the explicit representation of cross-effects – multivariate probit (MVP). As discussed earlier, the major limitation of traditional models, such as the MVP, is the small number of categories that can be modeled, owing to computational constraints. Therefore, when comparing the CRBM method with the MVP, we limited the analysis to four categories: soup, pasta sauce, dog food, and pasta. These four categories have been widely used in multi-category research (pasta and pasta sauce are complements; dog food is considered to be independent of the other three categories).

The purchasing decisions for these 4 product categories for the previous 4 weeks, along with household size information, were included as independent variables in all models. We used the mean squared error (the Brier Score), as the error measure for both testing and training errors. The testing error should be a better metric for model comparison and model selection because it is a better approximation of the generalization error than are other metrics, such as AIC or BIC, which are based on training errors (Friedman et al. 2009). The results for the two models are displayed in Table 1.

Table 1 Model Comparison: Brier Scores for MVP and CRBM Models

Testing Error	Soup	Pasta Sauce	Dog Food	Pasta
MVP	0.1146	0.0651	0.0271	0.0486
CRBM	0.1143	0.0650	0.0274	0.0485
Training Error	**Soup**	**Pasta Sauce**	**Dog Food**	**Pasta**
MVP	0.1212	0.0601	0.0202	0.0399
CRBM	0.1212	0.0601	0.0206	0.0399

Both models have similar performance, in terms of both testing error and training error at relatively low dimension. However, the training time for the models was vastly different. We trained both models on the same computer and recorded the training time. For the MVP model, the training time was over 8 hours in R and around 7 hours in STATA. It took less than 30 seconds to train the proposed CRBM model. Using the CRBM instead of the MVP model results in an enormous improvement in efficiency in terms of model training time, and the drastic reduction in training time is critical when it comes to large data analysis. The comparison demonstrates that our model is much more efficient than the popular MVP models, even with relatively small dimensional data.

1.4.2 Large-scale Cross-category and Intertemporal Effects

By first comparing our model with other models, we demonstrated its significantly greater efficiency with low dimensional data. In order to capture purchasing patterns more comprehensively, we expanded our analysis to include all 100 product categories, and time horizons beyond 4 weeks. Estimation of the more traditional multivariate models is simply not feasible in this case, given the size of the data, and we focus instead on comparing different specifications of the CRBM model.

We first trained the model with 80 hidden variables, keeping the structure the same as in the four-product category model with a four- period time lag (Model 4P4T).[1] Next, we included all 100 categories with the same four-period time lag (Model 100P4T). Comparing the two models

---

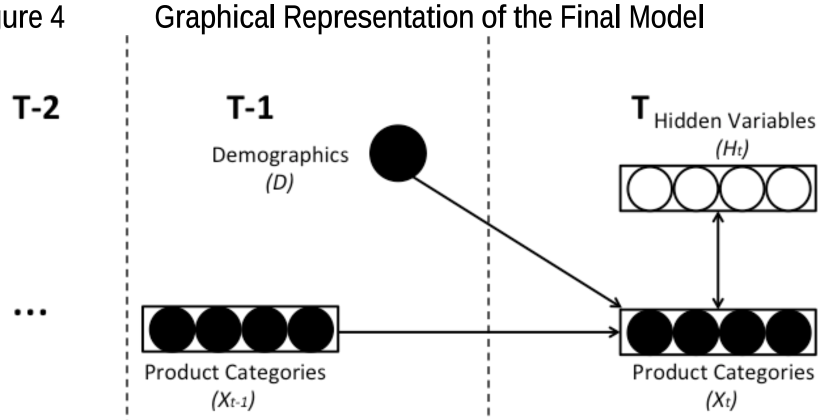
[1] The usual practice is to set the number of hidden variable to slight below the number of visible variables (100 in our case). Dropout training (which we use) essentially prevents overfitting from an initial specification of too many hidden variables.

in terms of their predictive performance (predicting the purchases in the four categories – soup, pasta sauce, dog food, and pasta), the testing errors decreased with the inclusion of the additional categories. For instance, the testing error for soup purchases dropped from 0.1143 to 0.1115, while the testing error across all categories dropped from 0.0638 to 0.0608, suggesting the presence of intertemporal cross-category effects beyond just the four categories. Clearly, such an expanded analysis would not be feasible using traditional multivariate models.

Next, we extended the number of time lags considered in the model to investigate the presence of intertemporal effects beyond the previous four weeks. We compared models with different time lags, and discovered that the testing error dropped as the number of lags was expanded to 12 weeks, and then stayed relatively flat. This may be explained by the observation that most products also have a shopping cycle of 12 weeks or less in the data. Therefore, we next extended the model to incorporate a time lag of 12 weeks with both the four and all 100 categories (Models 4P12T and 100P12T).

We also trained and compared CRBM models with and without household demographics as moderators. Results from cross validation showed that demographics were not significant moderators according to the testing errors. The final model is shown in Figure 5.

Figure 4 Graphical Representation of the Final Model



After selecting the final model, we performed non-parametric bootstrapping to test the significance of the parameters, especially the parameters that represent intertemporal effects. Non-parametric bootstrapping makes fewer assumptions about the distributions of data than does parametric bootstrapping, and it does not rely on the correct specification of the model; thus it is a more conservative test than is parametric bootstrapping (Friedman et al. 2009). We drew samples from the data randomly with replacement, and created 100 sets of sample data the same size as the training data. We estimated parameters for each sample data set, and placed a 95% confidence interval on the parameters, to check if the parameter values are significantly different from zero. The bootstrapping process can be time-consuming if the simulations are conducted sequentially. Because bootstrapping samples are independent, we parallelized the sampling tasks, to reduce working time. We estimated a model with only significant parameters according to the bootstrapping confidence interval, and that model (Model 100P12T-SIG), had a smaller test error than did the model with the full set of parameters (Model 100P12T). Table 2 compares the testing errors across the five CRBM models examined – 4P4T, 100P4T, 4P12T, 100P12T, and 100P12T-SIG.

Table 2 Comparison of Model Testing Errors

Testing Error	Soup	Pasta Sauce	Dog Food	Pasta	All Products
4P4T	0.1143	0.0650	0.0274	0.0485	0.0638
100P4T	0.1115	0.0641	0.0274	0.0483	0.0608
4P12T	0.1116	0.0640	0.0261	0.0482	0.0625
100P12T	0.1118	0.0640	0.0264	0.0483	0.0602
100P12T-SIG	0.1112	0.0638	0.0261	0.0482	0.0599
Training Error	**Soup**	**Pasta Sauce**	**Dog Food**	**Pasta**	**All Products**
4P4T	0.1212	0.0601	0.0206	0.0399	0.0605
100P4T	0.1183	0.0592	0.0204	0.0395	0.0595
4P12T	0.1167	0.0586	0.0192	0.0390	0.0584
100P12T	0.1158	0.0580	0.0186	0.0385	0.0576
100P12T-SIG	0.1156	0.0582	0.0189	0.0388	0.0576

Note: 4P and 100P indicate CRBM models with 4 and 100 product categories respectively. 4T and 12T represent 4 and 12 time lags in the models. SIG means the model only includes significant parameters according to the bootstrapping results. The All Products column shows the mean errors over all product categories in the model.

The results further support the presence of intertemporal cross-category effects beyond just a limited number of products and time periods. However, the full (more complex) model tends to overfit the training data if all effects are included. Thus, the appropriate model in this case is 100P12T –SIG, which includes only statistically significant effects (based on the bootstrapping-based confidence intervals).

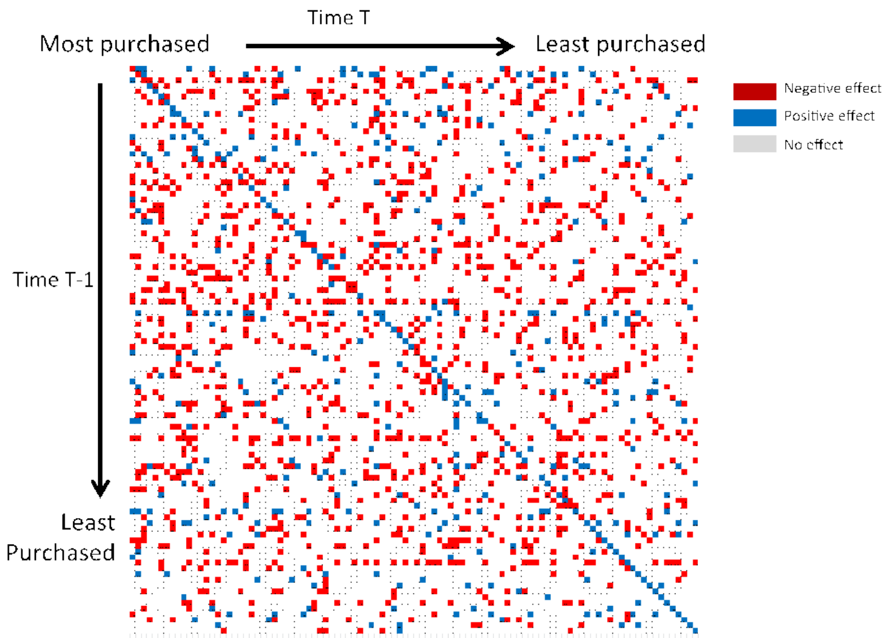
1.4.3 Examination and Discussion of Intertemporal Cross-Category Effects

Note that the proposed CRBM model is designed to predict shopping behavior of a household on the basis of past behavior (say, over the last 12 weeks), the hidden variables associated with the household, and key demographics. In this sub-section, we discuss the pattern of intertemporal cross-category effects based on the model parameter estimates. Given the very large number of such effects (with 100 categories and 12 time periods), presenting these results in a manner that makes sense for the reader poses a challenge. We start by observing the general purchasing pattern, and then drilling down afterwards, to examine the associations between specific categories over time. In order to follow that strategy, we first present visual representations of the intertemporal multi-category effects, shown in Figure 6. Since these effects tend to be weaker for longer time lags, and for ease of exposition, we focus on four (rather than all twelve) lags.

Each column represents the product category purchasing decision at time *T* and each row represents product category purchasing decisions in the past, from *T*-1 (Figure 6(a)) through *T*-4 (Figure 6(d)). The categories are ordered in decreasing purchase frequency from left to right and from top to bottom. The value in each cell represents the association between a category purchase decision in Period *T* and a category purchase decision in a prior period. The associations are shown as color-coded ‘pixels’ – blue for a positive association, red for a negative association, and white to indicate no significant association between the two purchase decisions (using the 95% confidence interval as discussed). The pattern formed by all the pixels provides an initial “big picture”.

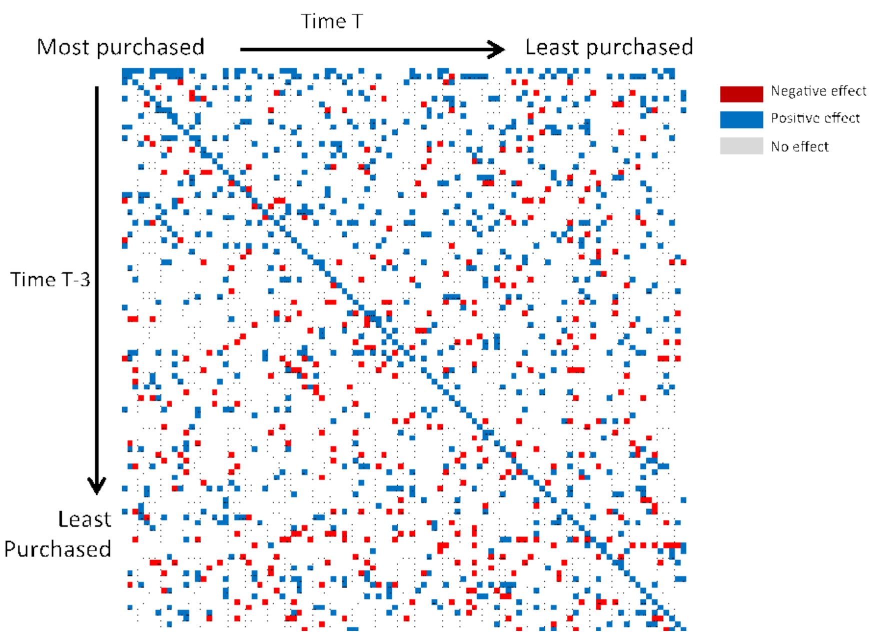
Figure 5 Plots of Intertemporal Effects from One to Four Week Lags

5(a) Intertemporal Effects from T-1 to T

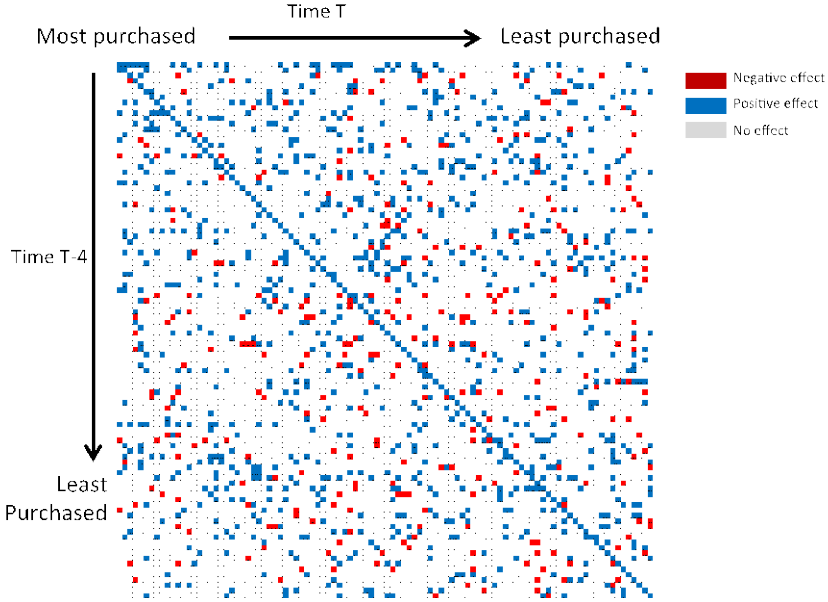


5(b) Intertemporal Effects from T-2 to T

5(c) Intertemporal Effects from T-3 to T



5(d) Intertemporal Effects from T-4 to T



In general, the varying patterns of the plots show the dynamic nature of intertemporal effects. Figure 6(a) has the most significant associations, with a large proportion of the cross-category associations being negative. The proportion of significant associations that are negative decreases from Figure 6(a) to Figure 6(d) (the proportions are 76.43%, 38.81%, 32.61% and 23.63% for Figures 6a, 6b, 6c, and 6d, respectively). The trend suggests that last week's shopping has a significant cross-category impact across several product categories, and the impact tends to be negative. For example, if a household bought milk the week before, it is less likely to buy orange juice this week. The own (same-category) effects, however, are mostly positive across all time periods, which indicates that if a household bought a certain product previously, then it is more likely repeat the purchase next week than a household that did not do so. The finding of predominantly positive own effects is consistent with previous research (Fader et al. 2005). The overall pattern at *T*-1 suggests that while a household may shop every week, a major shopping trip in one week is likely be followed by fewer purchases the next week. We plotted random households' weekly total number of purchases as non-model evidence, and the pattern of fluctuations supports the idea of households doing their heavy grocery shopping in alternate weeks.

We examined each column for significant coefficients at *T*-1, and found that the greatest negative associations exist between pairs of typical substitutes, such as between frozen novelties and ice cream, fresh lettuce and precut fresh salad mix, butter and margarine, and sausage and bacon. However, other significant negative effects at *T*-1 appear to capture buying habits and patterns (for example, households do not purchase certain products every week – this may be why the purchase of milk last week and orange juice this week are negatively associated). Other effects appear to reflect lifestyle differences – for example, households that bought canned green beans one week are less likely to buy frozen pizza the following week.

Examining the cross-category positive associations at *T*-1 in detail, several of the product pairs are complements, such as cereal and milk, and cold remedies and facial tissues. Other positive significant associations are not so obvious, such as between hot cereal and fresh carrots.

All four plots indicate that, while nearly all intertemporal own effects are significant and positive, there are also significant intertemporal cross-category effects. These results suggest that including cross-effects over multiple categories and over multiple periods in the model helps to capture a more complete picture of household purchasing patterns, which can be complex, asymmetric, and time-varying.

After observing the pattern at each time lag separately, we examined the intertemporal effects across multiple time periods jointly, to examine which cross-category associations remain significant over several time periods. As expected, fewer effects persist as the number of time period increases. The pairs for which the cross-category coefficients are significant for all four time lags are summarized in Table 3.

Table 3 Product Category Pairs with Significant Intertemporal Effects over all 4 Lags

Category at *T*-1, *T*-2, *T*-3, or *T*-4	Category at *T*	Cross-Category Intertemporal Coefficients			
		T-1	*T*-2	*T*-3	*T*-4
ICE CREAM - BULK	BUTTER	0.050	0.055	0.084	0.065
BAKERY - BREAD - FRESH	PEANUT BUTTER	0.082	0.073	0.071	0.048
BAKERY - BREAD - FRESH	CHEESE-PROCESSED SLICES-AMERICAN	0.066	0.062	0.073	0.078
MARGARINE AND SPREADS	BUTTER	-0.220	-0.205	-0.151	-0.098
CANDY-CHOCOLATE	CANDY-CHOCOLATE-MINIATURES	0.071	0.086	0.082	0.059
SOFT DRINKS - CARBONATED	FRESH STRAWBERRIES	-0.055	-0.067	-0.054	-0.043
BAKERY-BUNS-FRESH	FRANKFURTERS-REFRIGERATED	0.059	0.091	0.073	0.083
BAKERY - BREAD - FRESH	MARGARINE AND SPREADS	0.047	0.090	0.043	0.045
PRECUT FRESH SALAD MIX	SALAD DRESSING - LIQUID	0.178	0.189	0.138	0.074
CHEESE - PROCESSED - CREAM CHEESE	DAIRY - SOUR CREAM - REFRIGERATED	0.088	0.062	0.109	0.100
DAIRY-MILK-REFRIGERATED	EGGS-FRESH	0.042	0.070	0.078	0.055
BUTTER	SUGAR-GRANULATED	0.094	0.141	0.119	0.103
DAIRY-MILK-REFRIGERATED	CEREAL - READY TO EAT	0.087	0.110	0.090	0.064
PRECUT FRESH SALAD MIX	CHEESE - SHREDDED	0.053	0.058	0.064	0.038
DAIRY-MILK-REFRIGERATED	BAKERY - BREAD - FRESH	0.023	0.042	0.051	0.042
CANDY-CHOCOLATE-SPECIAL	FRESH STRAWBERRIES	0.125	0.102	0.198	0.260
FRANKFURTERS-REFRIGERATED	CANDY-CHOCOLATE-SPECIAL	-0.191	-0.109	-0.116	-0.108
FRESH LETTUCE	SALAD DRESSING - LIQUID	0.106	0.123	0.135	0.124
SAUSAGE-BREAKFAST	BACON-REFRIGERATED	-0.098	-0.061	0.085	0.078
SOFT DRINKS - POWDERED	CANDY-CHOCOLATE-SPECIAL	-0.145	-0.086	-0.074	-0.082
VEGETABLES-BEANS WITH MEAT-SHELF	CANDY-CHOCOLATE-SPECIAL	-0.139	-0.080	-0.159	-0.112
CANDY-CHOCOLATE-MINIATURES	SOUP-CANNED	0.050	0.046	0.116	0.063
BUTTER	MARGARINE AND SPREADS	-0.276	-0.181	-0.226	-0.180
CAT FOOD - DRY TYPE	CAT FOOD - WET TYPE	-0.112	0.098	0.201	0.252
ENTREES-REFRIGERATED	VEGETABLES - POTATOES - FROZEN	0.066	0.050	0.058	0.070

Some of the significant associations over the four time periods are between regularly purchased staples (milk and bread) and complementary products (cereal and milk, bread and peanut butter, and salad mix and salad dressing) or substitutes (butter and margarine). Not all the positive (negative) associations are the result of product pairs being complements (substitutes). There are persistent purchasing patterns that are seemingly influenced by each household's lifestyle, the unobservable hidden factor that can only be inferred from one's past purchasing history (Manchanda et al. 2009). For example, the persistent positive associations between purchasing refrigerated entrees and frozen vegetables is likely accounted for by households who want to fix meals quickly and conveniently without starting from scratch. Buying carbonated soft drinks is always a negative sign that the consumer will ever purchase fresh strawberries, which again appears to relate to lifestyle differences.

Certain significant associations change signs through the four time periods, indicative of the complex nature of intertemporal effects. Dry-type cat food and wet-type cat food may be considered substitutes. At time T-1, the purchase of dry type cat food does indicate a lower probability of purchasing wet type food at time T. However, if a household purchased dry-type cat food at times T-2, T-3 or T-4, then it is more likely to purchase wet-type food at time T. Another example: a household that bought breakfast sausage within the prior two weeks is less likely to purchase bacon this week, but the associations turn positive if the purchase of sausage was made three or four weeks earlier. Such behavior may be attributed to variety seeking. Candy chocolate and miniature candy chocolate, which seem to be perfect substitutes, show positive associations for all four time lags. This might reflect variety seeking and/or heterogeneity across households: given the presence of heavy users and light / non-users in this category, previous purchases of one kind of chocolate may suggest a greater likelihood of buying chocolate again, even of a different type.

As emphasized earlier, the key benefit of the proposed CRBM approach is its ability to predict future purchase probabilities at the household level, taking into account previous purchasing patterns, without having to impose limits on the number of product categories and the number of time lags for inclusion in the model. As further demonstration, we randomly picked a household in the dataset that bought groceries in the previous 12 weeks. Based on the household's purchasing history, the model predicted that the six products the household was most likely to buy in the next week were frozen pizza, meat, cat food, fresh salad mix, cake, and frozen dinners. We compared the predictions against the actual purchase data, and found that the household did indeed buy those six products. We made similar predictions for all households in the testing data, and found an average error rate of 18.03%, which means that, on average, approximately five out of every six were predicted correctly. We then compared those results with a prediction based on selecting the six most frequently purchased products for each household over the previous twelve weeks, which appears to be a reasonable benchmark. The average error rate under this household-level frequency-based prediction achieved a 22.09% error rate, which is higher than that for our model.

1.4.4 Contemporaneous Cross-Category Effects

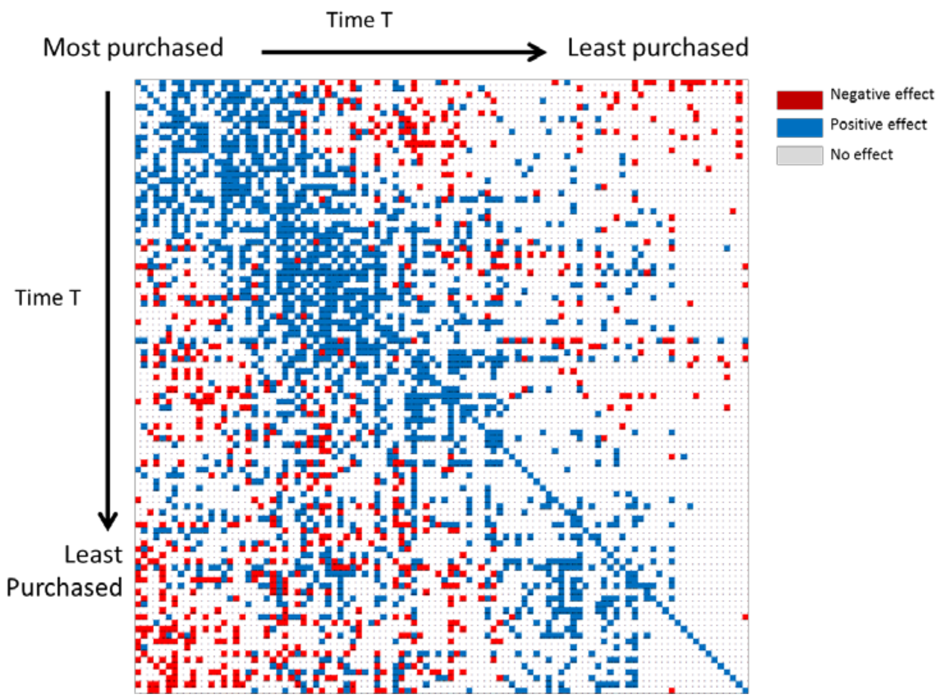
In CRBM (as in RBM), the parameters representing associations between hidden and visible variables representing contemporaneous category purchases are estimated explicitly. However, there are no parameters explicitly representing associations between the contemporaneous category purchases (see Section 2.1 and Hruschka 2014). Therefore, contemporaneous cross-category effects have to be estimated indirectly from the associations between visible and hidden variables. We did so using the method advocated by Hruschka (discussed earlier in Section 2.2), with variational inference (Appendix 3). The resulting plot, in Figure 7, shows that most positive cross-effects are adjacent to the principal diagonal (i.e., between products with similar purchase frequencies), while the negative associations tend to be between categories with disparate purchase frequencies. This pattern seems to confirm previous findings that products with similar purchase frequencies tend to have positive contemporaneous associations with each other (Manchanda et al. 1999). Overall, 38.5% of the significant associations are negative (relative to 76.43% in the case of one period lag, in Figure 6(a))[2]. As expected, there are positive

---

[2] We have not dropped the contemporaneous own associations along the principal diagonal which are (trivially) significantly positive. Since a vast majority of the intertemporal own associations are significantly positive (Figure 6), we retained them in this case (Figure 7) so that comparisons are valid.

associations between complements such milk and cookies, and also negative ones between substitutes, such as cookies and donuts. Interestingly, the contemporaneous association between milk and orange juice is positive; recall that the association with a one period lag is negative.

Figure 6 Contemporaneous Effects



1.4.5 Discussion

The primary purpose of our proposed CRBM model is to predict product purchases at the individual (household) level for non-durables when shopping baskets typically contain an assortment of items chosen from a large number of possible product categories, such as in the context of grocery (supermarket) shopping. In this context, our model allows for the inclusion of a very large number of product categories with potentially interrelated shopping patterns over multiple time periods. The inclusion of such large-scale intertemporal cross-category patterns (possible with our CRBM-based model) potentially allows for better predictions.

As shown in Section 4.1, the proposed model is able to predict at least as well as, and very much more efficiently than, traditional multivariate models, even for small datasets (which can be handled by the latter models). When we expanded the model to consider 100 product categories over 12 time periods (Section 4.2), we discovered significant intertemporal cross-category associations. This illustrates the benefit of not having to impose a priori limitations on the model specification, a freedom allowed by the far greater efficiency of the CRBM approach.

The intertemporal (contemporaneous) cross-category effects discussed in Sections 4.3 and 4.4 are interesting from theoretical and managerial perspectives. However, we note that while our model has a tractable and clear model structure that permits meaningful interpretation of the results, it does not explain why the particular patterns occur, so that our interpretations are speculative and plausible at best. As discussed in Section 4.4, our interpretations do not parcel out individual-level behavioral patterns from cross-household heterogeneity. There could potentially be other confounding effects, such as differences in preferences across members of a household. It should be noted, however, that while the model parameters are estimated at the aggregate level, the behavioral predictions are unique to each household, based on the direct (and potentially moderating) effects of demographics and the role of hidden variables in the restricted Boltzmann machine based approach.[3]

[3] As noted in Section 2.2, heterogeneity can be explicitly modeled in at least two ways, although in our application, neither reduced the testing error, and therefore were not incorporated in the final model.

1.5 MANAGERIAL IMPLICATIONS

The outstanding computational efficiency of the proposed model is critical for analyzing, quickly and inexpensively, the massive amounts of data available to managers, while capturing complex intertemporal cross-category purchasing patterns that would not be possible to discover using more traditional approaches. The ability to predict future behavior based on this analysis is what makes our model a potentially valuable tool for marketing decision making.

One significant application area for our model is personalized mobile couponing. The model can assist managers in multiple ways. It predicts what consumers are most likely to buy next, and marketing managers can send personalized coupons to consumers, based on the predictions, in order to increase the coupon redemption rate. For example, retailers can predict the top six products a consumer is most likely to buy next week, and send volume-based coupons for some of the products, to provide an incentive for the shopper to buy more units than she otherwise would.

The intertemporal aspect also helps marketers decide the timing of mobile coupon campaigns. As our analysis suggests, the intertemporal patterns of cross-category purchases are often complex and non-obvious. Classifying product pairs as complements or substitutes fails to capture this complexity. For instance, our analysis suggests that consumers who bought breakfast sausage are less likely to buy bacon the following two weeks, but more likely to do so after two weeks. Therefore, the coupon might be more effective if sent two weeks after a consumer bought sausage, rather than immediately after the purchase of bacon.

Marketers can also classify the products based on *purchasing patterns*, for better category management and more effective promotions. The intertemporal effects capture how a product purchase decision is influenced by previous purchases. Instead of a traditional classification scheme, marketers can adopt a consumer-oriented approach, and segment the products by their purchasing patterns. Products in the same segment may be promoted and even possibly displayed together, to facilitate the consumers' shopping process. For demonstration, we segmented the 100 products in the model based on their intertemporal purchasing patterns. Products with similar significant intertemporal effects (as captured by the appropriate model parameters) are thus likely to be assigned to the same cluster. We obtained 26 clusters using K-means, with the optimal number of clusters chosen based on the average silhouette width criterion (using a package in R). The results are shown in Appendix 5. While the classifications based on consumer purchasing patterns are largely similar to conventional classification schemes, there are obvious differences. For example, disposable dishes are grouped with carbonated soft drinks, fruit drinks, bottled water, etc. (in Cluster 6), presumably because these products tend to be purchased together for parties. Interestingly, liquid tea does not belong to this group, but is in a cluster of its own (#20).

The proposed model can be applied in more general settings than personalizing mobile coupons in marketing. Better predictions by our model would not only benefit the marketing manager, but also assist in inventory management, by providing more accurate demand forecasts to design stocking schedules.

Besides brick-and-mortar supermarkets, this model is especially applicable for online retailers, who sell a much more extensive line of products than do traditional stores. For online retailers, the model can be used as part of a consumer recommendation system, to help consumers identify what other products they might be interested in, based on their past purchasing histories. Collaborative filtering models based on RBM have already been used for online movie recommendations, and have proven to be very effective (Salakhutdinov et al. 2007).

For this paper, we conducted analysis at the product category level. The proposed model may also be used at the brand level directly, to capture brand associations in consumer purchasing patterns. Analysis at that level would enable managers to uncover associations among brands across categories, and to design cross-brand promotions based on the captured associations. Furthermore, at the brand level, price and promotion information can be incorporated in the model. Managers then can examine how price and promotions affect purchasing decisions across brands, and can determine optimal marketing mix strategies (Hansen et al. 2006). Our model has the power and efficiency to conduct the analysis necessary in very reasonable time, even for thousands of brands.

1.6 CONCLUSION

Our proposed CRBM based model is a significant extension to the marketing literature on multi-category joint purchasing decision models, by expanding the number of product categories and incorporating multiple time periods. The model possesses all the benefits provided by traditional models, but with the huge advantage of efficient scale-up for big data analysis. Our approach is comprehensive, in that it allows for the inclusion of intertemporal multi-category effects via a clearly specified model, without having to impose a priori constraints to ensure model tractability. We showed that, even with small data sets with limited product categories, our model fits the data at least as well, and predicts at least as accurately, as do traditional models, which do not scale up efficiently.

We demonstrated our model's ability to analyze high-dimensional consumer purchasing data, and to capture purchasing patterns helpful for marketing decisions, such as in the area of personalized coupon promotions. The intertemporal, cross-category associations discovered by the model can help determine what product coupons to send to a consumer and when to send them. We show that the intertemporal effects may be complex, and may span many product categories. The effects are dynamic and asymmetric, and challenge the conventional notion of complements or substitutes.

Our research may be extended in multiple ways. We conducted our analysis at the product category level. Our model can be used for brand level intertemporal cross-effects analysis, with additional input variables, such as price and promotion. The inclusion of marketing mix variables would lead to more useful and actionable insights.

We modeled consumer purchasing decisions as binary outcomes, and did not explore the associations of purchasing decisions in other forms, such as the amount spent in each product category or the quantity purchased. The amount spent may be modeled as a continuous variable, and an extension of RBM may be used to model the associations (Hinton 2006). The Poisson RBM may be used to model the quantities of products purchased by the consumer. It would be interesting to see if using spending or purchase quantity as the dependent variable (especially at the brand level) leads to new consumer insights.

Consumer heterogeneity can be incorporated either by adding individual-specific intercept terms or by integrating a hierarchical Bayesian model with CRBM to form a "Hierarchical Deep Model" (HD), as discussed in Section 2.2.   While we did not see any improvement in the models' predictive performance in our application, explicit incorporation of heterogeneity in our model is definitely an option.

Finally, from a consumer behavior perspective, there is a clear opportunity for experiment-based research to better understand (and otherwise verify) the complex effects suggested by the results of our model.

2.0 CHAPTER TWO. THE MANY FORMS OF GROUP BUYING: A NORMATIVE STUDY OF ALTERNATIVE MECHANISMS

2.1 INTRODUCTION

Group buying is an emerging approach to increase sales. Jing and Xie (2011) define group buying as a strategy in which a "seller offers discounted group rates to encourage individual consumers to purchase through buying groups." In just four years since its launch, Groupon – a flag bearer of group buying – has grown to over 260 million subscribers in over 45 countries, and with gross billing of $1.55 billion in their most recent quarter (Q1 of 2015, Groupon.com) for a 10% year-to-year increase. There were 311 online group buying sites in the USA in 2011 with another 176 sites in China (zixun.tuan800.com, June 2014). The growing popularity of group buying sites is in large measure due to the Internet's unmatched ability to connect consumers with each other and with the seller. Yet academic research on the topic is sparse (see Anand and Aron 2003; Dholakia 2011; Jing and Xie 2011).

A core premise of our normative study is that group buying is not a single, monolithic strategy. Group buying sites employ diverse strategies, with distinct implications. Groupon, Nuomi (China) and DailyDeal (Germany) follow the *volume strategy* by which each announces a deeply discounted price for a product to its subscribers and does not require any peer mobilization on the subscriber's part. LivingSocial, Beibeidai (China), and MyDala (India) rely on a *referral reward strategy* under which a buyer initially pays an announced price but can get the good for free if she convinces three of her peers to make the purchase. The now defunct Mercata, Tuanche (China), and Groffr (India) have relied on a true *collective buying strategy* wherein buyers get bigger discounts as their group size increases. Further, each of these strategies has subtypes that differ in the extent of market penetration and the underlying processes.

Our normative study examines the product-market conditions under which each of the above group buying mechanisms is optimal. We focus on a local monopolist seller and address the following research questions:

1. Under what conditions of consumer heterogeneity in product knowledge and intrinsic valuations should the seller choose group buying over the margin strategy?
2. When group buying is preferred, which specific strategy (and variant or subtype within the strategy) is optimal?

Our consideration of a broader strategy space and a more general form of consumer heterogeneity relative to extant studies such as Anand and Aron (2003) and Jing and Xie (2013) helps in offering fresh insights on optimal strategies and underlying processes. The baseline mechanism in our model is the *margin strategy* under which the seller would rather sell fewer units of the good at a higher unit margin. Our key conclusion is that no one strategy is best across product market conditions, and therefore the choice of strategy must be judiciously made on the basis of the prevailing product-market conditions.

Overall, our study helps to point out that while all group buying strategies seek greater market penetration (relative to the margin strategy), specific conditions favor a limited vs. aggressive pursuit of market penetration even within group buying strategies (e.g., 2-, 3- or 4-segment collective buying can each be optimal albeit under different product market conditions). We are also able to show that group buying is not about relying just on well informed, high valuation consumers to bring in their less informed, lower valued peers. Even the low valuation consumers who are well informed can be a channel for drawing other consumers, if the resulting drop in price is incentive enough for the lower valuation consumers to make the purchase.

We find that collective buying and referral reward strategies are actually more profitable than the alternatives over a sizable domain. Along with the two sources of consumer heterogeneity, other factors such as the efficiency of information sharing among consumers matter in determining the optimal online selling strategy. While the regions of optimality are complicated, we are able to derive some general results that are insightful and managerially relevant especially in today's online environment.

We start with a brief literature review that helps position our study. Next, we describe the model framework, and then present our results based on comparing the profitability of different strategies under various product-market conditions. In the final section, we discuss the theoretical contributions and managerial implications of our study, and provide future research directions.

2.2 LITERATURE AND POSITIONING

Although group buying has long been practiced and examined in B2B contexts, the online B2C group buying phenomenon has caught the attention of both business practitioners and academic researchers more recently (Chen et al. 2007). While there has been significant research on online shopping, there has been little work specifically on online selling strategies. The group buying mechanism works differently from traditional selling strategies, and its effectiveness has been examined mostly through analytical models (Chen et al. 2010; Chen and Li 2013; Hu et al. 2013; Kauffman and Wang 2001). Some researchers have considered group buying as a fixed price group discount without information sharing (Marvel and Yang 2008; Chen and Roma 2011; Surasvadi et al. 2014), while others have emphasized the effect of social interaction in group buying (Jing and Xie 2011; Ye et al. 2012). Without consumer interaction, strategic consumers either buy the product immediately or just wait until the group buying threshold is met (Liang et al. 2013). In this setting, displaying the number of consumers signed up for the deal would probably increase the profitability of group buying (Hu et al 2013).

Anand and Aron (2003) compare pricing under group buying with posted prices under different situations, and find that it is not always optimal for sellers to use group buying. The specific model they consider is the Mercata-type collective buying. They assume implicitly that all consumers have perfect information on the seller's product so that there is no information gap among consumers. Their analysis considers the effect of demand uncertainty on the outcome of group buying.

In contrast, Jing and Xie (2011) focus on the information gap among segments. Specific forms of group buying can help to spread product information via word-of-mouth and thereby change consumers' evaluation of the product. They classify consumers into 2 groups, one perfectly informed about the product and the other not. Their analytical model does not consider demand uncertainty or heterogeneity in intrinsic valuation among consumers.

Recent empirical research has been based on analysis of data collected from online group buying deals. Consistent with previous research, empirical findings show that consumer interactions take place during group buying deals, and product information is spread among consumers as a result (Byers et al. 2012). Aside from heterogeneity in product valuation and product information, there are other factors that would affect the outcome of a group buying deal, such as the threshold effect in terms of the surge of sign-ups when the minimum number of consumers required for a group buying deal is reached (Wu et al. 2014; Zhou et al. 2013). The "two-sided platform" is also a distinctive aspect of online group buying and has garnered some research interest (Rochet and Tirole 2003; Kim et al. 2012).

Dholakia (2011) examines the real world impact of daily deals on profits for different product categories, and local merchants' attitude towards daily deals after a first try. His survey results suggest that online daily deals yield disappointing consumer return rates, thereby lowering merchants' willingness to run another daily deal in the future. More recently, Edelman, Jaffe, and Kominers (2011) analyze the profitability of using Groupon to sell experience goods based on the parsimonious framework of Bils (1989).

Our study complements the above stream of research in at least two ways:

- Blending the perspectives of Anand and Aron (2003) and Jing and Xie (2009), we consider a market characterized by heterogeneity in consumers' intrinsic valuation *and* product knowledge. Our four-segment conceptualization in which consumers are high or low dimension helps unearth process explanations previously unexplored.
- Together with the above, our consideration of not only the three principal group buying strategies but also their subtypes helps clarify the optimal strategy and its underlying mechanism in a manner that is new. For example, we are repeatedly able to show that (a)

a group buying strategy need not be about maximizing penetration; (b) a suboptimal subtype of a specific strategy may undermine the appeal of the general strategy under specific market conditions; (c) under a strategy such as referral reward, the initial buyer need not just be the prime consumer (with the highest valuation) but even a subprime consumer who sees an opportunity to gain a referral reward through social interaction, thereby making the purchase incentive compatible for herself.

2.3 MODEL

Our model builds on Jing and Xie's (2011) framework. Like them, we focus on a profit maximizing monopolist who has ample quantity of a product to offer. The product's marginal cost is set to zero without loss of generality. Consumers are surplus maximizers. They are heterogeneous in their level of information about the product (i.e., product knowledge). We simply consider two consumer types on this basis, well informed (*Hi-Inf*) and less uninformed (*Lo-Inf*). The difference in the information levels between the consumer types is the information gap I (see Alba and Hutchinson 1987).

Specific group buying strategies such as referral reward depend on the influencers – the *Hi-Inf* consumers – informing the *Lo-Inf* consumers to induce the latter to buy the product by increasing their product valuation. Informing a *Lo-Inf* consumer requires effort on the part of the *Hi-Inf* consumer, who must be incentivized to make the effort. The referral reward, for example, is such an incentive. The *Hi-Inf* consumer incurs a cost of effort β in providing information to a *Lo-Inf* consumer to narrow the information gap by one unit. One unit of information boosts the *Lo-Inf* consumer's valuation by α. Thus, bridging the entire information gap I increases *Lo-Inf* consumer valuation by αI, while the cost incurred by the *Hi-Inf* consumer in providing the information is βI. The parameters α, β and I are likely to be product-market specific. For ease of reference, we list all the notations used in our model in Table 1.

Table 4 List of Notations

Notation	Description
Consumer/Segment Types	
Hi-Inf	Well informed consumer
Lo-Inf	Less informed consumer
Hi-Val	High intrinsic valuation consumer
Lo-Val	Low intrinsic valuation consumer
Segment 1	*Hi-Val* / *Hi-Inf* consumers
Segment 2	*Hi-Val* / *Lo-Inf* consumers
Segment 3	*Lo-Val* / *Hi-Inf* consumers
Segment 4	*Lo-Val* / *Lo-Inf* consumers
Parameters	
V	Product valuation of *Lo-Val* / *Lo-Inf* consumer (*Segment* 4)
I	Information gap between *Hi-Inf* and *Lo-Inf* consumers
ϕ	Intrinsic product valuation gap between Hi-Val and Lo-Val consumers
α	Impact of one unit of information on a consumer's product valuation
β	Cost incurred by *Hi-Inf* consumer to narrow the information gap with the *Lo-Inf* consumer by one unit
δ	Discount factor
Decision Variables	
P	Price
R	Referral Reward
ΔI	Amount of information conveyed by *Hi-Inf* consumers in educating *Low-Inf* consumers ($\Delta I \leq I$)

One of the ways in which we extend Jing and Xie's framework is by incorporating heterogeneity in intrinsic valuations among consumers, capturing the important idea that consumers with the same level of product information about the product are likely to have different valuations of the product. Consumers have either high or low intrinsic product valuation, with ϕ denoting the difference in intrinsic valuation between the two consumer types, denoted by *Hi-Val* and *Lo-Val*.

The resulting four segment structure is shown schematically in Figure 1, with the initial valuations before the seller introduces incentives that might bridge some of the gaps. As in Jing and Xie, we assume that each segment is of size 1. Relaxing this assumption is straightforward, but the greater complexity does not add any meaningful insights.

Figure 7 Consumer Segments and Their Initial Valuations

	Well Informed (*Hi-Inf*)	Less Informed (*Lo-Inf*)
High Intrinsic Valuation (*Hi-Val*)	*Segment* 1: $V + \phi + \alpha I$	*Segment* 2: $V + \phi$
Low Intrinsic Valuation (*Lo-Val*)	*Segment* 3: $V + \alpha I$	*Segment* 4: V

This four-segment structure provides conceptual and strategic elements of our framework vis-à-vis Jing and Xie. *First*, our model captures the point that the information dissemination aspect of group buying need not originate only from the well informed, high value consumer in Segment 1. Segment 3, informed but with low product valuation, can also be an effective source of information. In fact, informing Segment 2 and bringing it on board might actually make the product affordable to Segment 3. (An illustration of this is LivingSocial's referral reward strategy noted earlier. The information disseminator can get the product for free in that context.) *Second*, the above heterogeneity structure opens up a broader strategy space, notably in the subtypes of each strategy, as will be evidenced by the following discussion and subsequent analysis.

The seller's strategy space consists of the following:

- *Margin strategy*. Under this baseline strategy, the seller prices the product to clear only the one or two highest value segments and does not offer any incentive to promote group buying. At a minimum, Segment 1, with the highest valuation, will be cleared. The choice between Segments 2 and 3 depends on their respective valuations. Accordingly, 1-segment and 2-segment margin strategies must be distinguished.

- *Volume strategy*. Analogous to Groupon's model, the seller sets the price low enough to clear most or all of the segments. There is no monetary incentive to the fully informed segments to inform the other segments. Rather the lower price is sufficient to make the purchase incentive compatible to Segments 2 and/or 4. We consider 3-segment and 4-segment volume strategies. In the former case, Segment 4 forgoes purchase.

- *Referral reward strategy*. Akin to LivingSocial, the seller offers a discount to the *Hi-Inf* consumers (Segments 1 and 3) as an incentive for informing Segments 2 and/or 4 and thereby improving the latter's valuation. The reward must be sufficient for Segments 1 and/or 3 to inform while also making the purchase themselves. Of course, Segments 2 and 4 do not receive the reward. The segment(s) that buy after being informed do so at a later point in time, and the reward for a referral is also received at this later point, and hence discounted (by a factor δ) to reflect the time value of money. Of course, the time difference may (and typically will) be small, in which case $\delta \approx 1$.

- *Collective buying strategy*. Akin to Mercata's business model, the seller offers a discount tied to the number of consumers who band together to purchase the product. Similar to the referral reward strategy, collective buying is time consuming as consumers have to seek out other prospects. Unlike referral reward, all consumers (or segments) who buy collectively will receive the product at the same discounted price.

The strategies, subtypes and notation used henceforth are provided in Table 2. The profit-maximizing seller seeks to maximize profits under each strategy. The seller's objective is to choose the profit maximizing strategy and subtype for a given set of product-market conditions.

Table 5		Strategies, Subtypes, and Notation

Strategy	Subtype and Notation
Margin Strategy (MS)	• 1-segment: MS (1) • 2-segment: MS (1, 2) or MS (1, 3)
Volume Strategy (VS)	• 3-segment: VS (1, 2, 3) • 4-segment: VS (1, 2, 3, 4)
Referral Reward Strategy (RRS)	• 2-segment: RRS (1, 2) • 3-segment: RRS (1, 2, 3) • 4-segment: RRS (1, 2, 3, 4)
Collective Buying Strategy (CBS)	• 2-segment: CBS (1, 2) or (1, 3) • 3-segment: CBS (1, 2, 3) • 4-segment: CBS (1, 2, 3, 4)

2.4 ANALYSIS AND RESULTS

Our analysis proceeds in two stages. First, in §4.1, we derive the optimal (or profit maximizing) price and/or incentive under each strategy and the corresponding profit. Under any given strategy and subtype, the seller and consumers behave rationally to ensure that the targeted number of consumers make the purchase; indeed, doing so is Pareto optimal. The analysis in §4.1 provides the basis for the key results of our research, presented in §4.2, identifying which strategy (margin, volume, referral reward or collective buying, each with their subtypes) is optimal under a range of product-market conditions.

As in Jing and Xie, we set each segment size to 1. Since ϕ represents the intrinsic valuation gap between two equally informed consumers (between Segments 1 and 3, or 2 and 4), and αI represents the information-related valuation gap (between Segments 1 and 2, or 3 and 4), the magnitudes of ϕ and αI determine the attractiveness of the segments and the effectiveness of alternative strategies. The seller is aware that consumers' choices under a given strategy – and subtype –fulfill individual rationality and incentive compatibility constraints.

2.4.1 Optimal Prices and Profits

In this subsection, we present the optimal prices, profit, and (in the referral reward case) optimal reward for each strategy, considering all feasible subtypes under each strategy, in the form of Lemmas. These lemmas then inform our analysis leading to the central results of our research.

Margin and Volume Strategies. When $\phi <$ (or $>$) $I\alpha$, the valuation difference between segments 1 and 2 due to the information gap is greater (or less) than that between segments 1 and 3 due to intrinsic valuation heterogeneity. Given the profit-maximizing monopolist facing rational consumers, the analysis yields the following optimal prices and profits under each subtype of the margin and volume strategies:

Lemma 1. *The optimal prices and corresponding profits under margin (MS) and volume (VS) strategies are as follows:*

Strategy Subtype	Optimal Price	Profit
MS (1)	$V+\phi+\alpha I$	$V+\phi+\alpha I$
MS (1, 2)	$V+\phi$	$2(V+\phi)$
MS (1, 3)	$V+\alpha I$	$2(V+\alpha I)$
VS (1, 2, 3) ($\phi<\alpha I$)	$V+\phi$	$3(V+\phi)$
VS (1, 2, 3) ($\phi>\alpha I$)	$V+\alpha I$	$3(V+\alpha I)$
VS (1, 2, 3, 4)	V	$4V$

Referral Reward. While the product is priced at P, the *Hi-Inf* consumer gets a reward R if she is successful in informing the *Lo-Inf* consumer so that the latter's valuation of the product increases to make purchase incentive compatible at full price P. As the *Hi-Inf* consumer is required to make the purchase up front to avail of the reward, she effectively pays $P - \delta R$ for the product, since the reward is obtained at a later period and hence discounted by δ. As shown in Table 2, there are three subtypes of the referral reward strategy, involving Segments 1 and 2 only, Segments 1, 2, and 3, and all four segments. The derivations of the optimal prices, rewards, and firm profits for the three cases are provided in Appendix A.1.

Lemma 2. *The optimal prices and corresponding profits under the referral reward strategy (RRS) are as follows:*

Strategy Subtype	Optimal Price	Optimal Reward	Profit
RRS (1, 2)	$V+\phi+\alpha I$	βI	$(1+\delta)(V+\phi+\alpha I)-\delta\beta I$
RRS (1, 2, 3)	$V+\alpha I$	$\frac{\beta(\alpha I-\phi)}{\alpha}$	$(2+\delta)(V+\alpha I)-\frac{\delta\beta(\alpha I-\phi)}{\alpha}$
RRS (1, 2, 3, 4) ($\phi < \alpha I$, Option I)	$V+\phi$	$\frac{\beta\phi}{\alpha}$	$(3+\delta)(V+\phi)-\frac{\delta\beta\phi}{\alpha}$
RRS (1, 2, 3, 4) ($\phi < \alpha I$, Option II)	$V+\alpha I$	βI	$(2+2\delta)(V+\alpha I)-2\delta\beta I$
RRS (1, 2, 3, 4) ($\phi > \alpha I$)	$V+\alpha I$	βI	$(3+\delta)(V+\alpha I)-\delta\beta I$

Collective Buying. Under this strategy, *Hi-Inf* consumers must first educate the *Lo-Inf* consumers to increase the latter's valuation, and then join with them to drive down the seller's price (assuming that reducing the price is in the seller's interest). The lower price provides the incentive for Segments 1 and/or 3 to undertake the effort to inform Segments 2 and/or 4. The consumers in Segment 1 may not expend the effort to *fully* inform the consumers in Segment 2 (and/or 4), but only to the extent that is incentive compatible for the concerned parties.

As the effort to inform the *Lo-Inf* consumers takes time, profits from collective buying are slightly discounted by δ. Group buying sites typically work with shopping windows of only days or weeks, and hence δ should usually be close to 1. Table 1 indicates the three subtypes of the collective buying strategy, involving Segments 1and 2 (or 3), Segments 1, 2, and 3, and all four segments. The derivations of the optimal prices, rewards, and firm profits for the three cases are provided in Appendix A.2.

Lemma 3. *The optimal prices and corresponding profits under the collective buying strategy (CBS) are as follows:*

Strategy Subtype*	Optimal Price	Profit
CBS (1, 2)	$V+\phi+\dfrac{\alpha^2 I}{\alpha+\beta}$	$2\delta(V+\phi+\dfrac{\alpha^2 I}{\alpha+\beta})$
CBS (1, 2, 3) ($\phi<\alpha I$)	$V+\alpha I+\dfrac{\beta(\phi-\alpha I)}{2\alpha+\beta}$	$3\delta(V+\alpha I+\dfrac{\beta(\phi-\alpha I)}{2\alpha+\beta})$
CBS (1, 2, 3, 4) ($\phi<\alpha I$)	$V+\dfrac{2\alpha^2 I+\beta\phi}{2\alpha+2\beta}$	$4\delta(V+\dfrac{2\alpha^2 I+\beta\phi}{2\alpha+2\beta})$
CBS (1, 2, 3, 4) ($\phi>\alpha I$)	$V+\dfrac{2\alpha^2 I}{2\alpha+\beta}$	$4\delta(V+\dfrac{2\alpha^2 I}{2\alpha+\beta})$

* CBS (1, 2) is not feasible for $\phi < I\alpha$, and CBS (1, 2, 3) is not feasible for $\phi > I\alpha$.

Some observations on the optimal prices and profits for different strategies are in order, especially with regard to the impact of α, β, ϕ, and I. Traditional selling strategies (margin and volume) do not involve consumer information sharing, and optimal prices and profits are therefore independent of the consumer's cost of information-sharing β. For group buying strategies (RRS and CBS), the optimality is affected not only by the heterogeneity in product information determined by α and I, but also by the heterogeneity in intrinsic valuation ϕ. The relative magnitude of these two sources of heterogeneity in valuations has significant implications for optimal prices and profits under RRS or CBS strategies. Since the magnitudes of α, I, β and ϕ determine the profit levels, it is clear that they will determine which strategy (and subtype within a strategy) is optimal, which is the focus in the next subsection.

Comparative statics. We briefly examine how the prices and profits for different strategies are affected by α, I, β and ϕ. The comparative statics (see Appendix A.3) indicate that the unlike the traditional margin or volume strategies, the parameters have a nonlinear impact on optimal prices and profits. There is also a distinct difference between CBS and RRS. CBS is a more nuanced group buying strategy than RRS: the optimal price for CBS is affected by all parameters, while for RRS the optimal price is always equal to the valuation of the referrer. Therefore, the optimal price under RRS is independent of β, and generally (but not always) changes positively with α or ϕ. Since reward under RRS is higher when β is larger, β negatively impacts the profit.

For CBS, the optimal price will always be non-decreasing in ϕ, but does not does not change monotonically in α, I, or β. These parameters interact with each other in a complicated manner in determining the optimal price, and consequently their impact on prices and profits can be either negative or positive, depending on their levels.

2.4.2 Relative Attractiveness of Alternative Strategies

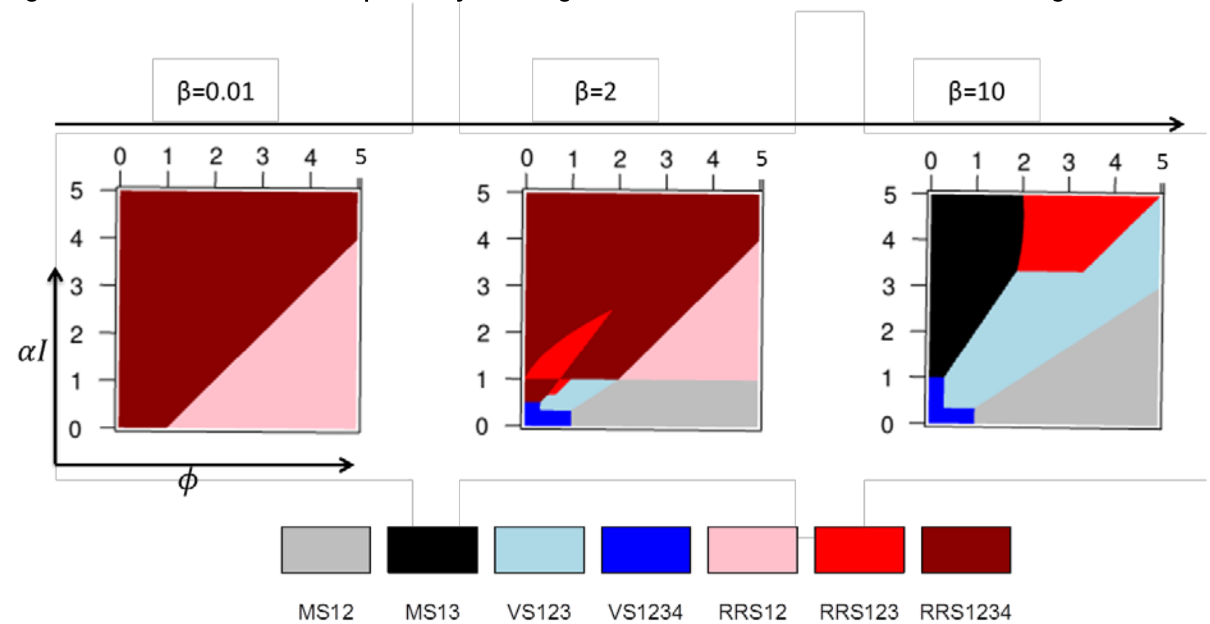
Our focus here is on identifying the optimal (i.e., profit maximizing) strategies under a range of conditions. Since each strategy has subtypes and there are several parameters that interact, we present our results in stages for more effective exposition. We compare (a) RRS alone with MS and VS in §4.2.1, (b) CBS alone with MS and VS in §4.2.2, and finally (c) both RRS and CBS with MS and VS (the full strategy space) in §4.2.3. This sequential presentation more effectively clarifies the domains of attractiveness of alternative strategies and the related intuition. To simplify the exposition, but without loss of generality, we set V to 1 which means the lowest product valuation among the consumers is 1. The values of other parameters can be scaled proportionately. In this sub-section, we set the discount factor δ to 1 which is a reasonable approximation, as discussed earlier. Discussion of the impact of a smaller δ is deferred to the following sub-section.

We present the domains of optimality in the form of phase diagrams. Each figure has panels for high, medium and low levels of β (cost to the *Hi-Inf* consumer of bridging the information gap by one unit). Each panel indicates the optimal strategy as a phase diagram in the two-dimensional space defined by ϕ (the measure of the heterogeneity in intrinsic valuation) on the horizontal axis and αI (the increase in reservation price of the less informed consumers when the information gap is fully bridged) on the vertical axis. The analytical demarcation of the boundaries between adjacent regions of optimality is discussed in Appendix B.

2.4.3 Relative Attractiveness of Margin, Volume and Referral Reward Strategies

Figure 2 provides the panel of phase diagrams (in color) showing specific thresholds for the domains of optimality in this case.

Figure 8 Domains of Optimality of Margin, Volume and Referral Reward Strategies



The implications are summarized below.

Result 1. *Among MS, VS and RRS: the domain of optimality of RRS is larger when β is lower. For higher β, VS is optimal when $\alpha I \approx \phi$; MS is optimal otherwise. Specifically:*

(a) When β is low, RRS dominates VS and MS. Within RRS:

(i) RRS (1,2,3,4) is optimal when $\alpha I > (\phi - 1)$;

(ii) RRS (1,2) is optimal otherwise.

(b) When β is moderate,

(i) RRS is optimal when $\alpha I > 1$. Within RRS, RRS (1,2) is optimal for $\alpha I < (\phi - 1)$; RRS (1,2,3,4) is optimal otherwise.

(ii) When $\alpha I < 1$, MS is optimal for $\left[\phi > 1 \text{ and } \phi > \dfrac{1+3\alpha I}{2}\right]$ *and VS is optimal otherwise.*

Within VS, VS(1,2,3,4) is optimal as αI and ϕ approach zero; VS (1,2,3) is optimal otherwise. Within MS, MS (1,2) is optimal.

(c) When β is large, the domain of optimality of RRS shrinks. VS is optimal for $\dfrac{2\phi-1}{3} < \alpha I < \dfrac{1+3\phi}{2}$. *MS is optimal otherwise. Further,*

(i) *Within VS: VS (1,2,3) is de facto optimal; VS (1,2,3,4) is optimal only when* αI *and* ϕ *approach zero.*

(ii) *Within MS: MS (1,2) is optimal when* $\alpha I < \frac{2\phi - 1}{3}$*; MS (1,3) is optimal when* $\alpha I > \frac{1+3\phi}{2}$.

RRS is a price discrimination strategy; MS and VS are not. Specifically, under RRS, the seller offers a discount for the *Hi-Inf* Segment 1 (and/or 3) in exchange for bridging the information gap with the *Lo-Inf* Segment 2 (and/or 4). Segments 2 and 4, when they make the purchase, pay full price.

Low β means information sharing is less burdensome for the *Hi-Inf* segments. Therefore, in exchange for a relatively small reward, the seller can entice one or both of these segments to inform the *Lo-Inf* segment(s). Segments 1 and 2 will always buy the product under RRS. The inclusion of Segments 3 and 4 depends further on the intrinsic valuation gap ϕ relative to the information-related valuation gap αI. When the latter is large enough relative to the former, it becomes profitable to reward information sharing with even Segment 4, making RRS (1,2,3,4) optimal.

When β is large, the seller would have to offer a significantly higher reward to the *Hi-Inf* segment(s) to motivate them to educate the *Lo-Inf* segment(s), thus reducing the appeal of RRS relative to MS or VS. Between the latter two strategies, first consider the case of $\alpha I \approx \phi$, which means that Segments 2 (*Hi-Val* / *Lo-Inf*) and 3 (*Lo-Val* / *Hi-Inf*) are roughly equally appealing to the seller who is better off not leaving these two segments untapped. Thus, VS (1,2,3) is optimal in this region. Now if αI and ϕ are both small, Segment 4 ((*Lo-Val* / *Lo-Inf*)) is no longer much lower than the other segments in its willingness to pay, and thus VS (1,2,3,4) becomes optimal for the seller. On the other hand, MS is optimal when αI and ϕ are significantly different from each other. MS (1,2) is optimal when $\alpha I << \phi$ because of segment 3's low valuation relative to Segment 2, and conversely MS (1,3) is optimal when $\alpha I >> \phi$.

Note, however, that even under high β, there is a region where RRS (1,2,3) is optimal – when αI and ϕ are both sufficiently large to make it worthwhile for the seller to incentivize the *Hi-Inf* segments to close the information gap for Segment 2. Given the large intrinsic valuation gap between Segments 2 and 4, it is not profitable for the seller to drop prices to include the latter segment. The relative magnitudes of αI and ϕ are important in defining the region of optimality for RRS (1,2,3). If ϕ is large relative to αI, then a simple volume strategy, VS (1,2,3) is a more profitable approach that includes both Segments 2 and 3 among buyers without having to reward the informed segments to share information with Segment 2.

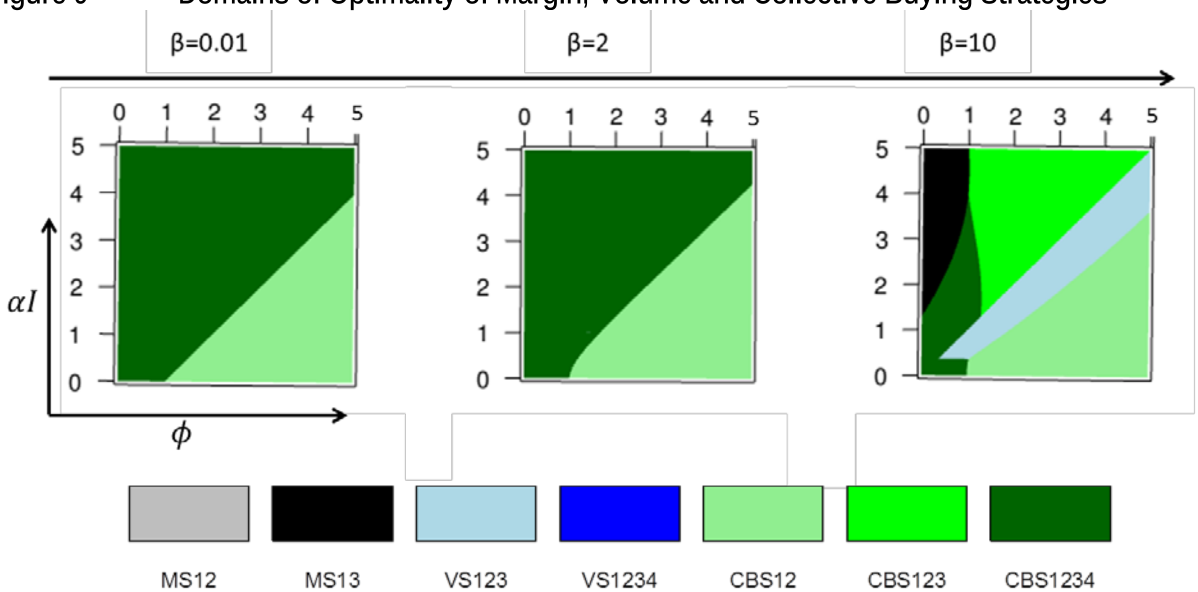
When β is moderate (Result 1(b); the middle panel in Figure 2), RRS is appealing but only when the information-related valuation gap is above some threshold (specifically, when $\alpha I > 1$). Moderate β means that referral is less efficient than under low β. The intuition is that a sufficiently high αI provides the opportunity to enhance the valuations of Segment 2 (and possibly 4). β is still moderate enough to facilitate the referral efforts. As discussed earlier in the case of low β, whether RRS (1,2) or RRS (1,2,3,4) is optimal depends on ϕ relative to αI. The latter is optimal when αI is relatively large, and vice versa. (Note that there is an RRS (1,2,3) "sliver" within the larger RRS (1,2,3,4) region, where the combination of the product-market parameter values makes it suboptimal to sell to Segment 4. The boundary conditions in this case are complex and not particularly insightful, and hence not discussed here.)

When the information-related valuation gap does not exceed the threshold, rewarding informed consumers is no longer profitable, and a "traditional" volume or margin strategy becomes optimal. Depending on the relative values αI and ϕ, MS (1,2), VS (1,2,3) or VS (1,2,3,4) is optimal – as discussed above, for the high β case.

4.2.2. *Relative Attractiveness of Margin, Volume and Collective Buying Strategies*

Figure 3 provides the panel of phase diagrams (in color) showing specific thresholds for the domains of optimality in this case.

Figure 9 Domains of Optimality of Margin, Volume and Collective Buying Strategies



The implications are summarized below.

Result 2. *Among MS, VS and CBS, the domain of optimality of CBS is larger when β is lower. For high β, VS is optimal when $\alpha I \approx \phi$; MS is optimal under low ϕ and high αI; CBS is optimal elsewhere. Specifically:*

(a) When β is low or moderate, CBS dominates VS and MS. Within CBS:

(i) CBS (1,2,3,4) optimal when $\phi < \left[1+\dfrac{\alpha^2 I(2\alpha+3\beta)}{(2\alpha+\beta)(\alpha+\beta)}\right]$;

(ii) CBS (1,2) is optimal otherwise.

(b) When β is large:

(i) MS (1,3) is optimal for high αI and low ϕ;

(ii) VS (1,2,3) is optimal when $\alpha I \approx \phi$. Specifically, VS (1,2,3) is optimal when

$$\left[\frac{\alpha+\beta+4\alpha^2 I}{3\alpha+\beta}\right] < \phi < \left[\frac{(\alpha+3\beta)\alpha I}{2(\alpha+\beta)} + \frac{1}{2}\right]$$

(iii) CBS is optimal otherwise. CBS (1,2) is optimal for $\phi >> \alpha I$; CBS (1,2,3,4) is optimal for low αI and low ϕ; and CBS (1,2,3) is optimal elsewhere.

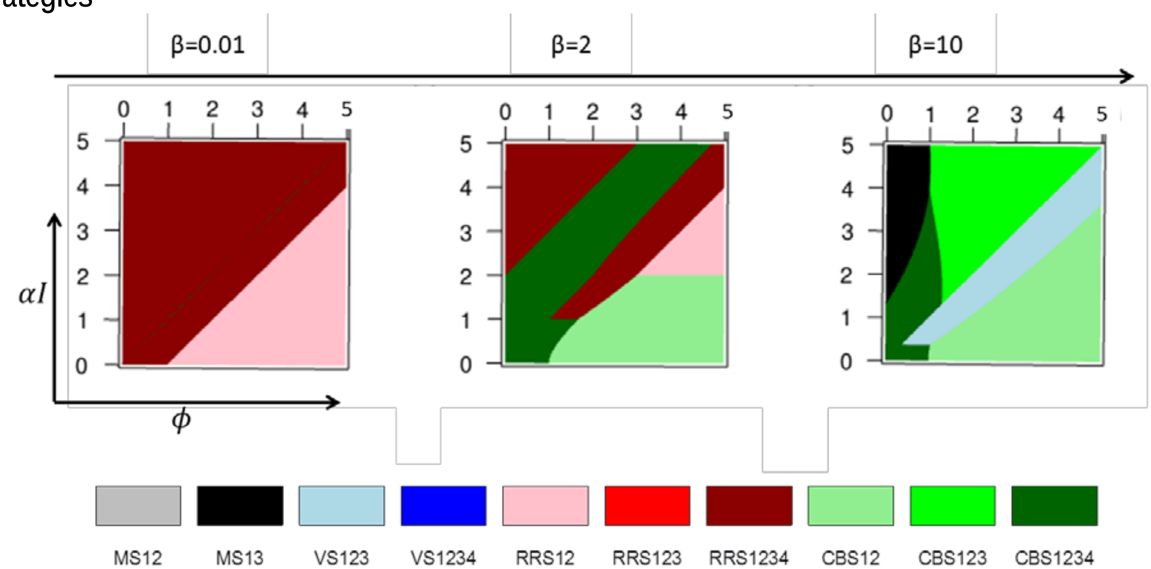
Comparing Figures 2 and 3, it is apparent that the regions of optimality for RRS (in Figure 2) and CBS (in Figures 3) versus VS and MS are qualitatively similar, although CBS has a larger domain of optimality than RRS does. The 2-, 3-, and 4-segment CBS subtypes are optimal in regions of the phase diagrams that are comparable to those for RRS, illustrating that in both group buying strategies, each of the subtypes can be optimal under appropriate product-market conditions defined by the model parameters. The rationale for the specific regions of optimality as shown in Figure 3 closely parallels that for the RRS case discussed §4.2.1 above, and therefore not repeated here.

While the intuition for the optimality of CBS is analogous to that for RRS, the two strategies work differently, which informs their relative attractiveness under different conditions when both options are available. The key insights in this regard are discussed next, as we expand the strategy space to include both RRS and CBS along with VS and MS.

2.4.4 Relative Attractiveness of Margin, Volume, Referral Rewards and Collective Buying Strategies

We now expand the strategy space to examine the domains of optimality for all four strategies under evaluation, with Figure 4 providing the panel of phase diagrams (in color).

Figure 10 Domains of Optimality of Margin, Volume, Referral Reward and Collective Buying Strategies



Result 3. *Among MS, VS, RRS and CBS: RRS is optimal for low β. As β increases, the domain of RRS shrinks and that of CBS expands. For high β, RRS is suboptimal; VS is optimal when $\alpha I \approx \phi$; MS is optimal under low ϕ and high αI; and CBS is optimal elsewhere. Specifically:*

(a) *When β is low, RRS dominates. RRS (1,2,3,4) is optimal when $\phi < (\alpha I + 1)$; RRS (1,2) is optimal otherwise.*

(b) *When β is moderate, the domain of optimality of RRS in (a) shrinks as follows:*

- (i) *RRS (1,2,3,4) remains optimal when $(\alpha - \beta)I < \phi < \alpha I$. Otherwise CBS (1,2,3,4) replaces RRS (1,2,3,4) as the optimal strategy.*
- (ii) *RRS (1,2) remains optimal when $\beta < \alpha$. Otherwise CBS (1,2) is optimal.*

(c) When β is large:

(i) MS (1,3) is optimal for high αI and low ϕ;

(ii) VS (1,2,3) is optimal for $\alpha I \approx \phi$;

(iii) CBS is optimal elsewhere. Within CBS, CBS (1,2) is optimal for $\phi >> \alpha I$; CBS (1,2,3,4) is optimal for low αI and low ϕ; CBS (1,2,3) is optimal elsewhere.

RRS is optimal when the cost of information sharing (β) is low, but as β increases, CBS takes over from RRS as the dominant strategy; at sufficiently high β there are regions where VS and MS are optimal along with CBS.

Under RRS, the *Hi-Inf* segment is rewarded more heavily for making the effort to educate *Lo-Inf* consumers to the extent needed (and thereby maximally raise the latter's valuation). On the other hand, under CBS, the price discount ("reward") for a *Hi-Inf* consumer decreases as the *Lo-Inf* segments are better educated, because this raises their product valuations which implies higher collective buy prices. *Hi-Inf* consumers in CBS have the incentive to educate their *Lo-Inf* peers as in RRS but do not have to expend as much effort because the product is discounted for *all* consumers based on the volume realized. Further, from the seller's perspective, the price at which *Lo-Inf* segments can be cleared under CBS is higher compared to the volume strategy because information sharing raises the valuation of the *Lo-Inf* consumers. Thus, CBS emerges as an effective compromise between RRS and volume strategies as β increases and RRS is no longer dominant owing to the greater cost of information sharing. (Note that in CBS, the seller chooses a price schedule that sets the price at a point between the valuations of the *Hi-Inf* and *Lo-Inf* segments.)

Consider Result 3(a) and the left panel of Figure 4. Given the low β, the *Hi-Inf* segments 1 (and 3) can educate Segments 2 (and 4) without much effort. Under RRS, a modest reward from the seller to the informed segments is sufficient to bridge the entire information gap, thus maximally raising valuations for segments 2 and 4. By contrast, with CBS, the *Hi-Inf* segments have the incentive to educate the *Lo-Inf* segments only to the extent that raises their valuations to make the purchase. As a result, the market's total valuation is less under CBS than under RRS, thereby making RRS more profitable for the seller. RRS, as a *price discrimination strategy*, helps the seller by getting the full-information price out of Segment 2 (and possibly 4) while offering a targeted discount to segment 1 (and possibly 3). The regions of optimality for RRS (1,2) versus RRS (1,2,3,4) and the underlying logic are exactly as discussed in §4.2.1, and therefore not repeated here.

Conversely, when β is high (Result 3(c) and the right panel of Figure 4), RRS is dominated by other strategies. Indeed, with RRS out of the picture, the phase diagram in this case is the same as the corresponding panel in Figure 3 in §4.2.2, and our earlier discussion explaining the results applies.

The interesting case is that of moderate β (the middle panel in Figure 4 and Result 3(b)) because it involves both RRS and CBS in their 2- and 4-segment subtypes. In effect, we need to understand why, as β increases from low to moderate, a part of RRS (1,2) is taken over by CBS (1,2) and a part of RRS (1,2,3,4) is taken over by CBS (1,2,3,4).

The boundaries between RRS (1,2,3,4) and CBS (1,2,3,4) and between RRS (1,2) and CBS (1,2) are formally stated in Result 3(b). CBS (1,2,3,4) appears as a diagonal strip within the RRS (1,2,3,4) region, defined by the condition stated in Result 3 (b) (i). In effect, there is a tradeoff between intrinsic valuation gap ϕ and the information-based valuation gap αI. As long as the intrinsic valuation gap exists within the range defined by Result 3(b)(i), CBS (1,2,3,4) is optimal, otherwise, RRS (1,2,3,4) is optimal. In the case of RRS (1,2) and CBS (1,2), the boundary condition is more straightforward (Result 3(b)(ii)) – when α (the valuation impact of a unit of information) is less than β (the cost of sharing one unit of information), CBS (1,2) takes over from RRS (1,2). Given our discussion of how the two group buying strategies operate, these results make sense.

From Figures 2, 3, and 4, we observe that a three-segment group buying strategy (CBS (1,2,3) or RRS (1,2,3)) is optimal only when β is large. The region of optimality is clearly defined on one boundary, with VS (1,2,3) – αI must be greater than ϕ to make the group buying strategy more attractive than the corresponding volume strategy. Also, for CBS (1, 2, 3) to be optimal, β must be sufficiently large. Otherwise, CBS (1,2,3,4) is optimal if αI is low (in which case segment 4 is included) and MS (1,3) is optimal if αI is high (the *Lo-Inf* segments are unattractive given the large information gap, and too expensive to inform, given high β).

2.4.5 Impact of Discount Factor on Optimal Strategies

Our analysis in §4.2 has assumed that the discount factor $\delta = 1$, effectively focusing on situations where there is little delay in the transactions (which covers a vast majority of real world cases – for example, most online deals only stay valid for a few days). Nevertheless, we extend our discussion in this subsection to consider the implications of the discount factor δ less than 1, focusing on broad, qualitative insights. (We will readily share the mathematical details with interested readers.)

In our analysis, δ comes into play only in the case of those group buying strategies where there is consumer interaction involved. Thus, margin or volume strategies are immune to the changes in δ. These traditional strategies become more attractive relative to group buying strategies as δ becomes smaller (that is, as the time value of money becomes more significant).

The impact of the discount factor is different between RRS and CBS. In the case of CBS, the transaction will not occur until (and unless) the information sharing process has reached the point where the valuation of the *Lo-Inf* consumer(s) has been raised sufficiently to motivate purchase. A lengthy information sharing process results in a lower (discounted) overall consumer surplus, implying lower profit for the seller.

In the case of RRS, the *Hi-Inf* consumer would buy the product first and then attempt to persuade the *Lo-Inf* consumer(s) to buy in order to get the referral reward. Therefore, only the part of the profit generated from the *Lo-Inf* consumers' transactions would be delayed and negatively affected. Thus, in general, CBS is more sensitive to the changes in discount factor than RRS, and thereby is at a relative disadvantage as δ decreases.

Even if RRS is impacted less relative to CBS by the discount factor, both are adversely affected, unlike MS or VS. Thus, the traditional volume or margin strategies dominate group buying strategies when the discount factor becomes sufficiently large. Therefore, the group buying strategies – CBS in particular – are unlikely to be optimal if the information sharing process is expected to span a very long period of time. Fortunately, this rarely happens in real world applications.

2.5 CONCLUSION

Three factors have motivated this study. First, while group buying sites have mushroomed in the real world, especially in the USA and Asia, academic work on the topic is still in its early stages. Second, a detailed but not exhaustive search online reveals that a large majority of the online group buying sites are based on the daily deal model of Groupon. The relevance of alternative mechanisms of group buying seems to be grossly overlooked at least among practitioners. Third, analytically-oriented academic studies on the topic have taken a rather restricted view of consumer heterogeneity and have ignored variants of specific strategies. Doing so limits the scope of the guidelines from extant studies.

Recall that we consider four core strategies – margin (MS), volume (VS), referral reward (RRS), and collective buying (CBS) – and their subtypes. We discuss below the scope and potential of these strategies. We also note the limitations and future research directions.

2.5.1 Theoretical Implications

Whereas Anand and Aron (2003) consider consumer heterogeneity in intrinsic valuation and Jing and Xie (2009) consider heterogeneity in valuation tied to consumers' product knowledge, we model both forms of heterogeneity. This is significant because it maps on to a broader array of products and segments, opens up a larger strategy space, and clarifies how a group buying strategy such as collective buying is not always about increasing sales volume.

The following are among the study's salient theoretical implications:

- Collective buying and referral reward, the two group buying strategies in which the seller is really able to co-opt some segments of consumers to persuade others, turn out to be the most widely optimal. A subtype of one of these strategies prevails except when the informed consumers find the process of informing their less informed peers very costly.

- Between collective buying and referral reward, the latter benefits from the power of price discrimination: the seller is able to target the reward at the consumers who persuaded their peers to buy. However, the strength of collective buying lies in its ability to find a balance between the power of information sharing and the power of an outright price discount. This balance makes CBS appealing when the cost of information sharing is not low.
- Our modeling approach and results underscore that group buying should not be seen primarily as selling to the most number of consumers. RRS or CBS can be optimal even when the bulk of the market is deliberately not served. This can happen when heterogeneity is large on both product knowledge and intrinsic valuation. Group buying succeeds here by informing and tapping the less informed consumers and forgoing the informed consumers who have lower intrinsic valuation. We see this facet of group buying as pull marketing at its best.
- The conventional wisdom on RRS appears to be that the highest valued consumers would buy first and can be roped in to attract their peers. While we do not dispute this, we show that RRS can appeal to the well informed but lower valuation consumers who anticipate that the referral reward will make it incentive compatible for them to engage in information sharing.

2.5.2 Managerial Implications

The principal motivation for our study comes from the proliferation and diversity of group buying sites around the world. Our modeling approach and findings have the following implications for practitioners exploring the potential of group buying:

- No one form of group buying is optimal and the choice of strategy is sensitive to the product market conditions (as captured by our model parameters). This suggests that firms relying on one group buying vehicle – e.g., Groupon – across products and markets are probably not tapping into the full richness of group buying.

- Our model incorporates the information gap and the cost of bridging this gap. We would surmise that the gap is greater for higher ticket experience goods such as consumer electronics and vacation packages. The cost of bridging the gap would be greater when network or interaction among consumers within the social system is more difficult. Under a combination of these circumstances, we would recommend CBS as with Tuanche or Mercata (now defunct) over Living Social (RRS) or Groupon (VS).
- When the information gap is low and the cost of information sharing is moderate, a LivingSocial-type RRS is appropriate. Car rentals or manicure services arguably fall in this category. But even here, we would not favor LivingSocial's formulaic "get service free in exchange for three referrals." This is a riskier all-or-nothing type approach. A better alternative might be to get x% off for each referral, as this reduces the burden of finding three individuals in a not-so-well networked social environment.

2.5.3 Research Limitations and Future Research Directions

Despite our arguably more general approach to modeling the group buying problem, we have relied on simplifying assumptions for the sake of model tractability. These limit the study's scope while offering opportunities for future research.

We have assumed that while the valuation of the less informed segment(s) increases through peer referral, the informed segment's valuation does not increase. However, group buying sites for homes such as Zhongtuan in China and Groffr in India are hoping to tap the network effect when friends and family come together to buy a block of apartments. Modeling the presence of network externality should make the analysis more relevant to such settings.

The seller in our model is assumed to pick one of the strategies in the study to sell the product. It could be argued that the seller can resort to a combination of group buying strategies such as referral reward and margin strategies. It would be interesting to examine which combination of strategies might work best in tandem.

Our approach is entirely analytical. The rise of group buying sites provides the opportunity for rich datasets for similar products but across strategies. It would be interesting to take some of the propositions to data so as to validate or modify the current guidelines.

APPENDIX 1 GLOSSARY OF NOTATION

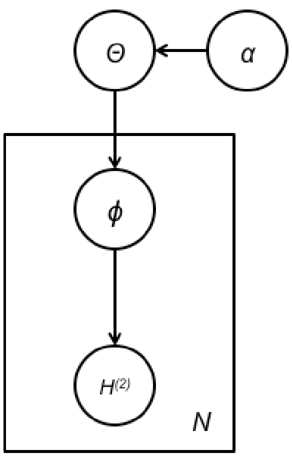
The symbols used to describe the model are summarized in the following table

$X_t = (x_{1t}, x_{2t}, \cdots, x_{(I-1)t}, x_{It})$	A I dimensional vector X_t representing a consumer's purchasing decision over I products at time t
$H_t = (h_{1t}, h_{2t}, \cdots, h_{(J-1)t}, h_{Jt})$	A J dimensional hidden vector H_t
z	Normalizing constant of a probability function
β_i, β_j	Constant terms in probability functions for x_{it} and h_{jt}
β_{ij}	Associations between product purchase decision x_{it} and hidden variable h_{jt}
β_{it}, β_{jt}	Dynamic terms in probability functions for x_{it} and h_{jt}
$D = (d_1, d_2, \cdots, d_{P-1}, d_P)$	A P dimensional vector D representing a consumer's demographic background
$x_{i'}^{t-m}$	Past purchasing decision for product *i'* and time *t-m*
β_{ijp}	Association between product purchase decision and hidden variable moderated by demographic variable d_p at time t
$\beta_{ii'p}^{t-m}$	Intertemporal effect of purchasing decision for product *i* at time *t-m* on the purchasing decision for product *i* at time *t,* moderated by demographic variable d_p. The product *i* at time *t-m* may refer to a different product in all *I* product categories than the product *i* at time *t*
$\beta_{ji'p}^{t-m}$	Intertemporal effect of purchasing decision for product *i* at time *t-m* on the hidden variable *j* at time *t,* moderated by demographic variable d_p.
β_{ip}	direct effect of demographic d_p on purchasing decisions x_{it}

$\beta_{ji'}^{t-m}$	Intertemporal effect of purchasing decision for product *i* at time *t-m* on the hidden variable *j* at time *t*
$\Delta\beta_{ij}(t)$	Increment for parameter update at iteration *t*
ε	Learning rate
α	Momentum term

APPENDIX 2 HIERARCHICAL BAYESIAN PRIOR

Figure A-2 Hierarchical Bayesian Prior



For simple illustration, intertemporal effects and demographic terms are not included in following models. In order to model heterogeneity among households, we can first place a second multinomial hidden layer on the first layer. (Salakhutdinov et al 2013):

$$P(X_{nt}, H_{nt}^{(1)} \mid H_{n}^{(2)}) = \frac{\exp(\sum_{J}\sum_{I}\beta_{ij}x_{nit}h_{njt}^{(1)} + \sum_{J}\sum_{M}\sum_{K}\beta_{kj}h_{nk}^{m(2)}h_{njt}^{(1)})}{z}$$

Where the multinomial hidden variable has *K* units, and is sampled *M* times.

For each consumer *n*, the multinomial hidden variable has a prior distribution parameterized by

Φ_n

$$h_{nk}^{(2)} \mid \Phi_n \sim Mult(1, \Phi_n)$$
$$\text{where } \Phi_{\text{n}} = (\phi_{n1}, \phi_{n2} ... \phi_{n(K-1)}, \phi_{nK})$$

Φ_n then are assumed to be generated from a global Dirichlet distribution

$$\phi_{nk} \mid \Theta \sim Dir(\Theta)$$
$$\text{where } \Theta = (\frac{\alpha}{K}, \frac{\alpha}{K} \cdots \frac{\alpha}{K}, \frac{\alpha}{K})$$

α is generated from a gamma distribution (Teh 2010).

If we integrate out the parameter Φ_n, the multinomial hidden variable follows a Dirichlet-Multinomial distribution

$$h_{nk}^{(2)} \mid \Theta \sim Dirichlet - Multinomial(\Theta)$$

For parameter inference, we can use MCMC sampling (Figure 8).

APPENDIX 3 VARIATIONAL (MEAN FIELD) INFERENCE FOR THE MODEL

According to KL divergence (Salakhutinov et al. 2013)

$$\log P(X_t; \beta) \geq \sum_H P(H_t | X_t) \log P(X_t, H_t; \beta) - \sum_H P(H_t \mid X_t) \log P(X_t \mid V_t)$$

We can use a factorized distribution Q to approximate the posterior of H_t as follows

$$Q(H_t \mid X_t; \mu) = \prod_J q(h_{jt} \mid X_t)$$

$$q(h_{jt} = 1 \mid X_t) = \mu_{jt}$$

Then the lower bound becomes

$$\log P(X_t;\beta) \geq \sum_H Q(H_t \mid X_t;\mu) \log P(X_t, H_t;\beta) - \sum_H Q(H_t \mid X_t;\mu) \log Q(H_t \mid X_t;\mu)$$

To maximize the lower bound, we set the derivative to zero and then

$$\mu_{jt} = \frac{1}{1+\exp(-\sum_I \beta_{ji} x_{it} - \beta_{jt})}$$

To predict X_t, we set $h_{jt} = \mu_{jt}, and$ estimate the expected value E(X_t) by solving simultaneously

the above equation and the conditional

probability $E(x_{it}) = P(x_{it} = 1 \mid H_t) = \dfrac{1}{1+\exp(-\beta_i - \sum_J \beta_{ji} h_{jt})}$

We compared the prediction results from variational inference with that of sampling methods.

Both methods have very similar performance, in terms of prediction accuracy.

APPENDIX 4 TRAINING ALGORITHM FOR THE FINAL MODEL

The model in this paper is coded in Matlab. The skeleton of the code appears in Taylor and Hinton (2009).

1. Divide the dataset into N mini-batches X^n $(r \times c)$, with corresponding past purchasing decisions $X^n_{<t}$and household demographics D. Randomly initialize the gradient matrices $G_{vis.hid,}$ $G_{vis,}$$G_{hid}$ and parameter matrices β. Set the momentum $\alpha,$ the learning rate ε, the dropout probability p, and the weight decay w
2. For q=1:Q (the number of iterations through the data set) do
3. For n=1:N (the number of batches) do

   Sample a binary matrix B ~p

   $M_{pos.h} = B * logit(X^n * \beta_{vis.hid} + \beta_{hid})$

   Sample a binary matrix $M_{posm} \sim M_{pos.h}$

$M_{neg.d} = logit(M_{posm} + X^n_{<t} * \beta_{past.current} + D * \beta_{demo.current} + \beta_{vis})$

$M_{neg.h} = B * logit(M_{neg.d} * \beta_{vis.hid} + \beta_{hid})$

$\beta^{t+1}_{vis.hid} = \beta^{t}_{vis.hid} + \alpha G_{vis.hid,} + \varepsilon(\frac{1}{r}\sum_R (X^n)^T M_{pos.h} - \frac{1}{r}\sum_R \left(M_{neg.d}\right)^T M_{neg.h} - w *$

$\beta^{t}_{vis.hid})$

$\beta^{t+1}_{vis} = \beta^{t}_{vis} + \alpha G_{vis} + \varepsilon(mean^{column}(X^n) - mean^{column}(M_{neg.d}) - w * \beta^{t}_{vis})$

$\beta^{t+1}_{hid} = \beta^{t}_{hid} + \alpha G_{hid} + \varepsilon(mean^{column}(M_{pos.h}) - mean^{column}(M_{neg.h}) - w * \beta^{t}_{hid})$

End for

4. End for

APPENDIX 5 CLUSTERING PRODUCTS ACCORDING TO INTERTEMPORAL EFFECTS

Product Category	Cluster	Product Category	Cluster
CHEESE COTTAGE	1	FRESH LETTUCE	9
FRANKFURTERS REFRIGERATED	1	PEANUT BUTTER	9
VEGETABLES BEANS WITH MEAT SHELF STABLE	1	VEGETABLES BEANS GREEN CANNED	9
TOILET TISSUE	2	POPCORN UNPOPPED	9
PAPER TOWELS	2	CIGARETTES	10
FACIAL TISSUE	2	DRY DINNERS PASTA	11
SEAFOOD TUNA SHELF STABLE	2	SPAGHETTI MARINARA SAUCE	11
DETERGENTS LIGHT DUTY	2	PASTA MACARONI	11
DINNERS FROZEN	3	SAUSAGE DINNER	11
ENTREES POULTRY 1	3	SEASONING DRY	11
ENTREES ITALIAN 1	3	SOFT DRINKS POWDERED	12
SANDWICHES REFRIGERATED FROZEN	3	COOKIES	13
EGGS FRESH	4	BAKERY CAKES FRESH	13
BACON REFRIGERATED	4	TOASTER PASTRIES	13
CHEESE PROCESSED SLICES AMERICAN	4	BAKERY DOUGHNUTS FRESH	13
SAUSAGE BREAKFAST	4	LUNCHMEAT SLICED REFRIGERATED	14
SUGAR GRANULATED	4	LUNCHMEAT DELI POUCHES REFRIGERATED	14
SOUP CANNED	5	SOFT DRINKS LOW CALORIE	15
BATTERIES	5	PIZZA FROZEN	15
CEREAL HOT	5	SNACKS TORTILLA CHIPS	15
NUTS CANS	5	SNACKS PRETZELS	15
VEGETABLES MIXED FROZEN	5	SNACKS REMAINING	15
NUTS BAGS	5	CRACKERS FLAVORED SNACK	15
SOUP MIXES DRY	5	CRACKERS CHEESE	15
SOFT DRINKS CARBONATED	6	FRESH MEAT	16
BAKERY BUNS FRESH	6	ENTREES REFRIGERATED	16
PRECUT FRESH SALAD MIX	6	CANDY CHOCOLATE	17
FRUIT DRINKS OTHER CONTAINER	6	CANDY NON CHOCOLATE	17
WATER BOTTLED	6	CANDY CHOCOLATE SPECIAL	17
FRUIT JUICE ORANGE	6	CANDY CHOCOLATE MINIATURES	17
FROZEN NOVELTIES	6	MAGAZINES SELECTED TITLES	18
VEGETABLES POTATOES FROZEN	6	VIDEO PRODUCTS PRERECORDED	18
SALAD DRESSING LIQUID	6	FRESH STRAWBERRIES	19
DISPOSABLE DISHES	6	FRESH TOMATOES	19
PAIN REMEDIES HEADACHE	7	TEA LIQUID	20
COLD REMEDIES ADULT	7	CAT FOOD WET TYPE	21
NUTRITIONAL SUPPLEMENTS	7	CAT FOOD DRY TYPE	21
DETERGENTS HEAVY DUTY	8	CHEESE SHREDDED	22
TOOTHPASTE	8	CHEESE PROCESSED CREAM	22
SHAMPOO AEROSOL LIQUID LOTION	8	DAIRY SOUR CREAM	22
DEODORANTS PERSONAL	8	YOGURT REFRIGERATED	23
GUM CHEWING SUGARFREE	8	FRESH CARROTS	23
DAIRY MILK REFRIGERATED	9	FRESH FRUIT REMAINING	23
BAKERY BREAD FRESH	9	FRESH APPLES	23
CEREAL READY TO EAT	9	FRUIT DRIED AND SNACKS	23
SNACKS POTATO CHIPS	9	DOG CAT TREATS	24
ICE CREAM BULK	9	DOG FOOD WET TYPE	24
MARGARINE AND SPREADS	9	BUTTER	25
FRESH POTATOES	9	DESSERTS RTS SINGLE SERVINGS	26
GROUND AND WHOLE BEAN COFFEE	9	GRANOLA YOGURT BARS	26

APPENDIX 6 OPTIMAL PRICES, REWARDS AND PROFITS UNDER ALTERNATIVE COLLECTIVE BUYING AND REFERRAL REWARD STRATEGIES

6.1. Referral Reward Strategy (RRS)

RRS (1, 2)

Here segment 1 receives a reward R contingent on the referred segment 2 making the purchase at the posted price P. The delay in the referral process is reflected in the revenue from segment 2 and the reward to segment 2 getting discounted at rate δ. Constraint (i) ensures incentive compatibility for the segment 1 to successfully refer segment 2; (ii) ensures incentive compatibility for segment 2; (iii) notes that the information shared is weakly less than the original information gap. (Similar constraints apply in the other RRS cases that follow.)

$$\underset{P,\Delta I,R}{Maximize}\ \pi = P + \delta(P - R)$$

$$\text{s.t. (i) } R - \beta\Delta I \geq 0; \text{ (ii)} \delta(\phi + V + \alpha\Delta I - P) = 0; \text{ (iii) } I \geq \Delta I \geq 0$$

This yields $P = V + \phi + \alpha I$ when $(1+\delta)\alpha > \beta$

and $P = V + \phi$ otherwise.

RRS (1, 2, 3)

This applies only for the case $\phi < \alpha I$ (i.e., segment 2 has a lower valuation than 3 to begin with.

$$\underset{P,\Delta I,R}{Maximize}\ \pi = 2P + \delta(P - R)$$

$$\text{s.t. (i) } R - \beta\Delta I \geq 0; \text{ (ii)} V + \phi + \alpha\Delta I - P = 0; \text{ (iii)} V + \alpha I \geq P; \text{ (iv) } I \geq \Delta I \geq 0$$

This yields

$$R = \beta\frac{\alpha I - \phi}{\alpha} \text{ and } P = V + \alpha I \quad \text{when } (2+\delta)\alpha > \beta$$

$$R = 0 \text{ and } P = V + \phi \text{ when } (2+\delta)\alpha \leq \beta$$

RRS (1, 2, 3, 4)

Case (a): $\phi < \alpha I$

Here, wherein segment 3's valuation is higher than those of segments 2 and 4 to begin with, there are two options: (i) in which is best to recruit segment 4 (the lowest valuation segment) through referral and (ii) in which segments 1 and 3 refer, inform and raise the valuations of both segments 2 and 4. Of course, segment 4's information gap has to be bridged more to make the segment appealing to the seller.

Option (i): Segment 4 is referred; segment 2's valuation is high enough to buy at posted price

$$\underset{P,\Delta I,R}{Maximize}\ \pi = 3P + \delta(P - R)$$

s.t. (i) $R - \beta\Delta I \geq 0$; (ii) $V + \alpha\Delta I - P = 0$; (iii) $I \geq \Delta I \geq 0$; (iv) $V + \phi \geq P$

This yields $P = V + \phi$ and $R = \dfrac{\beta\phi}{\alpha}$ when $(3+\delta)\alpha > \beta$

$P = V$ and $R = 0$ otherwise.

Option (ii): Segments 2 and 4 are referred

$$\underset{P,\Delta I,R}{Maximize}\ \pi = 2P + \delta(2P - 2R)$$

s.t. (i) $R - \beta\Delta I_1 \geq 0$; (ii) $R - \beta\Delta I_2 \geq 0$; (iii) $V + \alpha\Delta I_1 - P = 0$; $(iv)\ V + \phi + \alpha\Delta I_2 - P = 0$;
(v) $V + \alpha I - P \geq 0$; (vi) $I \geq \Delta I \geq 0$

This yields $P = V + \alpha I$ and $R = \beta I$ when $(2 + 2\delta)\alpha > \beta$ and

$P = V$ and $R = 0$ otherwise

Case (b): $\phi > \alpha I$

The price here is no higher than the valuation of segment 3, i.e., $V + \alpha I$.

$$\underset{P,\Delta I,R}{Maximize}\ \pi = 3P + \delta(P - R)$$

s.t. (i) $R - \beta\Delta I \geq 0$; (ii) $V + \alpha\Delta I - P = 0$; (iii) $I \geq \Delta I \geq 0$; (iv) $V + \alpha I \geq P$

This yields $P = V + \alpha I$ and $R = \beta I$ when $(3+\delta)\alpha > \beta$

$P = V$ and $R = 0$ otherwise

6.2. Collective Buying Strategy (CBS)

CBS (1, 2)

$$\underset{P,\Delta I}{Maximize}\ \pi = 2\delta P$$

s.t. (i) $V + \phi + \alpha I - P - \beta\Delta I \geq 0$; (ii) $V + \phi + \alpha\Delta I - P = 0$; (iii) $I \geq \Delta I \geq 0$

This yields $P = V + \phi + \dfrac{\alpha^2 I}{\alpha + \beta}$

Here constraint (i) under the objective function ensures that the price is low enough for the fully informed consumer in segment 1 to inform the less informed consumer in 2. Constraint (ii) ensures that the increase in price tied to the increased valuation of segment 2 is incentive compatible to the seller and segment 2. Constraint (iii) ensures the shared information can at most bridge the original information gap I. Similar constraints are imposed in the derivations to follow.

CBS (1, 2, 3)

Two cases arise in the 3-segment setting: $\phi < \alpha I$ (or respectively $\phi > \alpha I$) when segment 3 has a higher valuation (or respectively lower valuation) than segment 2 prior to information sharing. However, $\phi > \alpha I$ represents the case when the less informed segment 2 has a higher valuation to begin with than segment 3 (the fully informed segment with lower intrinsic valuation). Any further information between 1 and 2 will further widen the gap between segments 2 and 3. Thus, CBS (1, 2, 3) will collapse to CBS (1, 2). As such, the following derivation is when $\phi<\alpha I$.

$$\underset{P,\Delta I}{Maximize}\ \pi = 3\delta P$$

$$\text{s.t. (i)}\ V+\phi+\alpha I-\beta\Delta I-P\geq 0;\ \text{(ii)}\ V+\alpha I-\beta\Delta I-P\geq 0;\ \text{(iii)}\ V+\phi+2\alpha\Delta I-P=0$$

$$(iv)\ V+\alpha I\geq P;\ \text{(v)}\ I\geq\Delta I\geq 0$$

$$\text{This yields}\ P=V+\alpha I-\beta\frac{\alpha I-\phi}{2\alpha+\beta}$$

The constraints in the objective function ensure incentive compatibility for the seller and segments 1, 2 and 3 required for the sustainability of CBS (1, 2, 3).

CBS (1, 2, 3, 4)

Two cases matter: (a) $\phi<\alpha I$, i.e., when segment 3 has a higher valuation than segment 2 prior to information sharing and (b) $\phi>\alpha I$.

Case (a): $\phi<\alpha I$

Here segments 2 and 4 both have lower valuations than segment 3. It is optimal for segments 1 and 3 to inform 2 and 4, albeit differentially, to raise their valuations to the seller's price level. The constraints under the objective function below make it incentive compatible for all four segments to purchase the product at price P.

$$\underset{P,\Delta I}{\text{Maximize}}\ \pi = 4\delta P$$

$$\text{s.t. (i)}\ V+\phi+\alpha I-P-\beta\Delta I_1-\beta\Delta I_2\geq 0;\ \text{(ii)}\ V+\alpha I-\beta\Delta I_1-\beta\Delta I_2-P\geq 0$$

$$(iii)\ V+\phi+2\alpha\Delta I_1-P=0;\ \text{(iv)}\ V+2\alpha\Delta I_2-P=0;\ \text{(v)}\ I\geq\Delta I\geq 0;\ \text{(vi)}\ V+\alpha I\geq$$

$$\text{This yields}\ P=V+\frac{2\alpha^2 I+\beta\phi}{2\alpha+2\beta}$$

Case (b): $\phi>\alpha I$

Here segments 1 and 2 focus on reducing the information gap of segment 4 such that it becomes incentive compatible for 4 (and the other segments) to purchase the product at price P.

$$\underset{P,\Delta I}{Maximize}\ \pi = 4\delta P$$

$$\text{s.t. (i)}\ V+\phi+\alpha I-P-\beta\Delta I\geq 0;\ \text{(ii)}\ V+\alpha I-\beta\Delta I-P\geq 0;\ \text{(iii)}\ V+2\alpha\Delta I-P=0;$$

$$(iv)\ I\geq\Delta I\geq 0;\ \text{(v)}\ V+\alpha I\geq P$$

$$\text{This yields}\ P=V+\frac{2\alpha^2 I}{2\alpha+\beta}$$

Appendix 6.3. Comparative Statics

Strategy ↓	$\frac{\partial P}{\partial \alpha}$	$\frac{\partial P}{\partial \beta}$	$\frac{\partial P}{\partial \phi}$	$\frac{\partial \pi}{\partial \alpha}$	$\frac{\partial \pi}{\partial \beta}$	$\frac{\partial \pi}{\partial \phi}$
CBS (1, 2)	$\frac{2\alpha I}{\alpha+\beta}-\frac{\alpha^2 I}{(\alpha+\beta)^2}$	$\frac{-\alpha^2 I}{(\alpha+\beta)^2}$	1	$2\delta(\frac{2\alpha I}{\alpha+\beta}-\frac{\alpha^2 I}{(\alpha+\beta)^2})$	$\frac{-2\delta\alpha^2 I}{(\alpha+\beta)^2}$	2δ
CBS (1, 2, 3) $\phi<\alpha I$	$I-\frac{2\beta(\phi-\alpha I)}{(2\alpha+\beta)^2}-\frac{\beta I}{2\alpha+\beta}$	$\frac{\phi-\alpha I}{2\alpha+\beta}-\frac{\beta(\phi-\alpha I)}{(2\alpha+\beta)^2}$	$\frac{\beta}{2\alpha+\beta}$	$3\delta(I-\frac{2\beta(\phi-\alpha I)}{(2\alpha+\beta)^2}-\frac{\beta I}{2\alpha+\beta})$	$3\delta(\frac{\phi-\alpha I}{2\alpha+\beta}-\frac{\beta(\phi-\alpha I)}{(2\alpha+\beta)^2})$	$\frac{3\delta\beta}{2\alpha+\beta}$
CBS (1, 2, 3, 4) for $\phi<\alpha I$	$\frac{4\alpha I}{2\alpha+2\beta}-\frac{2(2\alpha^2 I+\beta\phi)}{(2\alpha+2\beta)^2}$	$\frac{\phi}{2\alpha+2\beta}-\frac{2(2\alpha^2 I+\beta\phi)}{(2\alpha+2\beta)^2}$	$\frac{\beta}{2\alpha+2\beta}$	$4\delta(\frac{4\alpha I}{2\alpha+2\beta}-\frac{2(2\alpha^2 I+\beta\phi)}{(2\alpha+2\beta)^2})$	$4\delta(\frac{\phi}{2\alpha+2\beta}-\frac{2(2\alpha^2 I+\beta\phi)}{(2\alpha+2\beta)^2})$	$\frac{4\delta\beta}{2\alpha+2\beta}$
CBS (1, 2, 3, 4) for $\phi>\alpha I$	$\frac{4\alpha I}{2\alpha+\beta}-\frac{4\alpha^2 I}{(2\alpha+\beta)^2}$	$-\frac{2\alpha I^2}{(2\alpha+\beta)^2}$	0	$4\delta(\frac{4\alpha I}{2\alpha+\beta}-\frac{4\alpha^2 I}{(2\alpha+\beta)^2})$	$-\frac{8\delta\alpha I^2}{(2\alpha+\beta)^2}$	0
RRS (1, 2)	1	0	1	$I(\delta+1)$	$-\delta I$	$\delta+1$
RRS (1, 2, 3)	1	0	0	$I(\delta+2)-\frac{\delta\beta I}{\alpha}-\frac{\delta\beta(\phi-\alpha I)}{\alpha^2}$	$\frac{\delta(\phi-\alpha I)}{\alpha}$	$\frac{\delta\beta}{\alpha}$
RRS (1, 2, 3, 4) for $\phi<\alpha I$ (Opt. i)	0	0	1	$\frac{\delta\beta\phi}{\alpha^2}$	$\frac{-\delta\phi}{\alpha}$	$\delta-\frac{\delta\beta}{\alpha}+3$
RRS (1, 2, 3, 4) for $\phi<\alpha I$ (Option (ii))	1	0	0	$I(2\delta+2)$	$-2\delta I$	0
RRS (1, 2, 3, 4) for $\phi>\alpha I$	1	0	0	$I(\delta+3)$	$-\delta I$	0

APPENDIX 7 PHASE DIAGRAMS ON OPTIMAL STRATEGIES: ANALYTICAL DEMARCATION OF BOUNDARIES

Tied to §4.2 of the paper on the relative attractiveness of alternative strategies, we make paired comparisons here of alternative strategies (and their subtypes) to demarcate the boundary conditions under which each pair of strategies generate the same profits. The boundary conditions will indicate how the product market conditions represented by the parameter values impact the choice of strategy. This appendix will help clarify why some boundaries in the phase diagrams are lines and others are not. In all following comparisons, segment 4's valuation V and discount factor δ are set to 1. We follow the same notation as in the text. (The paired comparisons shown here are limited to those strategies that are not dominated across the appropriate parameter space.)

MS vs. VS

When $\phi > \alpha I$

$\pi_{MS12} - \pi_{V1234} = 2(1+\phi) - 4 = 2\phi - 2$

Thus, $\pi_{MS12} > \pi_{V1234}$ when $\phi > 1$; $\pi_{MS12} < \pi_{V1234}$ otherwise.

$\pi_{MS12} - \pi_{V123} = 2(1+\phi) - 3(1+\alpha I) = -1 + 2\phi - 3\alpha I$

Thus, $\pi_{MS12} > \pi_{V123}$ when $\phi > \frac{3\alpha I + 1}{2}$; $\pi_{MS12} < \pi_{V123}$ otherwise.

When $\phi > \alpha I$

$\pi_{MS13} - \pi_{V123} = 2(1+\alpha I) - 3(1+\phi) = -1 + 2\alpha I - 3\phi$

Thus, $\pi_{MS13} > \pi_{V123}$ when $\alpha I > \frac{3\phi + 1}{2}$; $\pi_{MS13} < \pi_{V123}$ otherwise

CBS vs. VS

When $\phi > \alpha I$

$\pi_{V123} - \pi_{CBS12} = 3(1+\alpha I) - 2(1+\phi+\frac{\alpha^2 I}{\alpha+\beta}) = 1 + \frac{3\alpha^2 I + 3\alpha\beta I - 2\alpha\phi - 2\beta\phi}{\alpha+\beta}$

Thus, $\pi_{V123} > \pi_{CBS12}$ when $\phi < \frac{(\alpha+3\beta)\alpha I}{2(\alpha+\beta)} + \frac{1}{2}$; $\pi_{V123} < \pi_{CBS12}$ otherwise

When $\phi < \alpha I$

$$\pi_{\text{V123}}-\pi_{\text{CBS1234}}=3(1+\phi)-4(1+\frac{2\alpha^2 I+\beta\phi}{2\alpha+2\beta})=1+3\phi-\frac{8\alpha^2 I+4\beta\phi}{2\alpha+2\beta}$$

Thus, $\pi_{\text{V123}}>\pi_{\text{CBS1234}}$ when $\phi>\frac{4\alpha^2\text{I}+\alpha+\beta}{3\alpha+\beta}$; $\pi_{\text{V123}}<\pi_{\text{CBS1234}}$ otherwise

Within CBS

When $\phi>\alpha I$

$$\pi_{CBS12}-\pi_{\text{CBS1234}}=2(1+\phi+\frac{\alpha^2 I}{\alpha+\beta})-4(1+\frac{2\alpha^2 I}{2\alpha+\beta})=-2+2\phi-\frac{2\alpha^2 I(2\alpha+3\beta)}{(2\alpha+\beta)(\alpha+\beta)}$$

Thus, $\pi_{CBS12}>\pi_{\text{CBS1234}}$ when $\phi>1+\frac{\alpha^2 I(2\alpha+3\beta)}{(2\alpha+\beta)(\alpha+\beta)}$; otherwise $\pi_{CBS12}<\pi_{\text{CBS1234}}$

When $\phi<\alpha I$

$$\pi_{CBS12}-\pi_{\text{CBS1234}}=2(1+\phi+\frac{\alpha^2 I}{\alpha+\beta})-4(1+\frac{2\alpha^2 I+\beta\phi}{2\alpha+2\beta})=\frac{2\alpha(\phi-\alpha^2 I)}{\alpha+\beta}-2<0$$

Thus, $\pi_{CBS12}<\pi_{\text{CBS1234}}$

RRS vs. VS

When $\phi<\alpha I$

$$\pi_{\text{RRS123}}-\pi_{VS123}=3(1+\alpha I)-\frac{\beta(\alpha I-\phi)}{\alpha}-3(1+\phi)=(\frac{3\alpha-\beta}{\alpha})(\alpha I-\phi)$$

Thus, $\pi_{\text{RRS123}}>\pi_{VS123}$ when $3\alpha>\beta$; $\pi_{\text{RRS123}}<\pi_{VS123}$ otherwise

Within RRS

When $\phi<\alpha I$

$$\pi_{RRS1234(i)}-\pi_{RRS1234(ii)}=4(1+\phi)-\frac{\beta\phi}{\alpha}-(4(1+\alpha I)-2\beta I)=4(\phi-\alpha I)+\beta(\frac{2\alpha I-\phi}{\alpha})$$

Thus, $\pi_{RRS1234(i)}>\pi_{RRS1234(ii)}$ when $\beta<\frac{4(\alpha\text{I}-\phi)}{(\frac{2\alpha\text{I}-\phi}{\alpha})}$; $\pi_{RRS1234(i)}<\pi_{RRS1234(ii)}$ otherwise

$$\pi_{\mathrm{RRS1234(i)}} - \pi_{RRS123} = 4(1+\phi) - \frac{\beta\phi}{\alpha} - (3(1+\alpha I) - \beta(\frac{\alpha I - \phi}{\alpha})) = 1 + 4\phi - 3\alpha I + \beta I - \frac{2\beta\phi}{\alpha}$$

Thus, $\pi_{\mathrm{RRS1234(i)}} > \pi_{RRS123}$ when $\phi > \frac{(3\alpha I - \beta I - 1)}{4 - \frac{2\beta}{\alpha}}$; $\pi_{\mathrm{RRS1234(i)}} < \pi_{RRS123}$ otherwise

$$\pi_{RRS123} - \pi_{RRS12} = 3(1+\alpha I) - \beta(\frac{\alpha I - \phi}{\alpha}) - (2(1+\phi+\alpha I) - \beta I) = 1 + \alpha I - 2\phi + \frac{\beta\phi}{\alpha}$$

Thus, $\pi_{RRS123} > \pi_{RRS12}$ when $\alpha\mathrm{I} > \phi(2-\frac{\beta}{\alpha})-1$; $\pi_{RRS123} < \pi_{RRS12}$ otherwise

When $\phi > \alpha I$

$$\pi_{RRS1234(i)} - \pi_{RRS12} = 4(1+\alpha I) - \beta I - (2(1+\alpha I+\phi) - \beta I) = 2 + 2\alpha I - 2\phi$$

Thus, $\pi_{RRS1234(i)} > \pi_{RRS12}$ when $\phi < \alpha\mathrm{I}+1$; $\pi_{RRS1234(i)} < \pi_{RRS12}$ otherwise

CBS vs. RRS

When $\phi > \alpha I$

$$\pi_{CBS12} - \pi_{RRS12} = 2(1+\phi+\frac{\alpha^2 I}{\alpha+\beta}) - (2(1+\phi+\alpha I) - \beta I) = \frac{2\alpha^2 I}{\alpha+\beta} - 2\alpha I + \beta I = \frac{\beta I(\beta-\alpha)}{\alpha+\beta}$$

Thus, $\pi_{CBS12} > \pi_{RRS12}$ when $\alpha < \beta$; $\pi_{CBS12} < \pi_{RRS12}$ otherwise

When $\phi < \alpha I$

$$\pi_{\mathrm{CBS123}} - \pi_{\mathrm{RRS123}} = 3(1+\alpha I - \beta\frac{\alpha I - \phi}{2\alpha+\beta}) - (3(1+\alpha I) - \beta\frac{\alpha I - \phi}{\alpha}) = \frac{\alpha\beta\phi - \alpha^2\beta I + \alpha\beta^2 I - \beta^2\phi}{2\alpha^2 + \alpha\beta} = \frac{\beta(\alpha-\beta)(\phi-\alpha I)}{2\alpha^2+\alpha\beta}$$

Thus, $\pi_{\mathrm{CBS123}} > \pi_{\mathrm{RRS123}}$ when $\alpha < \beta$; $\pi_{\mathrm{CBS123}} < \pi_{\mathrm{RRS123}}$ otherwise

$$\pi_{CBS1234} - \pi_{RRS1234} = 4(1+\frac{2\alpha^2 I + \beta\phi}{2\alpha+2\beta}) - (4(1+\alpha I) - \beta I) = \frac{2\beta(\phi - (\alpha-\beta)I)}{\alpha+\beta}$$

Thus, $\pi_{CBS1234} > \pi_{RRS1234}$ when $\phi > (\alpha-\beta)\mathrm{I}$; $\pi_{CBS1234} < \pi_{RRS1234}$ otherwise

BIBLIOGRAPHY

Boztug, Y., L. Hildebrandt. 2008. Modeling joint purchases with a multivariate MNL approach. *Schmalenbach Business Review* **60** 400-422.

Burnap, A., Y. Ren, H. Lee, R. Gonzalez, P.Y. Papalambros. 2014. Improving Preference Prediction Accuracy With Feature Learning *ASME 2014 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. American Society of Mechanical Engineers, V02AT03A012-V002AT003A012.

Duvvuri, S.D., A. Ansari, S. Gupta. 2007. Consumers' price sensitivities across complementary categories. *Management Science* **53**(12) 1933-1945.

Duvvuri, S.D., T.S. Gruca. 2010. A Bayesian multi-level factor analytic model of consumer price sensitivities across categories. *Psychometrika* **75**(3) 558-578.

Fader, P.S., B.G.S. Hardie, K.L. Lee. 2005. “Counting your customers” the easy way: An alternative to the Pareto/NBD model. *Marketing Science* **24**(2) 275-284.

Friedman, J., T. Hastie, R. Tibshirani. 2001. *The Elements of Statistical Learning*. Springer series in statistics Springer, Berlin.

Hansen, K., V. Singh, P. Chintagunta. 2006. Understanding store-brand purchase behavior across categories. *Marketing Science* **25**(1) 75-90.

Hinton, G. 2010. A practical guide to training restricted Boltzmann machines. *Momentum* **9**(1) 926.

Hinton, G.E. 2002. Training products of experts by minimizing contrastive divergence. *Neural Computation* **14**(8) 1771-1800.

Hinton, G.E., R.R. Salakhutdinov. 2006. Reducing the dimensionality of data with neural networks. *Science* **313**(5786) 504-507.

Hruschka, H. 2014. Analyzing market baskets by restricted Boltzmann machines. *OR Spectrum* **36**(1) 209-228.

Keenan, D.M. 1982. A time series analysis of binary data. *Journal of the American Statistical Association* **77**(380) 816-821.

Lawrence, N.D. 2004. Gaussian process latent variable models for visualisation of high dimensional data. *Advances in Neural Information Processing Systems* **16**(3) 329-336.

Manchanda, P., A. Ansari, S. Gupta. 1999. The “shopping basket”: A model for multicategory purchase incidence decisions. *Marketing Science* **18**(2) 95-114.

Mehta, N. 2007. Investigating consumers' purchase incidence and brand choice decisions across multiple product categories: A theoretical and empirical analysis. *Marketing Science* **26**(2) 196-217.

Miles, S.. 2013. 7 strategies for boosting digital coupon conversions. Retrieved from: *http://streetfightmag.com/2013/10/03/7-strategies-for-boosting-digital-coupon-conversions/*

Pearl, J. 2014. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann.

Russell, G.J., A. Petersen. 2000. Analysis of cross category dependence in market basket selection. *Journal of Retailing* **76**(3) 367-392.

Salakhutdinov, R., G. Hinton. 2012. An efficient learning procedure for deep Boltzmann machines. *Neural Computation* **24**(8) 1967-2006.

Salakhutdinov, R., A. Mnih, G. Hinton. 2007. Restricted Boltzmann machines for collaborative filtering. *Proceedings of the 24th International Conference on Machine learning*. ACM, 791-798.

Salakhutdinov, R., J.B. Tenenbaum, A. Torralba. 2013. Learning with hierarchical-deep models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **35**(8) 1958-1971.

Seetharaman, P.B., S. Chib, A. Ainslie, P. Boatwright, T. Chan, S. Gupta, N. Mehta, V. Rao, A. Strijnev. 2005. Models of multi-category choice behavior. *Marketing Letters* **16**(3-4) 239-254.

Srivastata, N., G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov. 2014. Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research* **15**(1) 1929-1958.

Shocker, A.D., B.L. Bayus, N. Kim. 2004. Product complements and substitutes in the real world: The relevance of “other products”. *Journal of Marketing* **68**(1) 28-40.

Taylor, G.W., G.E. Hinton. 2009. Factored conditional restricted Boltzmann machines for modeling motion style *Proceedings of the 26th Annual International Conference on Machine Learning*. ACM, 1025-1032.

Teh, Y. W. 2010. Dirichlet process. *Encyclopedia of machine learning*. Springer US, 280-287

Alba, J.W., J.W. Hutchinson. 1987. Dimensions of consumer expertise. *Journal of Consumer Research* **13**(4) 411-454.

Anand, K.S., R. Aron. 2003. Group buying on the web: A comparison of price-discovery mechanisms. *Management Science* **49**(11) 1546-1562.

Bils, M. 1989. Pricing in a customer market. *The Quarterly Journal of Economics* **104**(4) 699-718.

Byers, J.W., M. Mitzenmacher, G. Zervas. 2012. Daily deals: Prediction, social diffusion, and reputational ramifications. *Proceedings of the fifth ACM International Conference on Web Search and Data Mining*. **5** 543-552.

Chen, J., X. Chen, X. Song. 2007. Comparison of the group-buying auction and the fixed pricing mechanism. *Decision Support Systems* **43**(2) 445-459.

Chen, R.R., C. Li, R.Q. Zhang. 2010. Group buying mechanisms under quantity discounts. Working paper, University of California, Davis.

Chen, R.R., P. Roma. 2011. Group buying of competing retailers. *Production and Operations Management* **20**(2) 181-197.

Chen, Y., X. Li. 2013. Group buying commitment and sellers' competitive advantages. *Journal of Economics & Management Strategy* **22**(1) 164-183.

Dholakia, U.M. 2011. How businesses fare with daily deals: A multi-site analysis of Groupon, Livingsocial, Opentable, Travelzoo, and BuyWithMe promotions. Working paper, Rice University, Houston.

Edelman, B., S. Jaffe, S.D. Kominers. 2011. To Groupon or not to Groupon: The profitability of deep discounts. *Marketing Letters* 1-15.

Hu, M., M. Shi, J. Wu. 2013. Simultaneous vs. sequential group-buying mechanisms. *Management Science* **59**(12) 2805-2822.

Jing, X., J. Xie. 2011. Group buying: a new mechanism for selling through social interactions. *Management Science* **57**(8) 1354-1372.

Kauffman, R.J., B. Wang. 2002. Bid together, buy together: On the efficacy of group-buying business models in Internet-based selling. CRC Press Boca Raton, FL, 99-137.

Liang, X., L. Ma, L. Xie, H. Yan. 2014. The informational aspect of the group-buying mechanism. *European Journal of Operational Research* **234**(1) 331-340.

Marvel, H.P., H. Yang. 2008. Group purchasing, nonlinear tariffs, and oligopoly. *International Journal of Industrial Organization* **26**(5) 1090-1105.

Rochet, J.-C., J. Tirole. 2003. Platform competition in two-sided markets. *Journal of the European Economic Association* 990-1029.

Surasvadi, N., C. Tang, G. Vulcano. 2014. Operating a group-buy mechanism in the presence of strategic consumers. Working Paper, New York University.

Wu, J., M. Shi, M. Hu. 2014. Threshold effects in online group buying. *Management Science*.

Ye, M., T. Sandholm, C. Wang, C. Aperjis, B.A. Huberman. 2012. Collective attention and the dynamics of group deals *Proceedings of the 21st International Conference Companion on World Wide Web*. ACM, 1205-1212.

Zhou, G., K. Xu, S.S.Y. Liao. 2013. Do starting and ending effects in fixed-price group-buying differ? *Electronic Commerce Research and Applications* **12**(2) 78-89.