Modeling Consumer Preferences and Price Sensitivities from Large-Scale Grocery Shopping Transaction Logs

Mengting Wan[*], Di Wang[†], Matt Goldman[†], Matt Taddy[†], Justin Rao[†],
Jie Liu[†], Dimitrios Lymberopoulos[†], Julian McAuley[*]

[*] University of California, San Diego, CA, USA, [†] Microsoft Corporation, WA, USA

[*]{m5wan, jmcauley}@ucsd.edu,
[†]{wangdi, mattgold, taddy, justin.rao, jie.liu, dlymper}@microsoft.com

ABSTRACT

In order to match shoppers with desired products and provide personalized promotions, whether in online or offline shopping worlds, it is critical to model both consumer preferences and price sensitivities simultaneously. Personalized preferences have been thoroughly studied in the field of recommender systems, though price (and price sensitivity) has received relatively little attention. At the same time, price sensitivity has been richly explored in the area of economics, though typically not in the context of developing scalable, working systems to generate recommendations. In this study, we seek to bridge the gap between large-scale recommender systems and established consumer theories from economics, and propose a nested feature-based matrix factorization framework to model both preferences and price sensitivities. Quantitative and qualitative results indicate the proposed *personalized*, *interpretable* and *scalable* framework is capable of providing satisfying recommendations (on two datasets of grocery transactions) and can be applied to obtain economic insights into consumer behavior.

Keywords

Recommender System; Consumer Behavior; Price Elasticity; Matrix Factorization

1. INTRODUCTION

Modeling consumer preferences and price sensitivities at scale is useful in both online and offline shopping worlds: Matching shoppers with the most desired products can help improve overall satisfaction, while providing appropriate promotions may lead to increased basket sizes (and revenue). Grocery shopping is one of the most frequent and regular shopping patterns in an individual or household's day-to-day activities. As a result, incredible volumes of data including transaction logs, product meta-data, and consumer demographics, can be collected from a number of offline (e.g. *Walmart*, *Kroger*, *Whole Foods*) and online (e.g. *AmazonFresh*, *walmart.com*) grocery stores and supermarkets.

WWW 2017, April 3–7, 2017, Perth, Australia.
ACM 978-1-4503-4913-0/17/04.
http://dx.doi.org/10.1145/3038912.3052568

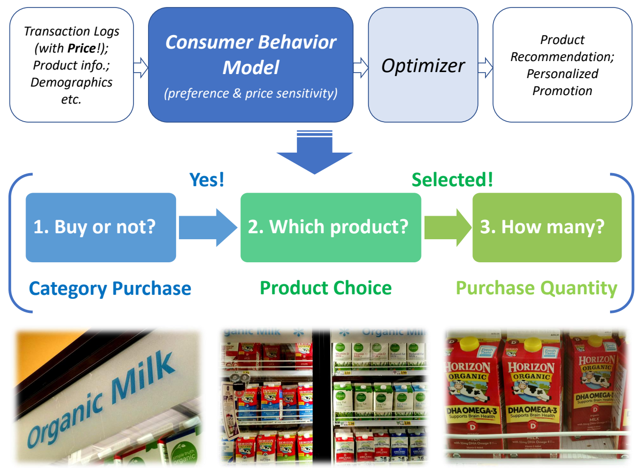

Figure 1: General workflow of the proposed three-stage purchase decision model.

Our goal in this paper is to study the problem of modeling consumer preferences and price sensitivities from large-scale grocery shopping data in order to support personalized and scalable recommendation and demand-forecasting systems.

Recommender & Demand Systems. The general workflow of the type of hybrid, large-scale recommendation and demand system we are considering is shown in Figure 1. We feed large transaction logs including product prices, meta-data, and consumer information to our behavioral model which generates feedback in the form of purchase predictions. On top of this model, we apply different optimization rules to provide user-specific results. For example, personalized ranked lists can be provided by matching preferences, or customized promotion strategies can be provided based on estimated price elasticity. Or hybrid personalized coupon lists can be provided by combining preference-matching and price-matching criteria. To achieve these goals, we need to consider consumer preferences in concert with price sensitivities in our behavioral model.

Preferences & Price Sensitivities. Product preferences are reflected by purchase incidence or purchase quantity in a consumer's shopping history. From item-based collaborative filtering [33] to matrix factorization techniques [21], various methods for consumer preference matching have been developed in the field of recommender systems. However, there are few studies where price is considered as a factor, let alone the relationship between preferences and price sensitivity.

On the other hand, price sensitivity has been richly studied in the areas of economics and marketing, from classic demand systems [13] to customized promotion models [37,38]. Demand systems are used to explore the relationship between product prices and quantities sold. In this context, price sensitivity is measured by the *'price elasticity'* value obtained from a demand system, which is defined as the unit change of purchase quantity (or probability) given a unit fluctuation in price [24]. In practice, elasticity-based consumer segments are considered and separate demand models are constructed for different consumer segments. Such segments can be regarded as useful signals for retailers and manufacturers to identify consumer groups to target. However, there are two limitations in current demand systems: 1) the data volumes involved are typically limited in terms of the number of products, categories and shopping trips[1] and 2) classic demand models are not able to be updated efficiently.

Therefore the major goal of this study is to construct an *interpretable* framework to model consumer preferences and price sensitivities *at scale*, by connecting well-developed techniques in recommender systems and well-established behavioral economic theories.

Three-Stage Purchase Decision Model. Different from modeling user preferences as a whole process, we follow the three-stage framework from recent customized promotion studies [37,38]. We notice that in real-world grocery shopping scenarios, products can be categorized either based on an existing commodity hierarchy or by clustering their associated characteristics (e.g. text descriptions). Each category should consist of some kind of products where consumers' purchase decisions share similar patterns. For example, one category might be 'organic milk' and two products in this category could be 'horizon organic whole milk' and 'organic valley whole milk.' As shown in Figure 1, we assume that for a given category, consumers' purchase decisions can be decomposed into three stages: 1) category purchase incidence, 2) product choice, and 3) purchase quantity. In a complete purchase decision-making process, stages are heterogeneous and consumers may behave quite differently across them. Given the fact that there are more than ten thousand distinct products in a typical grocery store [1], this three-stage model is more efficient compared with a flat model without fine-grained product categorization [2, 14], since it can be constructed in parallel across product categories and explicitly interpreted across different purchase stages.

Specifically, we first consider whether a consumer will make a purchase from a particular category in a certain shopping trip,[2] which can be regarded as a binary prediction problem. If so, we model their purchase from this category following a multinomial distribution. Third, we determine what quantity of the product will be purchased, which leads to a numeric prediction problem. This combination of binary, categorical, and numeric prediction is quite different from that used by traditional recommender systems, requiring new approaches to be developed. In particular, we develop a nested framework and extend state-of-the-art feature-based matrix factorization models to include price as a factor; this framework is embedded in the above prediction tasks with different link functions. We evaluate our model on two real-world grocery shopping datasets where our experiments reveal that the proposed framework is capable of providing high-quality preference predictions and personalized price sensitivity estimates.

Contributions. Specifically, our contributions are as follows:

- We model consumer preferences and price sensitivities for grocery shopping scenarios at scale, bridging a gap between large-scale recommender systems and established economic theories.
- We propose a nested feature-based matrix factorization framework, which is flexible enough to include a range of features, to fit different prediction scenarios (for different stages of purchase behavior), to be applied with scalable learning algorithms (e.g. stochastic gradient descent) and can be updated efficiently.
- By applying matrix factorization techniques, separate consumer segments no longer need to be extracted in advance and personalized price elasticity can be obtained from the model directly.
- By applying the proposed framework, we can provide economic insights from the results in our experiments. These insights include: 1) price does not significantly affect category purchase decisions, suggesting that if the general category of interest is not known, then 'deal' based promotions will be ineffective; 2) price is an important factor in the product choice stage while there is wide variance of price elasticities across categories, products and consumers, which indicates that if the category of a consumer's interest is known, it is effective to target appropriate products and consumers in order to improve the fruitfulness of promotions.

2. RELATED WORK

Preference matching has been richly studied in the area of recommender systems, where two kinds of approaches of interest have been developed: 1) content-based approaches [23,29], where explicit user profiles or item information are used as features, and 2) collaborative filtering approaches where preference predictions mainly rely on users' previous behavior [21,33]. By combining multiple techinques, hybrid recommender systems can be developed to handle a variety of complex scenarios [8,17]. Matrix factorization techniques have been widely applied for recommender systems due to their accuracy and scalability [6, 16, 21, 35]. Of particular interest, *feature-based* matrix factorization techniques have been proposed [3,4,27,28,30] and efficient tools (e.g. **SVDfeature**, **libFM**) have been developed [9,31]. Such ideas have been included in a recently proposed generalized linear mixed model (**GLMix**) [39], which has been deployed in the LinkedIn job recommender system with a scalable parallel block-wise coordinate descent algorithm. We build upon **GLMix** and adapt it to fit different prediction settings, such as multi-class classification.

[1]Typical demand system studies [13, 18, 37, 38] usually involve only several products, categories and several hundred transactions.

[2]A 'shopping trip' in this study is represented as a (consumer, timestamp) pair.

Demand systems and price sensitivity have been an ongoing focus of economists [2,13,14]. Three-stage purchase decision decomposition (i.e., category purchase; product choice; purchase quantity), such as we consider here, has been explored in several studies [5, 10, 18, 37, 38]. Customized promotion techniques have been recently proposed for offline and online shopping behavior [12, 37, 38] where individual purchase behavior is considered and optimal promotions are derived. However, these are not completely personalized demand systems and consumer segmentation is required beforehand. In addition, none of these models is considered in the context of large-scale predictive systems.

The idea of price sensitivity in recommender systems for e-commerce has been mentioned as a potential direction in a classic survey [34], though surprisingly we find that this factor has received relatively little attention. Optimization of online promotions in the context of recommendation has been recently studied [19,20], where the reservation price (i.e., the highest price a customer is willing to pay) is assumed as known information and a complete behavioral model is missing. The most related work is perhaps the price-sensitive recommender system developed in [36]. In their study, however, price is discretized into different levels rather than evaluated numerically and personalization is not thoroughly explored. Such a system thus struggles when quantitatively estimating personalized price sensitivities and cannot effectively support customized promotion strategies.

3. BACKGROUND

In this section, we introduce a generalized feature-based matrix factorization approach, which can be adjusted and applied in different purchase prediction stages. The basic notation used in this paper is provided in Table 1.

3.1 A Unified Feature-Based Matrix Factorization Model

We extend the state-of-art **GLMix** approach [39] and consider a generalized **Feature-Based Matrix Factorization (FMF)** model:

$$link(\boldsymbol{Y}(t)) = \boldsymbol{L}(t) \approx \boldsymbol{\Phi}(t)^T \boldsymbol{\Psi}(t). \tag{1}$$

Here $\boldsymbol{Y}(t)$ is the time-aware label matrix, where each element $y_{i,u}(t)$ indicates the label for an item i and a user u at timestamp t. $y_{i,u}(t)$ could be a binary label when predicting category purchase or product choice, or a numeric label when predicting purchase quantity. By applying a link function $link(\cdot)$ (e.g. the logit function, or logarithm function), we can transform the original label matrix into a numeric matrix $\boldsymbol{L}(t)$ and decompose $\boldsymbol{L}(t)$ as a product of $\boldsymbol{\Phi}(t)$ and $\boldsymbol{\Psi}(t)$. Here $\boldsymbol{\Phi}(t)$ and $\boldsymbol{\Psi}(t)$ capture both explicit features and latent factors from items and users. Specifically for each element $l_{i,u}(t)$ in $\boldsymbol{L}(t)$, we have

$$\begin{aligned} l_{i,u}(t) &\approx \langle \boldsymbol{\phi}_i(t), \boldsymbol{\psi}_u(t) \rangle \\ &= \underbrace{\left\langle \boldsymbol{w}, \overbrace{\tilde{\boldsymbol{g}}_{i,u}(t)}^{\text{global features}} \right\rangle}_{\text{global effect}} + \underbrace{\left\langle \overbrace{\tilde{\boldsymbol{\phi}}_i^{(o)}(t)}^{\text{item features}}, \boldsymbol{\psi}_u^{(o)} \right\rangle + \left\langle \boldsymbol{\phi}_i^{(o)}, \overbrace{\tilde{\boldsymbol{\psi}}_u^{(o)}(t)}^{\text{user feature}} \right\rangle}_{\text{observed item/user-specific effect}} \\ &+ \underbrace{\left\langle \boldsymbol{\phi}_i^{(l)}, \boldsymbol{\psi}_u^{(l)} \right\rangle}_{\text{latent item-user interaction}}, \end{aligned} \tag{2}$$

where $\langle \cdot, \cdot \rangle$ indicates the inner product. Here we decompose each prediction into three components: global effects, observed item/user-specific effects and latent item-user interactions.

Notation	Description
i, u, t	item, user, timestamp
(u, t)	shopping trip associated with user u at time t
$\tilde{\boldsymbol{g}}_{i,u}$, $\boldsymbol{w}$	explicit global feature, global coefficient
$\tilde{\boldsymbol{\phi}}_i^{(o)}(t)$, $\tilde{\boldsymbol{\psi}}_u^{(o)}(t)$	explicit item feature, explicit user feature
$\boldsymbol{\phi}_i^{(o)}$, $\boldsymbol{\phi}_i^{(l)}$	item random coefficient, item latent factors
$\boldsymbol{\psi}_u^{(o)}$, $\boldsymbol{\psi}_u^{(l)}$	user random coefficient, item latent factors
$\mu_u(t)$	probability of user u selecting a category
$\eta_{i,u}(t)$	conditional prob. of user u purchasing product i
$\hat{q}_{i,u}(t)$	u's conditional expected quantity of product i
$\beta_{i,u}^{(\cdot)}(t)$	coefficient associated with i's price
$e_{i,u}^{(cate)}(t)$	price elasticity in the category purchase stage
$e_{ii,u}^{(prod)}(t)$	self price elasticity in the product choice stage
$e_{i,u}^{(quant)}(t)$	price elasticity in the purchase quantity stage

Table 1: Notation.

- **Global effects.** Here $\tilde{\boldsymbol{g}}_{i,u}(t)$ includes a set of provided features for (i, u, t) and $\boldsymbol{w}$ includes a set of global coefficients which need to be estimated and should be consistent for $\forall(i, u, t)$. Such features may include general temporal and spatial factors, such as day-of-week and store location.
- **Observed item/user-specific effects.** The next term can be regarded as an analogy of the random coefficient model [22,25,36,39], which involves explicit features whose coefficients are item- or user-dependent. Here $\tilde{\boldsymbol{\phi}}_i^{(o)}(t)$ and $\tilde{\boldsymbol{\psi}}_u^{(o)}(t)$ are explicit item- and user-related features (such as item information, user demographics) while $\boldsymbol{\phi}_u^{(o)}$ and $\boldsymbol{\psi}_i^{(o)}$ are (latent) item- and user-dependent coefficients.
- **Latent item-user interactions.** The last component is designed to capture the remaining latent effects in terms of low-rank user and item factors, where both $\boldsymbol{\phi}_i^{(l)}, \boldsymbol{\psi}_u^{(l)}$ are latent parameters that need to be estimated.

Note that considering the identity link function $link(x) = x$, and discarding explicit features and timestamps, the above formulation extends typical matrix factorization formulations:

$$y_{i,u} = b_0 + b_i + b_u + \left\langle \boldsymbol{\phi}_i^{(l)}, \boldsymbol{\psi}_u^{(l)} \right\rangle. \tag{3}$$

4. METHODOLOGY

As discussed, we assume that purchase decisions can be predicted in three stages: category purchase incidence, product choice, and purchase quantity. In this section, we propose a nested framework to holistically model the interdependence of these three stages, adopting the above **FMF** model as a building block in each stage.

4.1 A Nested Factorization Framework

We notice that in different categories, consumers' purchase patterns are different, which requires us to establish a distinct behavioral model for each category. Given a product i in category c, a consumer u, and a timestamp t, suppose we have the following definitions:

$$\begin{aligned} C_u(t) &: \text{consumer } u \text{ selects the category } c \text{ at time } t; \\ B_{i,u}(t) &: \text{consumer } u \text{ purchases product } i \text{ at } t; \\ Q_{i,u}(t) = q &: \text{consumer } u\text{'s purchase quantity of } i \text{ at } t \text{ is } q. \end{aligned}$$

Thus if we focus on the category c, a consumer's preferences can be represented by the joint probability of buying a certain quantity of a particular product in category c, i.e.,

$$\begin{aligned} &P(Q_{i,u}(t) = q, B_{i,u}(t), C_u(t)) \\ &= \underbrace{P(C_u(t))}_{\substack{\text{category} \\ \text{preference}}} \times \underbrace{P(B_{i,u}(t)|C_u(t))}_{\substack{\text{conditional} \\ \text{product preference}}} \times \underbrace{P(Q_{i,u}(t) = q|B_{i,u}(t), C_u(t))}_{\text{conditional quantity preference}}. \end{aligned} \tag{4}$$

This joint probability can be regarded as a product of three conditional probabilities which represent the preferences in previous purchase stages. By adopting different link functions in the previous **FMF** formulation, these three preferences can be estimated by **L**ogistic, **C**ategorical, and **Q**uantity-based **FMF** models.

- **Category Purchase (L-FMF).** For a given category c, we have the following *logistic* probability

$$\mu_u(t) \coloneqq P_{\Theta_{cate}}(C_u(t)) = \sigma(s_u^{(cate)}(t)), \tag{5}$$

  where $\sigma(\cdot)$ is the sigmoid function. Here $s_u^{(cate)}(t)$ is a category preference score, factorized using (2), where we have only one general 'item,' i.e., the category c.

- **Product Choice (C-FMF).** Next we estimate the probability of selecting a product within a category as a multinomial distribution via a softmax formulation:[3]

$$\eta_{i,u}(t) \coloneqq P_{\Theta_{prod}}(B_{i,u}(t)|C_u(t)) = \frac{\exp(s_{i,u}^{(prod)}(t))}{\sum_{i'} \exp(s_{i',u}^{(prod)}(t))}. \tag{6}$$

  Similarly, we apply (2) to factorize the product preference score $s_{i,u}^{(prod)}(t)$.

- **Purchase Quantity (Q-FMF)** Purchase *quantity* can be represented as a positive integer in $\{1, 2, \dots\}$ and follows a shifted Poisson distribution:

$$P_{\Theta_{quant}}(Q_{i,u}(t) = q|B_{i,u}(t), C_u(t)) = \frac{z_{i,u}(t)^{q-1} \exp(-z_{i,u}(t))}{(q-1)!}, \tag{7}$$

  where $z_{i,u}(t) = \exp(s_{i,u}^{(quant)}(t))$. Again we apply (2) to factorize the quantity preference score $s_{i,u}^{(quant)}(t)$. Notice that the conditional expectation of purchase quantity can be calculated as

$$\hat{q}_{i,u}(t) \coloneqq \mathbb{E}_{\Theta_{quant}}(Q_{i,u}(t)|B_{i,u}(t), C_u(t)) = z_{i,u}(t) + 1, \tag{8}$$

  which can be regarded as an estimate of $Q_{i,u}(t)$.

Finally, we let $\Theta_{cate}, \Theta_{prod}, \Theta_{quant}$ denote the sets of parameters involved in category purchase incidence, product choice and purchase quantity prediction respectively.

[3]Note that given the fine-grained categories in our data (e.g. 'organic milk'), the multinomial assumption can be justified in most cases. If this were badly violated when users purchase several different products in the same category, this formulation is still helpful as providing the preference-based product ranked list is sufficient in the personalized promotion and recommendation scenario.

4.2 Inference

Since the three purchase stages are heterogeneous, we assume $\Theta_{cate}, \Theta_{prod}, \Theta_{quant}$ are separate parameter sets. Models for each stage can then be inferred independently. The proposed framework inherits the scalability of matrix factorization techniques, where efficient algorithms such as stochastic gradient descent can be applied [7]. We optimize all terms following the principle of maximum likelihood estimation (**MLE**). For a given category, we have the following likelihood functions for category purchase, product choice and purchase quantity:

$$\begin{aligned} \mathcal{LL}_{cate} &= \sum_{u,t} [c_u(t) \log \mu_u(t) + (1 - c_u(t)) \log(1 - \mu_u(t))], \\ \mathcal{LL}_{prod} &= \sum_{i,u,t} b_{i,u}(t) \log \eta_{i,u}(t), \\ \mathcal{LL}_{quant} &= \sum_{i,u,t} [(q_{i,u}(t) - 1) \log z_{i,u}(t) - z_{i,u}(t)] + const, \end{aligned} \tag{9}$$

where $const$ is a term independent of the parameters Θ_{quant}, $c_u(t)$, $b_{i,u}(t)$ and $q_{i,u}(t)$ are corresponding labels for $C_u(t)$, $B_{i,u}(t)$ and $Q_{i,u}(t)$.[4]

Particularly for product choice, consumer purchase behavior is a kind of implicit feedback, in the sense that *not* purchasing a particular product does not necessarily indicate that a consumer dislikes it. Thus rather than predicting if a product is selected via **MLE**, we can instead optimize a criterion that says purchased products are simply 'more preferred' than non-purchased ones. This type of optimization criterion is captured by Bayesian Personalized Ranking (**BPR**) [32], a state-of-the-art technique that approximately optimizes the area under the curve in terms of product rankings, i.e.,

$$AUC^* = \frac{1}{N} \sum_{u,t} \frac{1}{|\mathcal{P}_{u,t}^+||\mathcal{P}_{u,t}^-|} \sum_{i \in \mathcal{P}_{u,t}^+, i' \in \mathcal{P}_{u,t}^-} \delta(s_{i,u}^{(prod)}(t) > s_{i',u}^{(prod)}(t)), \tag{10}$$

where N is the total number of shopping trips for all consumers, $\mathcal{P}_{u,t}^+$ is composed of the products selected by consumer u at timestamp t and $\mathcal{P}_{u,t}^-$ includes (a random sample of) products which were not selected. Here $\delta(\cdot)$ is an indicator function ($\delta(x) = 1$ if x is true; $\delta(x) = 0$ otherwise). $\delta(s_{i,u}^{(l)}(t) > s_{i',u}^{(l)}(t)) = 1$ indicates that the consumer u prefers product i to product i' (at timestamp t). In practice, we maximize following objective function

$$\mathcal{LL}_{BPR} = \sum_{i,u,t} b_{i,u}(t) \sum_{i' \neq i} \log p_{i > i',u}(t) \tag{11}$$

where $p_{i>i',u}(t) = \sigma(s_{i,u}^{(prod)}(t) - s_{i',u}^{(prod)}(t))$.

When optimizing the parameters above we adopt a simple ℓ_2 regularization procedure in order to avoid overfitting.

4.3 Price Elasticity Estimation

We introduce the concept of 'price elasticity' to model the product price sensitivity, which is a popular measure in economics and can be defined as the responsiveness of a product's purchase quantity (or probability) to changes in its price ('self elasticity') or another product's price ('cross elasticity') [11, 18]. Self elasticity values are usually negative. Larger absolute values of elasticity indicate higher

[4]$c_u(t) = 1$ indicates the incidence of $C_u(t)$ and $b_{i,u}(t) = 1$ indicates the incidence of $B_{i,u}(t)$

price sensitivity, which means if the product price drops, its purchase probability or purchase quantity will increase accordingly. Since products within a category are often the same kind of commodities (and likely to be substitutes), the cross elasticity values in the product choice stage are usually positive, which indicates that if the product price drops, purchase probabilities of other products within the same category will decrease.

Suppose product prices are involved in previous **FMF** models by logarithmic transformations, and $P_i(t)$ is defined as the price of product i at timestamp t. Due to the linear representation of **FMF**, for a product i, we can represent the previous preference scores $s_u^{(cate)}(t), s_{i,u}^{(prod)}(t), s_{i,u}^{(quant)}(t)$ as

$$\begin{aligned} s_u^{(cate)}(t) &= r_u^{(cate)}(t) + l_u^{(cate)} + \sum_i \beta_{i,u}^{(cate)}(t) \log P_i(t), \\ s_{i,u}^{(prod)}(t) &= r_{i,u}^{(prod)}(t) + l_{i,u}^{(prod)} + \beta_{i,u}^{(prod)}(t) \log P_i(t), \\ s_{i,u}^{(quant)}(t) &= r_{i,u}^{(quant)}(t) + l_{i,u}^{(quant)} + \beta_{i,u}^{(quant)}(t) \log P_i(t). \end{aligned} \tag{12}$$

where $\beta_{i,u}^{(\cdot)}(t)$ is the coefficient associated with the price of product i, $r_{\cdot,u}^{(\cdot)}(t)$ captures (temporal and spatial) contextual information of the shopping trip (e.g. day-of-week, store location) and $l_{\cdot,u}^{(\cdot)}$ captures consumer u's category loyalty or product loyalty which is independent of the product's price and the environment of the shopping trip.[5] Then we can define the price elasticity of demand in different purchase stages.

- **Category Purchase.** For the probability of category purchase incidence and the price of product i in this category, we can define the elasticity as[6]

$$e_{i,u}^{(cate)}(t) \coloneqq \frac{d\mu_u(t)}{\mu_u(t)} \Bigg/ \frac{dP_i(t)}{P_i(t)} \approx (1-\mu_u(t))\beta_{i,u}^{(cate)}(t). \tag{13}$$

  Based on (13), if we assume that $\beta_{i,u}^{(cate)}(t)$ does not have significant variations and $e_{i,u}^{(cate)}(t) < 0$, the absolute value of $e_{i,u}^{(cate)}(t)$ will decrease as the preference prediction $\mu_u(t)$ increases.

- **Product Choice.** An advantage of our choice-based model is that product competition within a category can easily be modeled. That is, we can model the effect of a product's price change not just to its own purchase probability but other products' purchase probabilities. To do so we define the self elasticity of i as

$$e_{ii,u}^{(prod)}(t) \coloneqq \frac{d\eta_{i,u}(t)}{\eta_{i,u}(t)} \Bigg/ \frac{dP_i(t)}{P_i(t)} \approx (1-\eta_{i,u}(t))\beta_{i,u}^{(prod)}(t). \tag{14}$$

  As with (13) if $\beta_{i,u}^{(prod)}(t)$ does not vary significantly and $e_{ii,u}^{(prod)}(t) < 0$, the absolute value of $e_{ii,u}^{(prod)}(t)$ will decrease as the associated preference prediction increases. For two products i and i', we have the cross elasticity (how a price change for i affects the sales of i')

$$e_{ii',u}^{(prod)}(t) \coloneqq \frac{d\eta_{i',u}(t)}{\eta_{i',u}(t)} \Bigg/ \frac{dP_i(t)}{P_i(t)} \approx -\eta_{i,u}(t)\beta_{i,u}^{(prod)}(t). \tag{15}$$

  Notice that $\eta_{i,u}(t)e_{ii,u}^{(prod)}(t) + \sum_{i' \neq i} \eta_{i',u}(t)e_{ii',u}^{(prod)}(t) = 0$, which indicates that total choice shares must be conserved at the product selection level regardless of price fluctuations.

- **Purchase Quantity.** If we use the conditional expectation (8) as the estimation of the conditional purchase quantity, we have the following elasticity definition:

$$e_{i,u}^{(quant)}(t) \coloneqq \frac{d\hat{q}_{i,u}(t)}{\hat{q}_{i,u}(t)} \Bigg/ \frac{dP_i(t)}{P_i(t)} \approx (1 - \frac{1}{\hat{q}_{i,u}(t)})\beta_{i,u}^{(quant)}(t). \tag{16}$$

  In this scenario, if the variance of $\beta_{i,u}^{(quant)}(t)$ is limited and $e_{i,u}^{(quant)}(t) < 0$, the absolute value of price elasticity will increase as consumers' preferences increase.

Notice that an advantage of the nested **FMF** framework is that these three elasticities are additive. If we consider the price elasticity for the whole shopping trip, since

$$\begin{aligned} \mathbb{E}Q_{i,u}(t) =& \mathbb{E}(Q_{i,u}(t)|B_{i,u}(t), C_u(t))P(B_{i,u}(t)|C_u(t))P(C_u(t)) \\ =& \hat{q}_{i,u}(t)\eta_{i,u}(t)\mu_u(t) \end{aligned} \tag{17}$$

then this elasticity can be decomposed as

$$e_{i,u}^*(t) = \frac{d\mathbb{E}Q_{i,u}(t)}{\mathbb{E}Q_{i,u}(t)} \Bigg/ \frac{dP_i(t)}{P_i(t)} = e_{i,u}^{(cate)}(t) + e_{ii,u}^{(prod)}(t) + e_{i,u}^{(quant)}(t). \tag{18}$$

5. EXPERIMENTS

We evaluate the proposed nested feature-based matrix factorization framework for consumer preference prediction and price sensitivity estimation on two real-world grocery store transaction datasets. For consumer preferences, we evaluate the proposed **FMF** model's ability to make satisfying purchase predictions in terms of category purchase incidence, product choice and purchase quantity estimation. In addition, we provide analysis of the price elasticity estimations and discuss the economic insights behind these observations.

5.1 Datasets

We consider two real-world datasets of supermarket transactions. *MSR-Grocery* is a new dataset of convenience store transactions from a grocery store in the Seattle area; since this dataset is proprietary, we also evaluate our method on the public *Dunnhumby* dataset to ensure the reproducibility and extensibility of our results. Note that both datasets contain instances of variability in the price of a given product due to promotions, making them an ideal platform to study the effect of price variability on consumer behavior.

- ***Dunnhumby.*** The first dataset is the *The Complete Journey* dataset published by *Dunnhumby*.[7] This dataset includes transactions over two years from around two thousand households who are frequent shoppers at multiple stores of a retailer. Three-level category information is provided in this dataset: department, commodity description, and sub-commodity description. Here we regard the most specific one as the category indicator. We filter out small stores, infrequent shoppers, rare products, tiny categories, and finally obtain around 531 thousand product transactions[8] from 98 thousand shopping trips

[5] Notice that $r_{\cdot,u}^{(\cdot)}(t)$, $l_{\cdot,u}^{(\cdot)}$ and $\beta_{i,u}^{(\cdot)}(t)$ can be composed of both implicit parameters and explicit features.

[6] This equation can be derived based on the fact that $d(\log(x)) \approx dx/x$.

[7] `https://www.dunnhumby.com/sourcefiles`

[8] Each product transaction is for a specific product in a shopping trip.

	#product transactions	#shopping trips	#users	#trips per user
Dunnhumby	531,201	98,020	799	123
MSR-Grocery	152,021	53,075	1,228	43

	#products	#stores	#categories	#products per category
Dunnhumby	4,247	108	104	42
MSR-Grocery	1,929	1	55	35

Table 2: Basic dataset statistics.

by 799 consumers at 108 stores, across 4,247 products and 104 categories. Consumer demographic information (household age, marital status, income, homeowner description, household size, number of children, etc.) and product related information (retailer price, coupon information, manufacturer, brand, size, description, etc.) are also included. We follow the dataset specification and calculate the actual product price based on the retailer price and promotion information. By comparing the actual price and retailer price, we find that 62% of the products in transaction logs associated with these frequent shoppers were sold on sale.

- ***MSR-Grocery.*** We collected eight months of transactions from a single (anonymous) convenience/grocery store in the Seattle area. After removing invalid transactions, infrequent shoppers, rare products, tiny categories, we keep about 152 thousand product transactions from 53 thousand distinct shopping trips by 1,228 frequent consumers across 1,929 popular products in 55 categories. Some product-related features (actual price, package size, size, description) are included, though we cannot obtain any consumer demographics due to the lack of a loyalty program. Since the complete retailer price history is not available, we regard the maximum price in the transaction logs as the retailer price and compare it with the actual price. Ultimately around 50% of the products were sold on sale in this dataset.

Detailed statistics of above two datasets are included in Table 2.

5.2 Feature Instantiation

Recall that in the general **FMF** representation in (2)

$$\underbrace{\langle \boldsymbol{w}, \tilde{\boldsymbol{g}}_{i,u}(t)\rangle}_{\text{global effect}} + \underbrace{\left\langle \boldsymbol{\phi}_i^{(o)}, \tilde{\boldsymbol{\psi}}_u^{(o)}(t)\right\rangle + \left\langle \tilde{\boldsymbol{\phi}}_i^{(o)}(t), \boldsymbol{\psi}_u^{(o)}\right\rangle}_{\text{observed item/user-specific effects}} + \underbrace{\left\langle \boldsymbol{\phi}_i^{(l)}, \boldsymbol{\psi}_u^{(l)}\right\rangle}_{\text{latent interaction}}$$

both observed features and latent variables are involved. In this section, we will describe the general philosophy of feature design in the context of the consumer behavior model, and the specific features used in each purchase stage for each dataset.

- **Category Purchase.** For category purchase prediction, three global features ($\tilde{\boldsymbol{g}}_{i,u}(t)$) are considered: 1) consumer u's previous category purchase frequency, which is used to capture u's category preference; 2) category purchase quantity in u's last shopping trip, which is included to capture u's inventory information; 3) prices of products in the given category c. Since popular products may have more significant effects compared with unpopular products from the same category, we transform product prices into log-scale, weighted by their cumulative sold quantities. Since we have only one general 'item' (i.e., the selected category) at this stage, we only consider a simple consumer bias term and ignore latent item-user interactions.
- **Product Choice.** Similarly for a product i, we include the following global features: 1) previous product purchase frequency by the consumer u; 2) current price of the product i ($\log P_i(t)$). Product biases and consumer biases are included. $\log P_i(t)$ is also considered in the item features ($\tilde{\boldsymbol{\phi}}_i^{(o)}(t)$) such that each consumer and each product has their own price-related coefficients. Latent item-user interaction can be considered if provided product-related and consumer-related features are not sufficient.
- **Purchase Quantity.** For a product i, we consider the consumer u's previous average purchase quantity of the product and its current price as a global effect.

Besides the above mentioned features, additional feature configurations on the *Dunnhumby* dataset and the *MSR-Grocery* dataset can be found in Table 3.

5.3 Price History Recovery

In real cases, the complete product price history may be unavailable. Given the transaction logs, we can only observe the prices of those products sold at a certain timestamp. However, as we claimed in the previous section, prices of unsold products ought to be included in the model as well, which requires us to attempt complete price history recovery. Specifically, we applied a simple 'hot deck' method [26] for imputing these missing prices, where the transactions are sorted by timestamps and the last observed price of the same product is carried forward to the current missing price. Note that this approach can be implemented efficiently but may generate biased values if people rarely buy products at their original price. Thus we claim that developing stronger approaches to recover the complete price histories could be another important problem which can potentially be explored as future research.

5.4 Baselines and Evaluation Methodology

Baselines. Consumers' previous category purchase frequencies, product purchase frequencies and average purchase quantities can be adopted as three simple baselines – **cateFreq**, **prodFreq** and **avgQuant** for category purchase, product choice and purchase quantity predictions.

We also consider standard logistic regression (**L-Reg**) for category purchase where all the global features in Table 3 are included. For product choice, matrix factorization as in (3) (**MF-mle**) is applied to fit the multi-class classification setting.[9] **L-Reg** and **MF-mle** thus yield two learning-based recommendation benchmarks for category purchase and product choice.

Finally, we apply two sets of **FMF**-based methods for all of these three prediction stages: 1) **L-FMF-b**, **C-FMF-b-mle** and **Q-FMF-b** are three **FMF** baselines where all features in Table 3 except for product prices are included and the **MLE** optimization criterion is applied; 2) **L-FMF-p**, **C-FMF-p-mle** and **Q-FMF-p** are three full **FMF** models where product prices are added back. Comparing these two sets of baselines, the importance of product prices can be

[9]The dimension of $\phi_i^{(l)}, \psi_u^{(l)}$ is set to 5.

Dataset	**global features** ($\tilde{\boldsymbol{g}}_{i,u}(t)$)	**item features** ($\tilde{\boldsymbol{\psi}}_u^{(o)}(t)$)
Dunnhumby	category purchase freq., last purchase quant., day-of-week, storeID, household demographics, **prices** of all products	intercept
MSR-Grocery	category purchase freq., last purchase quant., day-of-week, **prices** of all products	intercept

(a) Category Purchase

Dataset	**global features** ($\tilde{\boldsymbol{g}}_{i,u}(t)$)	**item features** ($\tilde{\boldsymbol{\phi}}_i^{(o)}(t)$)	**user features** ($\tilde{\boldsymbol{\psi}}_u^{(o)}(t)$)
Dunnhumby	product purchase freq., product **price**, **price***freq., **price***day-of-week, **price***storeID	intercept, product **price**, product info. (brand, manufacturer, size description)	intercept, product **price**, household demographics
MSR-Grocery	product purchase freq., product **price**, **price***freq., **price***day-of-week	intercept, product **price**, product info. (package size, size description)	intercept, product **price**

(b) Product Choice

Dataset	**global features** ($\tilde{\boldsymbol{g}}_{i,u}(t)$)
Dunnhumby	avg. purchase quant., day-of-week, storeID, product info., household demo., product **price**, **price***(avg. quant.), **price***day-of-week, **price***storeID, **price***(product info.), **price***(household demo.)
MSR-Grocery	avg. purchase quant., day-of-week, product info., product **price**, **price***quantity, **price***day-of-week, **price***product info.

(c) Purchase Quantity

Table 3: Specific features applied in **FMF** on the *Dunnhumby* and *MSR-Grocery* datasets. Notice that coefficients for the item intercept and user intercept indicate consumer bias and product bias respectively.

evaluated. In addition to **MLE**, we adopt another method **C-FMF-p-bpr** for product choice where the **BPR** criterion (11) is used to optimize the personalized product ranking (i.e., the AUC^*) directly.

Evaluation Methodology. Note that the number of purchase incidences for each category is usually much smaller than the total of those for the remaining categories in the complete transaction logs. Therefore we apply the area under the curve (AUC) metric to evaluate the performance of category purchase prediction, which is suited to imbalanced binary prediction tasks [15].

For product choice, in real-world recommender systems, one is often interested in providing satisfactory ranked lists instead of simply predicting incidence. Thus we directly adopt the AUC^* defined in (10), which measures if the selected product is preferred to those products that were not selected in each shopping trip.

Since purchase quantity estimation is a numeric prediction task, we apply the mean absolute error (MAE) to evaluate performance where

$$MAE = \frac{1}{N^*} \sum_{i,j,t} |\hat{q}_{i,u}(t) - q_{i,u}(t)|, \tag{19}$$

and N^* indicates the total number of successful product transactions in the given category. One advantage of this measure is that the MAE is more robust to outliers than the root mean squared error (RMSE).

5.5 Results

We chronologically partition shopping trips into 70/10/20 training/validation/test splits. Because of the number of item- and user-related parameters is very large, we set two different coefficients on the ℓ_2 regularizers of the global parameters (λ_1) and the item-/user-related parameters (λ_2). These coefficients are selected on the validation set.[10] All results in this section are reported on the test data.

[10] λ_1 is selected from $\{0.1, 0.5, 1, 5\}$ and λ_2 is selected from $\{1, 5, 10, 50\}$.

5.5.1 *Preference Prediction*

We evaluate the performance for preference prediction based on the measures described in the previous section.

- **Category Purchase.** Results of category purchase prediction in terms of the AUC for binary classification are shown in Table 4a. Compared with the baseline **cateFreq**, category prediction can be significantly improved by incorporating additional features and consumer biases. However, we notice that price has little impact on performance, indicating that it may be difficult to drive consumers' category purchase decisions by altering product prices.
- **Product Choice.** For product choice prediction, we evaluate the product-ranking AUC (i.e., AUC^* in (10)). Results across different categories are provided in Table 4b. Compared with **prodFreq** and **MF-mle**, performance can be improved by incorporating more features and latent factors. We particularly notice the significance of the price feature on the *Dunnhumby* dataset by comparing the performance of **C-FMF-b-mle** and **C-FMF-p-mle**. Also in general, **C-FMF-p-bpr** reliably outperforms other **MLE**-based methods by directly optimizing ranking scores.
- **Purchase Quantity.** We include results for purchase quantity prediction in Table 4c. Again, performance can be improved by including additional features in the **Q-FMF** model but product price features do not help substantially.

5.5.2 *Price Elasticity Estimation*

Next we consider price elasticity estimation. Table 5 shows summary results (median, mean and standard deviation) of the elasticity distribution across all shopping trips in each purchase stage. Elasticity for product choice is calculated from **C-FMF-p-bpr**.

Based on the results in Table 5, price elasticity for category purchase prediction is limited. This indicates that it is hard to drive consumers' desire to purchase items from a

Dataset	*Dunnhumby* mean	*Dunnhumby* s.e.	*MSR-Grocery* mean	*MSR-Grocery* s.e.
cateFreq	0.661	0.006	0.643	0.009
L-Reg	0.722	0.006	0.657	0.009
L-FMF-b	0.782	0.005	**0.747**	0.008
L-FMF-p	**0.783**	0.005	0.746	0.007

(a) Category purchase prediction (AUC).

Dataset	*Dunnhumby* mean	*Dunnhumby* s.e.	*MSR-Grocery* mean	*MSR-Grocery* s.e.
prodFreq	0.726	0.006	0.727	0.008
MF-mle	0.723	0.006	0.641	0.006
C-FMF-b-mle	0.824	0.005	0.802	0.007
C-FMF-p-mle	0.830	0.005	0.802	0.007
C-FMF-p-bpr	**0.832**	0.005	**0.808**	0.007

(b) Product choice (AUC*).

Dataset	*Dunnhumby* mean	*Dunnhumby* s.e.	*MSR-Grocery* mean	*MSR-Grocery* s.e.
avgQuant	0.706	0.033	0.386	0.021
Q-FMF-b	**0.372**	0.023	0.123	0.022
Q-FMF-p	**0.372**	0.025	**0.115**	0.021

(c) Purchase quantity prediction (MAE).

Table 4: Mean and standard error of preference prediction results across different categories in different purchase stages on the *Dunnhumby* and *MSR-Grocery* datasets.

Dataset	*Dunnhumby* median	*Dunnhumby* mean	*Dunnhumby* s.d.	*MSR-Grocery* median	*MSR-Grocery* mean	*MSR-Grocery* s.d.
cate. purchase	-0.001	-0.012	0.029	-0.025	-0.055	0.062
product choice	-0.798	-0.842	0.683	-0.117	-0.242	0.551
purchase quant.	-0.141	-0.196	0.213	-0.004	-0.024	0.067

Table 5: Summary of self price elasticity estimation.

particular category by a single product promotion (at least for grocery shopping). Compared with category and quantity prediction, product choice is the most price sensitive stage (in terms of elasticity) in the decision making process, while price still serves as an important, but less significant, feature for quantity prediction (especially on the *Dunnhumby* dataset). We also notice that consumers in the *Dunnhumby* dataset are more price-sensitive than those in the *MSR-Grocery* dataset in the product choice and purchase quantity stages. One possible reason is that the *MSR-Grocery* dataset is collected from a convenience store, where people usually have certain targets in mind and are less likely to seek a large inventory of products. On the other hand, the *Dunnhumby* dataset is composed of household-level shopping transactions where consumers may be more likely to redeem promotions and purchase more products.

In Table 6, we provide details of the eight most price-sensitive categories in each purchase stage. From Table 6b, we notice that consumers tend to select the most inexpensive products when shopping for meat (bacon, pork rolls, tuna), eggs, drinks (water, juice, coffee), cereal and snacks (potato chips, candy). In addition, from Table 6c we observe that consumers are more likely to stock products which have relatively long shelf lives (e.g. frozen food, soft drinks) if appropriate promotions are offered. Some featured categories (categories with promotions and located in designated areas) in the *MSR-Grocery* dataset appear in Table 6a and Table 6b, which indicates that a combination of promotions and advertisements may help to affect consumers' purchase decisions.

6. CASE STUDY: BACON

Besides showing the overall preference prediction performance and the price elasticity distribution across the 104 categories in the *Dunnhumby* and 55 categories in the *MSR-Grocery* dataset, we provide detailed explorations of the most price sensitive category from the *Dunnhumby* dataset in the product choice stage: 'bacon (economy).' A summary of product prices in this category is included in Table 7, where price variabilities can be observed for all products except *product 10*. We also include the total quantity sold for each product in Table 7, where we notice that products with moderate prices are more popular than others.

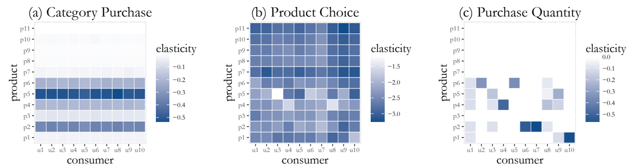

Figure 2: Heatmaps of consumer-specific price elasticity in different purchase stages for the example category 'bacon (economy).' Darker blocks indicate higher price sensitivity.

Preference vs. Representative Features. We find that the estimated coefficient on 'bacon (economy)' for *category frequency* in the category purchase stage is 0.28, which indicates that a consumer's previous category purchase frequency is still positively related to the category preference. Also the estimated coefficient for *last purchase quantity* is -0.22, which means if consumers purchased a substantial volume of economy bacon products in their previous shopping trips, they may avoid making the same category purchase in their current shopping trip. For the product choice stage and the purchase quantity stage, we find that the estimated coefficients for *product frequency* and *average purchase quantity* are 0.28 and 0.20, which indicates these two features are positively correlated with preferences as well.

Personalized Product-Specific Price Sensitivity. Among the 11 products in the example category 'bacon (economy)' there are 448 consumers who have purchased products in this category. We randomly select 10 consumers and calculate their average price elasticity for each 'bacon (economy)' product in terms of category purchase, product choice and purchase quantity decisions. Heatmaps of the results are shown in Figure 2. We notice that within the 'bacon (economy)' category, different consumers and products may have significantly different price sensitivities in each of the three stages, though the personalized elasticity is not obvious in Figure 2a since the user-specific price coefficient is not considered in the first stage. By setting appropriate thresholds for price elasticity, we can easily uncover those price sensitive consumer-product pairs in Figure 2 and customize promotion strategies accordingly. In addition, we find that in Figure 2a, consumers are more sensitive to the prices of *products 2–6* in the category purchase stage, which indeed are popular products as we observed in Table 7. This implies that while it is hard to increase the possibility of category purchase incidence, promotions on popular products will be more effective than others in terms of category purchase.

Preference vs. Price Sensitivity. Recall we claimed that if the variances of price-associated coefficients in (13), (14), (16) are limited, then consumers with high preference scores will be relatively insensitive to price changes as far as cat-

Dunnhumby		MSR-Grocery	
bacon (economy)	-0.04	broth	-0.19
soft drinks (20/24pk)	-0.04	spices	-0.17
beef (lean)	-0.01	popcorn	-0.15
garbage compactor	-0.01	energy drinks	-0.15
hot dogs	-0.01	chocolate	-0.14
pork rolls	-0.01	pizza	-0.11
salad	-0.01	tortilla chips*	-0.10
baby diapers	-0.01	protein bars	-0.10

(a) Category Purchase

Dunnhumby		MSR-Grocery	
bacon (economy)	-2.59	eggs	-1.65
milk (white)	-1.96	coffee	-1.18
butter	-1.84	chips	-1.07
cereal (family)	-1.79	bottle water	-1.06
juice	-1.78	tortilla chips*	-0.98
tuna	-1.77	bottled dill	-0.82
cereal (kids)	-1.67	fresh bread	-0.80
pork rolls	-1.67	candy	-0.80

(b) Product choice

Dunnhumby		MSR-Grocery	
mac & cheese	-0.47	chocolate	-0.12
soft drink (12/15/18pk)	-0.44	candy	-0.06
bacon (economy)	-0.41	energy drinks	-0.05
hot dogs	-0.41	mac & cheese	-0.04
milk (white)	-0.34	sausage	-0.04
pork rolls	-0.33	frozen fruit	-0.04
frozen dinner	-0.32	spices	-0.04
facial tissue	-0.32	soft drinks	-0.04

(c) Purchase Quantity

Table 6: The eight most price sensitive categories regarding three different purchase stages on the *Dunnhumby* and *MSR-Grocery* datasets. Values are the median self-elasticities within each category. Categories marked with * are composed of featured products at special locations of the store.

Product	prod. 1	prod. 2	prod. 3	prod. 4	prod. 5	prod. 6	prod. 7	prod. 8	prod. 9	prod. 10	prod. 11
mean	\$1.95	\$2.99	\$2.88	\$3.09	\$3.02	\$3.08	\$3.38	\$3.83	\$3.90	\$5.99	\$5.66
standard deviation	\$0.14	\$0.72	\$0.62	\$0.67	\$0.71	\$0.77	\$0.86	\$0.66	\$0.54	\$0.00	\$0.53
minimum	\$1.49	\$2.00	\$2.00	\$2.00	\$1.50	\$2.50	\$2.50	\$0.99	\$2.49	\$5.99	\$3.99
maximum	\$1.99	\$3.99	\$3.99	\$3.99	\$3.99	\$4.39	\$4.39	\$4.49	\$4.99	\$5.99	\$5.99
# unique values	4	4	4	4	8	5	5	4	5	1	3
quantity sold	204	373	215	455	730	507	173	64	65	32	30

Table 7: Summary of product prices and sold quantities on 'bacon (economy)'.

egory and product choice is concerned, but they tend to be price sensitive with respect to purchase quantity. Again taking 'bacon (economy)' as an example, in Figure 3 we show the relationship between preferences and price sensitivities in different purchase stages. We notice that all elasticity values are negative, which is consistent with the intuition that purchase probability will increase if product price drops. Here absolute price elasticity values are generally negatively correlated with preferences in category purchase and product choice, but positively correlated with purchase quantity, which indeed verifies our previous arguments about the relationship between preference and price sensitivity. In Figure 3b, we notice that 'low-preference' consumers have larger variations in price sensitivity than 'high-preference' consumers. This is possibly because high-preference consumers' preferences dominate purchase decisions (i.e., $1-\eta_{i,u}(t)$ is close to zero in (14)) and they tend to purchase a product no matter its price. On the other hand, if a product is not preferred by a consumer, this could be either because the price is too high to trigger a purchase, or because the consumer simply dislikes the product. In Figure 3c, we observe that those consumers with strong preferences are not the most price-sensitive consumers. This observation is consistent with the intuition that aggressive buyers are more likely to exhaust the potential of purchase quantity due to budget limits so that it would be difficult to increase their purchase quantities by adjusting price.

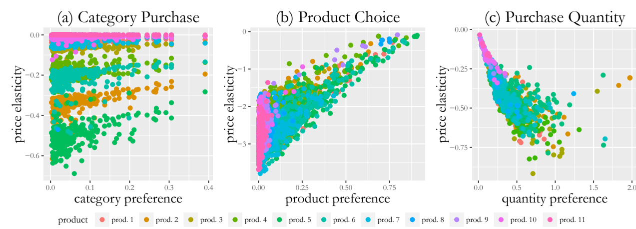

Figure 3: Scatter plots between preference prediction and price elasticity estimation in different purchase stages for the example category 'bacon (economy)'. Note that axes within each subfigure are scaled based on their own ranges.

7. CONCLUSIONS AND FUTURE WORK

We systematically studied the problem of modeling consumer preferences and price sensitivities, and proposed a nested feature-based matrix factorization framework to support personalized and scalable recommendation and demand systems. We verified that the proposed model is capable of providing high quality preference predictions and specific price elasticity can be appropriately estimated for each shopping trip. By applying the proposed framework on two real-world datasets, we provided economic insights which may benefit both data mining and economics communities. Particularly, we noticed that price affects product choice but has limited effects on category purchase or product quantity, which means coupons are primarily effective "within category". Grocery shopping behavior is particularly explored in this study but the nested multi-stage framework and the relationship between preference and price sensitivities can be translated to other domains (e.g. clothes shopping, online advertising).

Price sensitivity in large-scale systems is an important problem and a number of possible topics can be explored along this trajectory. For example, temporally-aware models could be developed to allow long-term purchase patterns to be carefully studied. Cross elasticity has been introduced but not completely explored in this work; this could be studied in detail in future work where not only product substitution but product complementarity could be modeled. In addition, since the straightforward imputation method we applied to recover price history will be problematic if people rarely buy products at their original price, another possible direction could be to develop more sophisticated approaches for price history recovery by combining preference prediction and missing price inference. In the context of hybrid recommender and demand systems, we have so far only studied consumer behavior in this work, but the optimization strategies could be adapted to generate personalized coupons.

8. REFERENCES

[1] Supermarket Facts. Food Marketing Institute, Accessed Oct. 2016. `http://www.fmi.org/research-resources/supermarket-facts`.

[2] D. A. Ackerberg. Advertising, learning, and consumer choice in experience good markets: an empirical examination. *International Economic Review*, 44(3):1007–1040, 2003.

[3] D. Agarwal and B.-C. Chen. Regression-based latent factor models. In *SIGKDD*, 2009.

[4] A. Ahmed, B. Kanagal, S. Pandey, V. Josifovski, L. G. Pueyo, and J. Yuan. Latent factor models with additive and hierarchically-smoothed user preferences. In *WSDM*, 2013.

[5] N. Arora, G. M. Allenby, and J. L. Ginter. A hierarchical bayes model of primary and secondary demand. *Marketing Science*, 17(1):29–44, 1998.

[6] L. Baltrunas, B. Ludwig, and F. Ricci. Matrix factorization techniques for context aware recommendation. In *RecSys*, 2011.

[7] L. Bottou. Large-scale machine learning with stochastic gradient descent. In *COMPSTAT*, pages 177–186, 2010.

[8] R. Burke. Hybrid recommender systems: Survey and experiments. *UMUAI*, 12(4):331–370, 2002.

[9] T. Chen, W. Zhang, Q. Lu, K. Chen, Z. Zheng, and Y. Yu. SVDFeature: a toolkit for feature-based collaborative filtering. *JMLR*, 13:3619–3622, 2012.

[10] P. K. Chintagunta. Investigating purchase incidence, brand choice and purchase quantity decisions of households. *Marketing Science*, 12(2):184–208, 1993.

[11] L. G. Cooper and M. Nakanishi. *Market-share analysis: Evaluating competitive marketing effectiveness*, volume 1. Springer Science & Business Media, 1989.

[12] P. J. Danaher, M. S. Smith, K. Ranasinghe, and T. S. Danaher. Where, when, and how long: factors that influence the redemption of mobile phone coupons. *Journal of Marketing Research*, 52(5):710–725, 2015.

[13] A. Deaton and J. Muellbauer. An almost ideal demand system. *The American economic review*, 70(3):312–326, 1980.

[14] A. M. Degeratu, A. Rangaswamy, and J. Wu. Consumer choice behavior in online and traditional supermarkets: The effects of brand name, price, and other search attributes. *International Journal of research in Marketing*, 17(1):55–78, 2000.

[15] T. Fawcett. An introduction to ROC analysis. *Pattern recognition letters*, 27(8):861–874, 2006.

[16] P. Forbes and M. Zhu. Content-boosted matrix factorization for recommender systems: experiments with recipe recommendation. In *RecSys*, 2011.

[17] A. Gunawardana and C. Meek. A unified approach to building hybrid recommender systems. In *RecSys*, 2009.

[18] S. Gupta. Impact of sales promotions on when, what, and how much to buy. *Journal of Marketing research*, pages 342–355, 1988.

[19] Y. Jiang and Y. Liu. Optimization of online promotion: a profit-maximizing model integrating price discount and product recommendation. *International Journal of Information Technology & Decision Making*, 11(05):961–982, 2012.

[20] Y. Jiang, J. Shang, Y. Liu, and J. May. Redesigning promotion strategy for e-commerce competitiveness through pricing and recommendation. *International Journal of Production Economics*, 167:257–270, 2015.

[21] Y. Koren, R. Bell, C. Volinsky, et al. Matrix factorization techniques for recommender systems. *IEEE Computer*, 42(8):30–37, 2009.

[22] N. T. Longford. Random coefficient models. In *Handbook of statistical modeling for the social and behavioral sciences*, pages 519–570. Springer, 1995.

[23] P. Lops, M. De Gemmis, and G. Semeraro. Content-based recommender systems: State of the art and trends. In *Recommender systems handbook*, pages 73–105. Springer, 2011.

[24] N. Mankiw. *Principles of Microeconomics*. Economics Series. Cengage Learning, 2011.

[25] C. E. McCulloch and J. M. Neuhaus. *Generalized linear mixed models*. Wiley Online Library, 2001.

[26] F. J. Molnar, B. Hutton, and D. Fergusson. Does analysis using "last observation carried forward" introduce bias in dementia research? *Canadian Medical Association Journal*, 179(8):751–753, 2008.

[27] X. Ning and G. Karypis. Sparse linear methods with side information for top-n recommendations. In *RecSys*, 2012.

[28] S. Park, Y.-D. Kim, and S. Choi. Hierarchical bayesian matrix factorization with side information. In *IJCAI*, 2013.

[29] M. J. Pazzani and D. Billsus. Content-based recommendation systems. In *The Adaptive Web*, pages 325–341. Springer, 2007.

[30] I. Porteous, A. U. Asuncion, and M. Welling. Bayesian matrix factorization with side information and dirichlet process mixtures. In *AAAI*, 2010.

[31] S. Rendle. Factorization machines with libfm. *TIST*, 3(3):57, 2012.

[32] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme. BPR: Bayesian personalized ranking from implicit feedback. In *UAI*, 2009.

[33] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. Item-based collaborative filtering recommendation algorithms. In *WWW*, 2001.

[34] J. B. Schafer, J. Konstan, and J. Riedl. Recommender systems in e-commerce. In *EC*, 1999.

[35] A. P. Singh and G. J. Gordon. A unified view of matrix factorization models. In *ECMLPKDD*, 2008.

[36] P. Umberto. Developing a price-sensitive recommender system to improve accuracy and business performance of ecommerce applications. *International Journal of Electronic Commerce Studies*, 6(1):1, 2015.

[37] J. Zhang and L. Krishnamurthi. Customizing promotions in online stores. *Marketing Science*, 23(4):561–578, 2004.

[38] J. Zhang and M. Wedel. The effectiveness of customized promotions in online and offline stores. *Journal of Marketing Research*, 46(2):190–206, 2009.

[39] X. Zhang, Y. Zhou, Y. Ma, B. Chen, L. Zhang, and D. Agarwal. Glmix: Generalized linear mixed models for large-scale response prediction. In *SIGKDD*, 2016.