Predictive modeling of consumer color preference: Using retail data and merchandise images

Songtao Li\(^1\) | Ruoran Chen\(^1\) | Lijian Yang\(^2\) | Dinglong Huang\(^3\) | Simin Huang\(^1\)

\(^1\)Department of Industrial Engineering, Tsinghua university, Beijing, China
\(^2\)Center for Statistical Science and Department of Industrial Engineering, Tsinghua University, Beijing, China
\(^3\)Shenzhen Malong Artificial Intelligence Research Center, Shenzhen, China

Abstract

The popularity of a fashion item depends on its color, shape, texture, and price. For different items (with all attributes identical except color) of a specific product, fashion retailers need to learn consumer color preference and decide their order quantities accordingly to match their products to consumer demand. This study aims to predict consumer color preference using the knowledge learned from merchandise images, historical retail data, and fashion trends. In our work, merchandise images are analyzed to extract color features, and the retail data of a sportswear retailer are used to reveal consumer choices among items with various colors. Choice behavior is described by a multinomial logit model, whose utility function captures the relationship between color features and popularity. Both linear functions and neural networks are applied to represent the utility function, and their out-of-sample prediction performances are compared. According to the out-of-sample performance test, our model shows reasonable predictive power and can outperform order decisions made by fashion buyers.

KEYWORDS

color features, consumer color preference, discrete-choice model, fashion retailing, neural networks

1 INTRODUCTION

In the fashion apparel retail industry, short product life cycles and uncertain market reactions always cause mismatch between supply and demand (Fisher, 1997). This mismatch either leads to “stockouts” of popular products or causes excessive inventory for slow-moving ones. According to Fisher, Hammond, Obermeyer, and Raman (1994), in the fashion industry the cost of supply–demand mismatch can actually exceed the total cost of manufacturing, which makes reducing mismatch a critical issue for fashion retailers.

Despite the importance of this “mismatch issue,” many fashion retailers do not use scientific methods to deal with it. According to Şen (2008), many fashion buyers, who are responsible for ordering products, were “far away from efficient use of sales data in their merchandise selections.” Fortunately, the development of new technologies is changing this situation. With advancements in computer vision and the vast amount of data collected in the fashion industry, data analytics and artificial intelligence have become promising tools to address the mismatch issue. According to an article on the Forbes website (Marr, 2018), fast fashion retailer H&M is using big data and artificial intelligence to gain insights about “fashion trends and their customers’ preferences.” Much of the recent literature has also developed various methodologies to understand fashion or to forecast fashion sales.

This literature can be divided into two streams according to their research domain. The first stream lies in the interdisciplinary field of fashion design and computer vision. It uses fashion product images as the main data source and applies computer vision models to discover fashionable visual patterns. These patterns are then fed to a machine learning model to train a fashion recommendation or classification system. Works in this stream include Chen,
Among the operations of a fashion apparel retailer, the “fashion buying” process, in which a fashion buyer places orders for new products, involves intensive knowledge about consumer color preference. A fashion buyer’s order decision can be decomposed into order quantities at three levels: the “product” level, “item” level, and SKU level. Here, a product is a collection of merchandise with an identical shape; an item is a collection of merchandise belonging to one product and with a specific color; and an SKU is a specific size of some item. Figure 1 illustrates an example of this classification and the corresponding order quantities, showing that a fashion buyer can help fashion retailers to learn consumer color preference. In the fashion apparel retailing industry, the color of an item can have a major effect on its popularity and market share (T.-M. Choi, Hui, Liu, Ng, and Yu (2014)). Although these forecasting modules can help fashion retailers, they do not provide insights about fashion trends or consumer preference that are contained in the visual features of fashion items.

Both sales data and fashion item images contain information about fashion trends and consumer preference. For this reason, our research makes a first attempt at integrating the two data sources and helping fashion retailers to learn consumer color preference. In the fashion apparel retailing industry, the color of an item can have a major effect on its popularity and market share (T.-M. Choi, Hui, Liu, Ng, & Yu, 2011). In fact, large differences in sale volumes can exist between different color items even when their other attributes are identical. Therefore, a fashion apparel retailer, who aims to increase sales and reduce excessive inventory, needs knowledge about consumer color preference.

To assist fashion buyers in balancing the quantity of different color items and to bridge the gaps between theory and practice, this study focuses on the predictive modeling of consumer color preference. It differs from the literature mentioned above in three aspects: (1) merchandise color features play a major role in the consumer preference model; (2) a quantitative model, not hypothesis testing results, is established by using real data; and (3) the input of our model can be a real product image instead of a single color.

Our research uses a sophisticated 2-year data set that comes from the operations of our partner sportswear retailer, which owns more than 500 brick-and-mortar stores. Purchase data, along with the inventory status of each store, offer a useful source of consumer preference
information. To separate the effect of merchandise color from that of other attributes, a conditional color choice set is defined for each consumer and it contains all the available items belonging to the specific product bought by them. As a result, the choice behavior in a conditional color choice set contains clean information on the color preference of the consumer. In addition, we extract color features from real merchandise images and use a utility-based discrete-choice model to describe the choice behavior in the conditional color choice set. Both linear functions and neural networks are used to model the utility function and their performance is compared.

During the 2-year study period, we collected $4.9 \times 10^5$ observations of choice behavior in the conditional color choice set, involving 2,606 male sports shoes. The data during the first three sale seasons (18 months) were used for model training, and the data during the last sale season (6 months) were used for performance evaluation. The results show that our best model can predict the proportion of consumers choosing different color items reasonably well. Its mean absolute percentage error (MAPE) is 28.8% when there are a sufficient number of consumers. In addition, our model appears to outperform the fashion buyers’ ordering decisions.

The rest of this article is arranged as follows. Section 2 reviews the related literature. Section 3 describes the data used in this paper. The data-processing techniques and color choice model are introduced in Section 4 and the results are presented in Sections 5 and 6. We conclude our work in Section 7. In addition, many of the details are given in the Appendices.

2 | LITERATURE REVIEW

In the fashion industry, a good trend prediction is the key to success for both manufacturers and retailers. Kim, Fiore, and Kim (2013) introduced the traditional process and methods about fashion trend prediction. Cobb and Scully (2012) specifically focused on color forecasting in the fashion industry. Compared with the qualitative methods introduced in these two books, data mining techniques have also been applied to predict color trend. Lin et al. (2010) applied a gray model to predict the color suggestion ratios of different single colors. Y. Yu et al. (2012) studied the same suggestion ratio prediction problem and compared the performance of several forecasting models. T.-M. Choi et al. (2011) applied different forecasting models to predict the yearly sales of different color fashion garments; their methods can be applied to situations with very few historical data. These papers conducted trend forecasting for a single color, whereas our approach uses more complicated color features of a real fashion item.

In addition to the prediction of fashion trends, learning consumer preference is another way to assist fashion buyers in reducing “mismatches between product attributes and consumer preferences” (Huang, Liang, & Wang, 2017). A widely accepted method of measuring consumer pref-
erence in marketing research is conjoint analysis (Green and Srinivasan, 1978, 1990). However, only a few conjoint analysis studies have considered a product’s exterior color directly. These approaches commonly describe the color attributes by using dummy variables (Deliza, Macfie, & Hedderley, 2003; Silayoi & Speece, 2007), which are not capable of capturing the color features of real merchandise. P. Choi et al. (2016) utilized CIELAB color space to accurately represent color features of a single-color backpack. They estimated a choice model to obtain the color preference and heterogeneity of the target market, which could assist firms in deciding which color to offer.

Merchandise color has also been analyzed by a stream of the marketing literature (Labrecque et al., 2013), but their research topics were different from ours. These works primarily investigated psychological responses of consumers to colors in a marketing environment (e.g., package color, logo color). Several recent studies (Deng et al., 2010; O’Donovan et al., 2011; Schloss & Palmer, 2011; Shieh & Yeh, 2015) in psychophysics are also related to our work. Deng et al. (2010) used a product self-design task to examine aesthetic color combinations; they stated that “people generally like to combine colors that are relatively close or exactly match.” Schloss and Palmer (2011) evaluated people’s perceptual responses to color pairs. The empirical results showed that color pair preference increases with hue similarity and relies on both component color preference and lightness contrast. Shieh and Yeh (2015) reported the results of an experiment in which subjects evaluated their preference for sports shoe color. L. Yu et al. (2018) used experiments and online surveys to obtain participants’ choices among six different hues. They introduced “colour consistency rate” to measure the correlation between personal color preference and purchase decisions, and they found that this correlation varied across different product categories.

O’Donovan et al. (2011) applied a least absolute shrinkage and selection operator (LASSO) regression to predict the quality of a five-color set (called a color theme). Our study differs from the above in its methodology and data set. In particular, the above literature focuses on hypothesis testing and does not use the color features of real merchandise as the stimuli.

3 | DATA SOURCE

We obtained retail data from our partner sportswear retailer that operates more than 500 nationwide specialty stores. In this study, a specific product category, namely male sports shoes, is under consideration. Restricting the product category does not cause a loss of generality because the methodology can be easily transferred to other product categories.

These retail data cover all the operations of each store, including inventory allocation, sales, transshipment, and excess stock return, from August 1, 2014, to August 8, 2016. The data consist of five data sets, each of which describes one kind of store operation. Based on these data, it is possible to track the SKU inventory at any store at any time with negligible discrepancy. The SKU level inventory tracking approaches and a detailed data description can be found in Appendix A.

In addition to the retail operations data, we collected images of male sports shoes manually from Google and used them as the source for the color information. We intentionally chose figures containing a side view of the sports shoes on a white background, which made it easier to recognize the “shoe region” of these figures (Figure 2 shows side views of sports shoes via five sketches).

Finally, we collected data about fashion color trends. In our opinion, consumer color preference not only relies on the merchandise color itself but is also influenced by the trend in fashion colors. We obtain each season’s top 10 trending colors established by the Pantone Color Insti-
tute. These colors are represented by coordinates in the CIELAB color space. A direct view of these trending colors is provided in Appendix B.

## 4 | METHODOLOGY

This research builds a predictive model of consumer color preference. Figure 3 presents the framework of our approach, in which three types of data were utilized: retail data, merchandise images, and fashion color trends. Preprocessing steps were conducted to extract color features and choice behavior and to quantify color trends. These structured data were then fed into the key component, a logit choice model describing consumer choice behavior. This model was estimated and then used to predict the proportion of consumers preferring a specific color item.

To make it easier to follow the main thread of this paper, we provide the main contents of each section here in order of occurrence. Sections 4.1–4.3 describe data preprocessing steps: (1) extracting color features; (2) quantifying color trends; and (3) extracting consumer color choices. Section 4.4 introduces the consumer color choice model, and its utility function is described in Section 4.5. Section 5 uses real data to estimate the color preference model. Finally, Section 6 reports the prediction performance.

### 4.1 | Quantification of color features

This subsection introduces the two-step procedure of quantifying the color features of sports shoes: (1) extract the h main colors from each sports shoe figure and (2) generate a visual codebook with M visual words and then calculate each shoe’s color histogram over it.

Figure 4 illustrates the details of procedure 1, which generates h main colors for each shoe figure. The ith main color is denoted by \((L_i, a_i, b_i)\), \(i \in \{1, 2, \ldots, h\}\) (coordinates in CIELAB color space), and its area percentage is denoted by \(p_i\). Appendix C describes this procedure in more detail.

Procedure 2 produces a more compact representation of color features compared with the h main colors. Suppose that our study covers \(\Gamma\) different shoes; then, there will be \(\Gamma \times h\) main colors in total. We partition these \(\Gamma \times h\) colors into \(M\) clusters by using a suitable clustering method. The \(M\) cluster centers, denoted by

\[
(L_m', a_{m}', b_{m}'), \quad m \in \{1, 2, \ldots, M\}.
\]

are treated as the M visual words in our visual codebook. Figure 5 shows an example of generating one visual codebook. It involves 2,606 shoes (\(\Gamma = 2, 606\)), each with 10 main colors (\(h = 10\)). Figure 5a is a scatter plot of the 26,060 colors in the CIELAB color space, and Figure 5b shows the 30 colors (or 30 visual words) in the codebook.

Based on the resulting visual codebook \((L_m', a_m', b_m')\), \(m \in \{1, 2, \ldots, M\}\), we can express a shoe’s color feature as a histogram:

\[
C_l = \sum_{i=1}^{h} p_i \times \left( \arg\min_{m=1,\ldots,M} (L_i - L_m')^2 + (a_i - a_m')^2 + (b_i - b_m')^2 \right),
\]

where \(C_l, l \in \{1, 2, \ldots, M\}\) is the area percentage of color \((L_l', a_l', b_l')\) contained by a shoe, and \(\sum_{l=1}^{M} C_l = 1\). Figure 6 visualizes the color histogram of an example shoe on the visual codebook in Figure 5b.

### 4.2 | Quantification of color trends

In our study, a variable \(\text{FashArea}_s\) is introduced for each shoe for each season. We define \(\text{FashArea}_s\) as the percentage of a shoe’s color that can be treated as a trending color in season \(s\), where the top 10 trending colors established by the Pantone Color Institute are denoted by

\[
(L_k^s, a_k^s, b_k^s), \quad k \in \{1, 2, \ldots, 10\}.
\]

Intuitively, if we define a distance metric between a shoe’s \(h\) main colors and the trending colors, then we can introduce a distance threshold \(D'\) to judge whether a main color is a trending color or not. In this article, we use a Euclidean distance metric in CIELAB color space. The distance threshold \(D'\) can be selected through cross-validation on the training set whose details are described in Section 5.3. For a given \(D'\), the calculation process of \(\text{FashArea}_s\) is provided in Appendix D.

### 4.3 | Extraction of consumer choices among different color items

As discussed in Section 4.1, a shoe’s color features in season \(s\) are described by a numeric array \((C_{1s}, C_{2s}, \ldots, C_{Ms}, \text{FashArea}_s)\). After quantifying these color-related features, we need to extract consumer choice behavior (from the retail data) to learn their color preference. We propose a creative approach in which consumer color preference is inferred from their choice behavior among the different items belonging to one product.

To introduce our approach, we first provide an illustrative example of consumer choice behavior representation, and then we provide the corresponding notation.
In Figure 7a, consumer \( k \) buys the item in red, whereas three different color items of the same product in his shoe size are available for him at the same store. A notion “conditional color choice set” is used to denote the set of color features of the three items. For consumer \( k \) in Figure 7a, the conditional color choice set is denoted by \( X_g = \{x_{g1}, x_{g2}, x_{g3}\} \), where \( x_{g1}, x_{g2}, \text{and} x_{g3} \) are the color features of three different color items. Now we can describe the color choice behavior of consumer \( k \) as “choosing \( x_{g3} \) among three alternatives in his conditional color choice set \( X_g = \{x_{g1}, x_{g2}, x_{g3}\} \).” To describe the color preference of the whole sports shoes market, in Figure 7b we summarize all the consumers faced with identical conditional color choice set to that of consumer \( k \) and count the number of consumers choosing each alternative. According to this figure, \( q_g \) consumers buying this specific shoe product only have three different colors of shoes to choose from. Among these \( q_g \) consumers, \( q_{g1} \) of them prefer color feature \( x_{g1} \), \( q_{g2} \) prefer color feature \( x_{g2} \), and \( q_{g3} \) prefer color feature \( x_{g3} \). Given this, we can describe consumers’ color choices by using the actual purchase counts \( (q_{g1}, q_{g2}, q_{g3}) \) of the different color items \( \{x_{g1}, x_{g2}, x_{g3}\} \) of a specific product.

We need various notations for a more formal description of the consumer choice behavior mentioned in the previous example. We use the notion of a consumer group to represent all the consumers that share the same conditional color choice set. For example, the \( q_g \) consumers in Figure 7b form consumer group \( g \) because they all have the same conditional color choice set \( \{x_{g1}, x_{g2}, x_{g3}\} \). Several additional symbols are defined, as are summarized in Table 1. The color features in the conditional color choice set for consumer group \( g \) is denoted by \( X_g \), whose \( j \)th element is \( x_{gj} = 1, 2, \ldots, |X_g| \). Here, \( q_g \) denotes the total number of consumers in consumer group \( g \) and \( q_{gj} \) denotes the number of consumers buying alternative \( x_{gj} \in X_g \).

In summary, we cluster consumers with an identical conditional color choice set into one consumer group and use the purchase count of each alternative to capture the color choice behavior of consumers. Because all the items in the conditional color choice set are from the same shoe product, the difference in shoe shape, material, or price can be well controlled (our partner retailer always offers identical prices and price discounts for a different color of the same pair of shoes).

### 4.4 Discrete-choice model of color features

The previous two subsections introduce how we obtain the three data components (color features, trending information, and consumer choice behavior) in Figure 3. To link these data, we now introduce the central segment in Figure 3, the discrete-choice model.
FIGURE 6  Color histogram ($C_1, C_2, \ldots, C_{30}$) of an example shoe [Colour figure can be viewed at wileyonlinelibrary.com]

FIGURE 7  Example of consumer choice behavior on a conditional color choice set [Colour figure can be viewed at wileyonlinelibrary.com]

For a consumer $k$ who is clustered into consumer group $g$ and has conditional color choice set $X_g = \{x_{g1}, x_{g2}, \ldots, x_{g|X_g|}\}$, assume his color choice behavior can be described by a multinomial logit model with utility function $f(\cdot)$ (mapping a color feature vector to a utility value). As a result, the probability of the consumer choosing each alternative in the conditional color choice set is

$$P(\text{buy } x_{gj} | x_{gj} \in X_g) = \frac{\exp[f(x_{gj})]}{\sum_{l \in \{1,2,\ldots,|X_g|\}} \exp[f(x_{gl})]}.$$  \hfill (3)

Given the color features $X_g = \{x_{g1}, \ldots, x_{g|X_g|}\}$ and the choice count vector $Q_g = (q_{g1}, \ldots, q_{g|X_g|})$ for each consumer group $g$, the log-likelihood value of all the observed consumer color choice behaviors for some specific utility function $f(\cdot)$ is displayed in the following equation, where

$G$ is the total number of different consumer groups:

$$l(f) = \sum_{g=1}^{G} \left[ \log \prod_{j=1}^{|X_g|} \left( \frac{\exp[f(x_{gj})]}{\sum_{l=1}^{|X_g|} \exp[f(x_{gl})]} \right)^{q_{gj}} \right].$$

To estimate $f(\cdot)$, we need to find a suitable function (denoted by $\hat{f}(\cdot)$) maximizing $l(f)$ in Equation (4).

The estimated utility function $\hat{f}(\cdot)$ describes the utility of a shoe’s color for the average population. Generally speaking, if a shoe’s color feature $x$ has higher utility $\hat{f}(x)$, a larger proportion of consumers will choose to buy it. However, $\hat{f}(\cdot)$ only provides a preference metric for color; it does not take other important attributes (e.g., price, material, functioning) into account. For this reason, it is mostly useful to compare consumer preference among items belonging to the same product. For example, assume there are 1,000 consumers clustered into a consumer group with two alternatives in the conditional color choice set, whose color utility is 1.4 and 0.9, respectively. Then, we
can predict that the proportion of consumers choosing the two items are 62.2% \( \left( e^{1.4}/(e^{1.4} + e^{0.9}) \right) \) and 37.8% \( (e^{0.9}/(e^{1.4} + e^{0.9})) \).

### 4.5 Estimation of utility function \( f(\cdot) \)

A proper functional form of \( f(\cdot) \) is necessary before we estimate it. In our article we propose three alternatives: (1) linear utility function; (2) linear utility function with interaction terms; and (3) a neural-network-based utility function.

The first alternative is a simple linear utility function with the functional form

\[
f(FashArea_s, C_2, \ldots, C_M) = \beta_1 FashArea_s + \beta_2 C_2 + \beta_3 C_3 + \ldots + \beta_M C_M. \tag{5}
\]

The \( C_1 \) term is omitted because of collinearity (\( \sum_{i=1}^{M} C_i = 1 \)).

The second alternative has a similar form to Equation (5), but additionally includes interaction terms such as \( C_iC_j \) to increase its flexibility. By maximizing the log-likelihood function, we can obtain the estimation for parameters in these two linear form utility functions (to reduce overfitting issues, we also include ridge penalty terms in estimation, and the details are refereed to in Section 5.4).

The third alternative is an artificial neural network. Bentz and Merunka (2000) provided a detailed methodology of using a neural-network-based utility function in a multinomial logit-choice model. Following their approach, we constructed a neural network model to describe consumer color choice behavior. Figure 8 shows the structure of the neural network model, where \( NN(\cdot) \) denotes a neural network with a color feature vector as its input and the color utility as its output (with \( K \) denoting the size of set \( X_g \)).

Given the color features \( X_g = \{x_{g1}, x_{g2}, \ldots, x_{g|X_g|}\} \) for consumer group \( g \), the artificial neural network produces the utility for each alternative as \( U_{gl} = NN(x_{g1}), \ldots, U_{g|X_g|} = NN(x_{g|X_g|}) \). The choice probability is identically defined as the multinomial logit model. For the consumer group \( g \) in Figure 8, the cross-entropy loss function of the neural network in Equation 8 can be calculated by

\[
\text{Loss} = -\sum_{j=1}^{|X_g|} q_{gj} \log \left( \frac{\exp(U_{gj})}{\sum_{l=1}^{|X_g|} \exp(U_{gl})} \right). \tag{6}
\]

The loss function in Equation (6) is nearly identical (except for the negative sign) to the log-likelihood function in Equation (4), so the newly introduced neural network model has an identical objective of maximizing the log-likelihood value to that of the previous multinomial logit model.

### 5 TRAINING THE COLOR PREFERENCE MODEL

After having introduced the methodology, we now report the results of color preference model estimation. First, some summary statistics of the consumer groups are displayed. Second, we chose an appropriate number \( M \) that denotes the number of visual words in the codebook and chose a proper distance threshold to calculate \( FashArea_s \).

Third, the utility functions of the color preference model are estimated. The descriptions follow.

#### 5.1 Summary statistics of consumer choice behaviors

Our data set comes from the operations of a sportswear retailer during a 2-year period. It covers 2,606 male shoe items from 662 different shoe products. Table 2 provides some statistics on the consumer groups in each season (Appendix E provides sample data on several consumer groups). Row 4 lists the number of consumer groups in each season. Row 5 calculates the average size of each group’s conditional color choice set. Rows 6–8 count the number of products, items, and consumers involved in each season.

#### 5.2 Selection of a good visual codebook

Section 4.1 describes the method to extract the \( h \) main colors from a shoe image and to construct an \( M \)-color visual codebook. In this article, we arbitrarily set \( h = 10 \). As most
FIGURE 8  Artificial neural network structure of consumer color choice
[Colour figure can be viewed at wileyonlinelibrary.com]

<table>
<thead>
<tr>
<th>Season</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of groups</td>
<td>1960</td>
<td>3235</td>
<td>3092</td>
<td>2,594</td>
</tr>
<tr>
<td>Average $</td>
<td>X_g</td>
<td></td>
<td>2.99</td>
<td>3.02</td>
</tr>
<tr>
<td>Number of products</td>
<td>195</td>
<td>314</td>
<td>335</td>
<td>204</td>
</tr>
<tr>
<td>Number of items</td>
<td>645</td>
<td>1,150</td>
<td>1,199</td>
<td>1,020</td>
</tr>
<tr>
<td>Number of consumers</td>
<td>105,236</td>
<td>126,592</td>
<td>150,037</td>
<td>96,525</td>
</tr>
</tbody>
</table>

TABLE 2  Summary statistics on the consumer groups in each season

sports shoes have fewer than 10 different colors, $h = 10$ is sufficient. Nevertheless, selecting an appropriate value for $M$ (thus generating a good visual codebook) is not that trivial. If $M$ is too small, the corresponding visual codebook cannot describe shoe color in detail. However, if $M$ is too large, the color feature vector $(C_1, C_2, \ldots, C_M)$ will be very long and our model may suffer from the curse of dimensionality.

To obtain the best visual codebook, a series of codebooks with different $M$ value were generated, and we chose that with $M = 30$, which achieves the best cross-validation performance on the training set (Appendix F provides the details). Figure 9 presents the resulting 30-color visual codebook, where we use $C_1, C_2, \ldots, C_{30}$ to name these colors. Table 3 summarizes the average values of color feature $(C_1, C_2, \ldots, C_{30})$ based on the current codebook.

5.3  Selection of threshold color distance to define $FashArea_s$

According to Section 4.2, a color distance threshold $D'$ is needed to calculate $FashArea_s$. A series of candidate threshold values $\{2, 3, 4, 5, 10\}$ were tested by a 10-fold cross-validation. During each fold of cross-validation, the utility function Equation (5) was estimated by maximizing the log-likelihood function with ridge penalty terms (whose regularization factor leads to the best performance on each validation set). Finally, the results are reported in Table 4.

The second column of Table 4 provides the cross-validation performance if $FashArea_s$ is not included, and the next several columns indicate the performance of including $FashArea_s$, which is calculated by different distance thresholds. The result is quite surprising because it indicates that $FashArea_s$ (which represents trending color information) does not increase our model’s performance. One possible explanation is that the men’s sports shoe industry is not led by the top 10 colors of Pantone. If another special source of trending colors for the sports shoe market were available, we could still apply our method to extract the $FashArea_s$ and add it into the utility function. Currently, according to the cross-validation results, this trend information is removed from our model. Although this may cause loss of trend information, the current approach is already good enough to learn color preference, considering that our observation period is relatively short and consumer color preference will not change dramatically. Moreover, in the conclusion of this article, obtaining a useful source of trending color information is treated as a future research direction.
5.4 Utility function training

According to Sections 4.3 and 4.4, \( X_g = \{ x_{g1}, x_{g2}, \ldots, x_{g|X_g|} \} \) and \( Q_g = \{ q_{g1}, q_{g2}, \ldots, q_{|Q_g|} \} \) are linked by a logit model with a utility function \( f(\cdot) \). Our task is to estimate the function \( f(\cdot) \) to best explain the observed \( X_g \) and \( Q_g \) pairs. The observations from September, 2014, to February, 2016, were used as training data.

First, we estimated the simple linear utility function, as displayed in Equation (5), except for the FashionArea term. The parameters \( \beta_2, \beta_3, \ldots, \beta_M \) were estimated by maximizing the log-likelihood function with ridge penalty terms:

\[
\text{Regularized } l(f) = \sum_{g=1}^{G} \left\{ \sum_{j=1}^{|X_g|} q_{gj} f(x_{gj}) - q_g \log \left( \sum_{j=1}^{|X_g|} \exp(f(x_{gj})) \right) \right\} - \lambda (\beta_2^2 + \beta_3^2 + \ldots + \beta_M^2).
\]

A 10-fold cross-validation on the training set was conducted to select a good ridge regularization factor \( \lambda \), and the best value was selected as \( \lambda = \exp(4.2) \) (the details of the cross-validation are referred to in Appendix G). With the specified regularization factor \( \lambda = \exp(4.2) \), we can obtain the estimated utility function parameter \( \beta_2, \beta_3, \ldots, \beta_M \) by maximizing Equation (7). The estimated value of each parameter is reported in the “Model 1” column in Table 6.

Second, the linear utility function with interaction terms was estimated. According to Schloss and Palmer (2011), a color combination’s preference depends not only on its component color but also on the contrast between these components. Adding interaction terms can address this issue and also increase the flexibility of our utility function. To restrict the number of parameters and make the utility function concise, we defined several variables in Table 5 to represent the area percentage of different color hues (RAL color list was used as a reference point to assign hue names; more details are given in Appendix H). By adding the interaction terms of these hue variables, the utility function is now

\[
f(C_1, C_2, \ldots, C_{30}) = \beta_2 C_2 + \beta_3 C_3 + \ldots + \beta_M C_{30} + a_1 \text{White} \cdot \text{Grey} + a_2 \text{White} \cdot \text{Black} + a_3 \text{White} \cdot \text{Oran} + \ldots + a_28 \text{Blue} \cdot \text{Green}.
\]

The regularized log-likelihood function is now

\[
\text{Penalized } l(f) = \sum_{g=1}^{G} \left\{ \sum_{j=1}^{|X_g|} q_{gj} f(x_{gj}) - q_g \log \left( \sum_{j=1}^{|X_g|} \exp(f(x_{gj})) \right) \right\} - \lambda (\beta_2^2 + \beta_3^2 + \ldots + \beta_M^2) - \gamma (a_1^2 + a_2^2 + \ldots + a_28^2).
\]

Again, we selected the two regularization factors \( \lambda \) and \( \gamma \) by 10-fold cross-validation on the training set. The process is given in Appendix I and the final regularization factor choice is \( \lambda = \exp(4.4) \) and \( \gamma = \exp(3.0) \). The estimated value of the parameters \( \beta_2, \beta_3, \ldots, \beta_{30} \) are reported in the “Model 2” column in Table 6, and the estimated value of the parameters \( a_1, a_2, \ldots, a_{28} \) are reported in Table 7.

Third, the neural network utility function \( NN(\cdot) \) was estimated. \( NN(\cdot) \) is a feedforward neural network with two hidden layers and logistic activation function. Appendix I describes its structure in formulas and documents the 10-fold cross-validation process of selecting the number of nodes in both layers. Finally, we chose the structure with 10 nodes in hidden layer 1 and 15 nodes in hidden layer 2, and we set the number of training epochs (how many times to repeat the training process using all training data) as 64, which is the average early stopping epoch during the 10-fold cross-validation.
6 PERFORMANCE EVALUATION ON THE TEST SET

The test set, which contains data on season $s = 4$, was used to test the prediction power of our color popularity model. In the rest of this section, first we introduce the accuracy metric of our model prediction; second, a benchmark of color popularity prediction is created; finally, we present the performance of our estimated model and compare it with benchmarks.

6.1 Accuracy metric definition

A performance metric was constructed based on the notion of a consumer group. As mentioned earlier, consumer group $g$ has $q_g$ consumers, whose conditional color choice set is $X_g = \{x_{g1}, x_{g2}, \ldots, x_{g|X_g|}\}$. The vector $Q_g = (q_{g1}, q_{g2}, \ldots, q_{g|X_g|})$ stores the number of consumers buying each alternative. We transform $Q_g$ into a probability vector $P_g$, as shown in Equation (10), where $p_{gj}$ is the observed proportion of consumers choosing alternative $j$. By using our choice model and the estimated utility function $\hat{f}(\cdot)$, we can also calculate the predicted value of $p_{gj}$ and $q_{gj}$, as shown in Equations (11) and 12, respectively:

$$P_g = (p_{g1}, p_{g2}, \ldots, p_{g|X_g|}) = \left(\frac{q_{g1}}{q_g}, \frac{q_{g2}}{q_g}, \ldots, \frac{q_{g|X_g|}}{q_g}\right), \quad (10)$$

$$\hat{P}_g = (\hat{p}_{g1}, \hat{p}_{g2}, \ldots, \hat{p}_{g|X_g|}), \quad \hat{p}_{gj} = \frac{\exp[\hat{f}(x_{gj})]}{\sum_{l=1}^{|X_g|} \exp[\hat{f}(x_{gl})]}, \quad j = 1, 2, \ldots, |X_g|, \quad (11)$$

$$Q_g = (q_{g1}, q_{g2}, \ldots, q_{g|X_g|}) = \left(q_{g1} \hat{p}_{g1}, q_{g2} \hat{p}_{g2}, \ldots, q_{g|X_g|} \hat{p}_{g|X_g|}\right). \quad (12)$$

These variables were then used to define the accuracy metric. If our model has strong predictive power, then the discrepancy between $\hat{P}_g$ and $P_g$ (or between $\hat{Q}_g$ and $Q_g$) should be small for the consumer groups in the test set. We introduce $MAPE_g$ to measure the prediction discrepancy for consumer group $g$:

$$MAPE_g = \frac{1}{|X_g|} \sum_{l=1}^{|X_g|} |\hat{p}_{gj} - p_{gj}|. \quad (13)$$

In addition, the average $MAPE_g$ of consumer groups was used as the overall accuracy metric.

6.2 Color popularity predicted by fashion buyers

We used fashion buyers’ ordering decisions to construct a benchmark. The order quantities can reflect the fashion buyer’s opinion about consumer color preference for different color items. Therefore, a benchmark model can be constructed intuitively. Let $P'_g = (p'_{g1}, p'_{g2}, \ldots, p'_{g|X_g|})$ denote the benchmark prediction, then we can set $P'_{gj}, j \in \{1, 2, \ldots, |X_g|\}$ proportional to each item’s procurement quantity. For example, a conditional color choice set of consumer group 1 consists of two different color items. To infer the value of $P'_1$, we obtained the procurement quantity of the two items, denoted by $I_1 = (I_{11}, I_{12})$. Finally, the benchmark prediction for the actual observed $P_1$ is

$$P'_1 = \left(\frac{I_{11}}{I_{11} + I_{12}}, \frac{I_{12}}{I_{11} + I_{12}}\right).$$

6.3 Performance of color preference model and benchmark

Finally, the performance evaluation of our model prediction is reported. The data in the original test set...
TABLE 6  Estimated parameters in the utility function: Part A

<table>
<thead>
<tr>
<th>$L, a, b$</th>
<th>Model 1</th>
<th>Model 2</th>
<th>$L, a, b$</th>
<th>Model 1</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>88, 0, 1</td>
<td>−0.32</td>
<td>−0.44</td>
<td>41, 48, −9</td>
<td>−0.11</td>
<td>−0.09</td>
</tr>
<tr>
<td>74, 3, 3</td>
<td>0.10</td>
<td>0.70</td>
<td>50, 34, −40</td>
<td>0.14</td>
<td>0.18</td>
</tr>
<tr>
<td>57, 2, 3</td>
<td>−0.03</td>
<td>−0.18</td>
<td>30, 34, −60</td>
<td>−0.34</td>
<td>−0.17</td>
</tr>
<tr>
<td>15, 2, −6</td>
<td>1.02</td>
<td>0.70</td>
<td>26, 14, −31</td>
<td>0.34</td>
<td>0.41</td>
</tr>
<tr>
<td>76, 3, 65</td>
<td>−0.32</td>
<td>−0.16</td>
<td>43, 20, −54</td>
<td>−0.60</td>
<td>−0.37</td>
</tr>
<tr>
<td>59, 13, 41</td>
<td>0.14</td>
<td>0.28</td>
<td>49, 5, −39</td>
<td>−0.88</td>
<td>−0.55</td>
</tr>
<tr>
<td>62, 44, 53</td>
<td>−0.18</td>
<td>0.05</td>
<td>35, 0, −8</td>
<td>0.54</td>
<td>0.43</td>
</tr>
<tr>
<td>70, 28, 26</td>
<td>−1.00</td>
<td>−0.77</td>
<td>58, −6, −24</td>
<td>0.43</td>
<td>0.35</td>
</tr>
<tr>
<td>53, 61, 40</td>
<td>−0.17</td>
<td>0.09</td>
<td>51, −21, −2</td>
<td>−0.15</td>
<td>0.06</td>
</tr>
<tr>
<td>33, 37, 21</td>
<td>0.24</td>
<td>0.53</td>
<td>78, −26, 1</td>
<td>−0.38</td>
<td>−0.14</td>
</tr>
<tr>
<td>44, 50, 28</td>
<td>−0.03</td>
<td>0.18</td>
<td>48, −30, 25</td>
<td>0.06</td>
<td>0.01</td>
</tr>
<tr>
<td>20, 20, 9</td>
<td>0.59</td>
<td>0.74</td>
<td>79, −38, 55</td>
<td>−0.62</td>
<td>−0.39</td>
</tr>
<tr>
<td>57, 29, 10</td>
<td>−0.21</td>
<td>−0.02</td>
<td>27, −4, 11</td>
<td>0.19</td>
<td>0.49</td>
</tr>
<tr>
<td>33, 30, 2</td>
<td>−0.15</td>
<td>−0.06</td>
<td>83, −11, 34</td>
<td>−0.56</td>
<td>−0.30</td>
</tr>
<tr>
<td>59, 57, −9</td>
<td>0.09</td>
<td>0.10</td>
<td>47, −8, 28</td>
<td>−0.23</td>
<td>−0.08</td>
</tr>
</tbody>
</table>

TABLE 7  Estimated parameters in the utility function: Part B

<table>
<thead>
<tr>
<th>Term</th>
<th>Value</th>
<th>Term</th>
<th>Value</th>
<th>Term</th>
<th>Value</th>
<th>Term</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red*Violet</td>
<td>0.27</td>
<td>Violet*Blue</td>
<td>0.11</td>
<td>Blue*Oran</td>
<td>0.55</td>
<td>Green*Black</td>
<td>−0.02</td>
</tr>
<tr>
<td>Red*Blue</td>
<td>0.80</td>
<td>Violet*Green</td>
<td>−0.04</td>
<td>Blue*Grey</td>
<td>−0.28</td>
<td>Oran*Grey</td>
<td>−0.04</td>
</tr>
<tr>
<td>Red*Green</td>
<td>0.46</td>
<td>Violet*Oran</td>
<td>0.08</td>
<td>Blue*White</td>
<td>−0.91</td>
<td>Oran*White</td>
<td>−0.41</td>
</tr>
<tr>
<td>Red*Oran</td>
<td>−0.41</td>
<td>Violet*Grey</td>
<td>−0.37</td>
<td>Blue*Black</td>
<td>0.70</td>
<td>Oran*Black</td>
<td>−0.14</td>
</tr>
<tr>
<td>Red*Grey</td>
<td>−1.14</td>
<td>Violet*White</td>
<td>0.13</td>
<td>Green*Oran</td>
<td>−0.42</td>
<td>Grey*White</td>
<td>1.14</td>
</tr>
<tr>
<td>Red*White</td>
<td>−0.81</td>
<td>Violet*Black</td>
<td>−0.18</td>
<td>Green*Grey</td>
<td>−0.10</td>
<td>Grey*Black</td>
<td>1.12</td>
</tr>
<tr>
<td>Red*Black</td>
<td>1.20</td>
<td>Blue*Green</td>
<td>1.94</td>
<td>Green*White</td>
<td>−1.12</td>
<td>White*Black</td>
<td>1.20</td>
</tr>
</tbody>
</table>

FIGURE 11  Scatter plot of $q_g$ compared with $\hat{q}_g$ in the test set [Colour figure can be viewed at wileyonlinelibrary.com]

were filtered to eliminate old products (that were found in the training set), as described in Appendix J. The final test set after filtering contained 567 consumer groups, which involved 415 items from 140 products. There were thus 63,732 consumer purchase observations.

To display our model’s prediction performance, we divided the test-set consumer groups into several sets according to their $q_g/|X_g|$ (which represents the average number of consumers buying each alternative) and report the average MAPE in each set. The prediction model could perform better for consumer groups with a larger $q_g$ because those groups’ observations contain less noise and are more in accord with the population’s color preference. Table 8 summarizes the preference prediction models introduced previously, and Table 9 displays the prediction performance.

In Table 9 the first column lists the $q_g/|X_g|$ range of each group set. Columns 2 and 3 report the number of consumer groups and number of consumers observed in each group set, respectively. The next three columns report the performance of our three color preference prediction models, and the last column reports the performance of our benchmark. We can draw three findings according to the result. First, our model’s prediction matches the observations better when more consumers are in a consumer group. Second, our model outperforms fashion buyers’ prediction. According to the last row of Table 9, there could be an 18% increase in average MAPE. Third, the linear utility function without any interaction terms had the
worst performance among the three models proposed. The neural-network-based utility function can capture some nonlinear components in consumer color preference and therefore had better prediction power than the previous simple linear utility function; thus it had similar performance to the linear utility function including hue interaction terms.

Finally, to gain an overview of the prediction performance of our estimated model, Figures 10 and 11 show scatter plots of \( \hat{p}_{gj} \) compared with \( p_{gj} \) and \( \hat{q}_{gj} \) compared with \( q_{gj} \) (for Model 2), respectively. In Figure 10, a specific data point’s coordination is \( (p_{gj}, \hat{p}_{gj}) \), where \( g \) is the index of the consumer group and \( j \in \{1, 2, \ldots, |X_g|\} \). The size of this data point was set proportional to \( \sqrt{|q_g|} \), the square root of the number of consumers in consumer group \( g \), because when consumer group \( g \) has a larger \( q_g \) (more consumers) its observed choice proportion vector \( P_g \) is more reliable and thus it should be given more importance. We also drew a 45° red line in Figure 10. All data points should lie on it if our model makes perfect predictions. Figure 11 was drawn in a similar way to present the scatter plot of consumer choices. The left part of Figure 11 offers an overview and the right part is a close-up view of the region \([0, 200] \times [0, 200]\).

Figure 10 shows that the predicted value \( \hat{p}_{gj} \) approximately matches the observed value \( p_{gj} \), which means that the prediction made by our model can generally capture the pattern of consumer color preference. Nevertheless, this prediction still has a significant error because many of the prediction values of \( \hat{p}_{gj} \) deviate from the observed \( p_{gj} \). Moreover, the errors for groups with a greater number of consumers (i.e., a larger \( q_g \)) are generally smaller, because few large points lie far from the red line. Figure 11 provides another view of prediction performance. Again, we find that for consumer group \( g \) the predicted consumer number choosing each alternative matches the observed data well. Further, many consumer groups have fewer than 100 consumers (i.e., with \( q_g < 100 \)); therefore, the data points \((q_{gj}, \hat{q}_{gj})\) are denser in the bottom left corner.

### 7 | CONCLUSION

This study analyzed consumer color preference in the fashion retailing industry to predict the popularity of a product’s items in various colors before any sales data were observed (popularity is represented by the proportion of consumers preferring each item). Predicting consumer color preference is an important factor in fashion retailers’ decision making, because a better understanding of consumer color preference for the upcoming season will help retailers decide on the procurement quantity of a product’s different items and thus reduce mismatch between supply and demand.

Most research on color preference tests hypotheses to validate some given rules rather than establishing a quantitative color preference model. Our study bridges this gap and makes several contributions. First, it proves the practicality of learning valuable knowledge from different data sources. In our study, three data sources were utilized: retail data, merchandise images, and color trend predictions. These data sources were integrated to build the color preference model. Second, a framework was built for color preference learning that contained three data components: merchandise color data, trending color information, and consumer choice behaviors. The first two components formed the color feature of each item and the last contained information about consumer color preference. The central part of our framework was a discrete-choice model, which linked the color features and consumer choice behavior. Third, we introduced an innovative way to obtain clean information about consumer color preference. For each consumer, we constructed a conditional color choice set, which corresponded to the items (that belonged to the same product that they bought) from which this consumer could choose. In this way, the influence of other attributes (e.g., material, price, functioning) are controlled for and the effects of color preference are separated. The data analysis presented here provides a good illustration of these contributions, and the estimated color preference model

<table>
<thead>
<tr>
<th>Model</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>Discrete-choice model with utility function in Equation (5)</td>
</tr>
<tr>
<td>Model 2</td>
<td>Discrete-choice model with utility function in Equation (8)</td>
</tr>
<tr>
<td>Model 3</td>
<td>Discrete-choice model with neural network-based utility function</td>
</tr>
</tbody>
</table>

**TABLE 8** Three preference prediction model and two benchmarks

| \( q_g / |X_g| \) range | Groups | Consumers | Model 1 | Model 2 | Model 3 | Benchmark |
|-----------------|--------|-----------|---------|---------|---------|-----------|
| [2.6, 7.4)      | 139    | 2,113     | 53.4%   | 52.8%   | 52.7%   | 71.4%     |
| [7.4, 16.5)     | 140    | 4,129     | 48.5%   | 46.2%   | 47.5%   | 60.3%     |
| [16.5, 42.8)    | 142    | 9,173     | 38.6%   | 36.7%   | 38.0%   | 46.1%     |
| [42.8, 1164.3)  | 139    | 48,308    | 30.1%   | 28.8%   | 29.2%   | 35.5%     |

**TABLE 9** Average MAPE on consumer groups with different consumer numbers
of the consumer population shows reasonable predictive power in our separate test set.

Although our model can improve the current order decisions made by the fashion buyers, it still has some insufficiencies that could be enhanced by future research. First, the performance of the neural networks can be improved if more training data are available. The current training data contain fewer than 3,000 shoes, which might limit the ability of neural networks. Second, if we could find a good trending color source leading the sports shoe market, we could add it to our model and improve its prediction power.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Songtao Li https://orcid.org/0000-0002-2530-8023

REFERENCES


**AUTHOR BIOGRAPHIES**

**Songtao Li** Ph.D. candidate at the department of Industrial Engineering, Tsinghua University. He received his bachelor's degree from the Department of Physics at Tsinghua University. Songtao Li's main research interest includes optimal searcher path problems, data-driven decisions in fashion supply chain and empirical studies in retail industries.

**Ruoran Chen** is the Post-Doctorate Fellow at the Department of Industrial Engineering, Tsinghua University. He received his PhD from the department of Industrial Engineering at Tsinghua University. Dr. Chen's main research interest includes procurement and risk management in commodity supply chains, empirical studies in supply chain and retailing industries, and data-driven decision methods.

**Lijian Yang** received Ph.D. in Statistics from the University of North Carolina at Chapel Hill in 1995. Currently, he is a Professor at the Center for Statistical Science and Department of Industrial Engineering, Tsinghua University, Beijing, China. He was previously Assistant, Associate and Full Professor at the Department of Statistics and Probability, Michigan State University, and Specially-Appointed Professor and Director of the Center for Advanced Statistics and Econometrics Research, at Soochow University in China. Lijian Yang won the Tjalling C. Koopmans Econometric Theory Prize, and was elected as Fellow of the American Statistical Association, Fellow of the Institute of Mathematical Statistics, and Member of the International Statistical Institute. He has published over 70 articles in peer reviewed journals such as Journal of the American Statistical Association, Journal of the Royal Statistical Society B, Annals of Statistics, Annals of Probability, Journal of Econometrics, Biometrics, Statistica Sinica, Journal of Time Series Analysis, on statistical inference of complex data that include time series, high dimensional data, functional data, and sample survey data, as well as applications of statistics in econometrics, food science, agronomy, genetics and neuroscience.

**Dinglong Huang** is co-founder and CEO of Malong Technologies, a global leader in artificial intelligence headquartered in Shenzhen, China. Raised in the small seaside community of Yantian during Shenzhen’s pre-boom years, Dinglong was a stand-out student, obtaining his Ph.D from department of Industrial Engineering, Tsinghua University. From there, Dinglong excelled in technology leadership roles at Google, Microsoft, Tencent and Tripadvisor. In 2014, Dinglong co-founded Malong Technologies, an awardwinning artificial intelligence (AI) startup in China. The company is on a mission to help its enterprise customers transform with AI to achieve higher efficiency, quality and safety, by creating machines that can “see” physical objects such as products.

**Simin Huang** is a Professor of the Department of Industrial Engineering, Tsinghua University. He received his Ph.D. in Operations Research at SUNY at Buffalo. His current research interests include

---

**How to cite this article:** Li S, Chen R, Yang L, Huang D, Huang S. Predictive modeling of consumer color preference: Using retail data and merchandise images. *Journal of Forecasting*. 2020;1–19. [https://doi.org/10.1002/for.2689](https://doi.org/10.1002/for.2689)

---

**APPENDIX A: RETAIL DATA DESCRIPTION**

Table A1 describes the content of each data set. For a specific store, these data capture the flow of SKUs both into and out of it, allowing us to track the inventory of a specific SKU at a specific store at any time. The process is briefly described in the following.

We followed two steps to achieve this goal: (1) Add an artificial time point “08:00:00” to the inaccurate time stamps in the out-transshipment data and return-to-warehouse data; and (2) extract all records related to one SKU at one store, sort them into increasing time order, and calculate the cumulative inventory. As the time stamps of the out-transshipment data and return-to-warehouse data have discrepancies, whenever these two operations are made it causes several hours of
TABLE A1  Five data sets for the different operations

<table>
<thead>
<tr>
<th>Data set</th>
<th>Operations</th>
<th>Content</th>
<th>Format of time stamp</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shipment data</td>
<td>Shipments received by each store</td>
<td>Store ID, SKU ID, quantity</td>
<td>2014/08/01 08:00:00</td>
</tr>
<tr>
<td>Point-of-sale data</td>
<td>Sales transactions at each store</td>
<td>Store ID, SKU ID, quantity</td>
<td>2014/08/01 08:00:00</td>
</tr>
<tr>
<td>In-transshipment data</td>
<td>Transshipment received by each store</td>
<td>Store ID, SKU ID, quantity</td>
<td>2014/08/01 08:00:00</td>
</tr>
<tr>
<td>Out-transshipment data</td>
<td>Transshipment delivered by each store</td>
<td>Store ID, SKU ID, quantity</td>
<td>2014/08/01</td>
</tr>
<tr>
<td>Return-to-warehouse data</td>
<td>Excessive inventory returned by each store</td>
<td>Store ID, SKU ID, quantity</td>
<td>2014/08/01</td>
</tr>
</tbody>
</table>

TABLE A2  Example of generating the SKU inventory time series

<table>
<thead>
<tr>
<th>Operations</th>
<th>Time</th>
<th>Inventory change</th>
<th>Real inventory</th>
<th>Calculated inventory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shipments received</td>
<td>2014/08/01 19:23:15</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Sale</td>
<td>2014/08/03 11:46:20</td>
<td>−1</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>Sale</td>
<td>2014/08/05 15:10:04</td>
<td>−1</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Sale</td>
<td>2014/08/08 20:30:55</td>
<td>−1</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Sale</td>
<td>2014/08/25 17:01:18</td>
<td>−1</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Sale</td>
<td>2014/08/29 08:00:00</td>
<td>0</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Out-transshipment</td>
<td>2014/08/29 11:30:24</td>
<td>−2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Sale</td>
<td>2014/09/02 18:49:47</td>
<td>−1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

inaccuracy in the time series. However, a store generally returns excessive inventory or delivers a transshipment of a specific SKU once (usually at the middle or end of the season); thus our method causes several hours of inaccurate time series. Compared with the several months sales span of one SKU, this error is negligible. Table A2 provides an example to illustrate the inventory tracking method and small discrepancy.

In Table A2 the first and second columns list the store’s operations and time of each operation, respectively. The third column records the inventory change caused by each operation and the inventory afterwards is in the fourth column. In the fifth column, we calculate the inventory time series. The discrepancy between our calculated inventory and the real one was caused by the artificial time “2014/08/29 08:00:00” of the out-transshipment (which actually happens at “11:30:24”). However, this only lasts for 3.5 hours, which is very short compared with the life cycle (several months) of this SKU.

APPENDIX B: PANTONE’S TOP 10 COLORS

Figure B1 displays the top 10 trending colors for men during our study period, established by the Pantone Color Institute. Table B1 lists the details of the top 10 trending colors for men during the fall/winter season of 2014. For each color, its coordinates in the CIELAB color space are listed in this table along with its name.

APPENDIX C: EXTRACTION OF THE H MAIN COLORS

The main color extraction procedure, as shown in Figure 4, involves three steps:

1. The left part of Figure 4 shows an example in which the shoe region is distinguished from the white background. By using the MATLAB Image Tool Box, we sequentially applied edge detection, image dilation, and interior hole filling to the original image to obtain a
two-dimensional binary matrix representing the shoe region.
2. Assume the shoe region contains \( N \) pixels. Their color information is stored in an \( N \times 3 \) matrix (shown in the middle of Figure 4). Row \( i = 1, 2, \ldots, N \) of this matrix, \( (L'_i, a'_i, b'_i) \), is the CIELAB coordinates of pixel \( i \).
3. Treat \( (L'_i, a'_i, b'_i), i \in \{1, 2, \ldots, N\} \) as the \( N \) points in the CIELAB color space. By applying K-means cluster algorithms to these points, \( h \) cluster centers were obtained, which represent the \( h \) main colors of this shoe. Further, the number of data points in each cluster was transformed into the area percentage of each main color. The main colors and their area percentages are denoted by \( (L_s, a'_k, b'_k), k \in \{1, 2, \ldots, 10\} \) of season \( s \):

\[
D_i^s = \min_{k=1,2,\ldots,10} \sqrt{(L_i - L_s^k)^2 + (a_i - a_k^s)^2 + (b_i - b_k^s)^2}. \tag{D1}
\]

After \( D_i^s \), \( i \in \{1, 2, \ldots, h\} \) was obtained, we artificially introduced a distance threshold \( D^s \) to decide whether main color \( i \) was a trending color or not. The logic of this process is described as follows:
1. If \( D_i^s \leq D^s \), main color \( i \) is a trending color during season \( s \).
2. If \( D_i^s > D^s \), main color \( i \) is not a trending color during season \( s \).

Finally, we can calculate the area percentage of the shoe covered by a trending color as

\[
\text{FashArea}_i = \sum_{i=1}^{h} p_i \times I\{D_i^s \leq D^s\}. \tag{D2}
\]

In this equation, \( I\{D_i^s \leq D^s\} \) has the value 1 if \( D_i^s \leq D^s \)
and 0 if \( D_i^s > D^s \).

**APPENDIX D: CALCULATION OF THE VARIABLE FASHAREAS**

For the main color \( i \in \{1, 2, \ldots, h\} \) of this shoe, we define a distance metric \( D_i^s \) in Equation (D1). Here, \( D_i^s \) measures the minimum distance between this main color \( i \) and the top 10 trending colors \( (L_i^k, a_i^s, b_i^s), k \in \{1, 2, \ldots, 10\} \) of season \( s \):

**APPENDIX E: SAMPLE DATA ON CONSUMER GROUP OBSERVATIONS**

In Table E1 we provide the sample observations for five consumer groups during season \( s = 1 \). Column 1 records the index of each consumer group. The next two columns are the product ID and item ID. Column 4 represents the number of consumers preferring each item (or the \( Q_g \) vector for consumer group \( g \)). For example, among those consumers clustered into consumer group 1, 429 of them choose the item whose id ends with “-038” and 233 of them choose the other item whose id ends with “-261.” That is, \( Q_1 = (429, 233) \).

**APPENDIX F: SELECTION OF A GOOD VISUAL CODEBOOK VIA CROSS-VALIDATION**

The final \( M \) value is chosen from a series of candidates \{10, 15, 20, 25, 30, 35, 40\}. For each value of \( M \), five different visual codebooks with \( M \) colors are generated randomly (these five visual codebooks are slightly different from each other because of the randomness in the K-means method). In total, we had 40 different visual codebooks as the candidates to be tested via cross-validation. The training data covered all the consumer groups observations before March of 2016, which were further divided into 10 subsamples to conduct a 10-fold cross-validation. In each round of the cross-validation process, nine subsamples were selected as the temporary training data and the linear utility function:

\[
f(C_2, \ldots, C_M) = \beta_2 C_2 + \beta_3 C_3 + \ldots + \beta_M C_M,
\]

was estimated to maximize the likelihood function with ridge penalty term. Afterwards, we used the estimated utility function to calculate the negative log-likelihood value on the last subsample. We repeated this process 10 times and the summation of the 10 negative log-likelihood values on temporary validation set were the cross-validation performance of the current candidate visual codebook. (In fact, the cross-validation performance depends on the regularization factor \( \lambda \) in the ridge penalty term. A series of cross-validation performances were obtained under a series of regularization factors \( \lambda \) and we only reserved the one under the best \( \lambda \) value.) Table F1 records the cross-validation performance of the 40 candidate visual

---

**TABLE B1** The top 10 trending colors, September 2014 to February 2015 (established by Pantone)

<table>
<thead>
<tr>
<th>Season</th>
<th>( L' )</th>
<th>( a' )</th>
<th>( b' )</th>
<th>Color name</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014 fall</td>
<td>38</td>
<td>47.8</td>
<td>1.9</td>
<td>Sangria</td>
</tr>
<tr>
<td>2014 fall</td>
<td>43.8</td>
<td>51.1</td>
<td>26.9</td>
<td>Aurora Red</td>
</tr>
<tr>
<td>2014 fall</td>
<td>52.9</td>
<td>40.6</td>
<td>-21</td>
<td>Radiant Yellow</td>
</tr>
<tr>
<td>2014 fall</td>
<td>76.4</td>
<td>1.6</td>
<td>49.1</td>
<td>Misted Yellow</td>
</tr>
<tr>
<td>2014 fall</td>
<td>37.2</td>
<td>-8.7</td>
<td>13.9</td>
<td>Cypress</td>
</tr>
<tr>
<td>2014 fall</td>
<td>39.3</td>
<td>2.4</td>
<td>-32.4</td>
<td>Bright Cobalt</td>
</tr>
<tr>
<td>2014 fall</td>
<td>32</td>
<td>19.7</td>
<td>-40.9</td>
<td>Royal Blue</td>
</tr>
<tr>
<td>2014 fall</td>
<td>62.9</td>
<td>9.2</td>
<td>-4.9</td>
<td>Sea Fog</td>
</tr>
<tr>
<td>2014 fall</td>
<td>62.2</td>
<td>-0.3</td>
<td>8.2</td>
<td>Aluminum</td>
</tr>
<tr>
<td>2014 fall</td>
<td>44.9</td>
<td>14.7</td>
<td>10.8</td>
<td>Cognac</td>
</tr>
</tbody>
</table>
codebooks. Each row represents one $M$ value, and columns 2–6 record the cross-validation performance (negative log-likelihood value) for each of the five randomly generated visual codebooks. The best choice seemed to be $M = 30$, because its average performance was the best and its case 1 visual codebook had the best performance of all the 40 candidates.

**APPENDIX G: CROSS-VALIDATION TO DECIDE THE $\lambda$ VALUE IN EQUATION 7**

A properly chosen $\lambda$ value can reduce the overfitting issue on the training set and thus increase the model’s prediction power. We apply a 10-fold cross-validation on the training set to select a good ridge regularization factor $\lambda$. The procedure is similar to the cross-validation referred to in Appendix F; thus we just report the cross-validation performance of different $\lambda$ choices in Table G1. The performance of the estimated utility function first increases with $\lambda$ because of overfitting reduction. After reaching the highest point, increasing the penalty strength will depress the model’s explanation power because all the parameters are pushed to zero. According to Table G1, the best $\lambda$ value $\lambda = \exp(4.4)$ was selected.

**APPENDIX H: DEFINITION OF HUE VARIABLES BASED ON THE RAL COLOR LIST**

RAL is a color matching system used in Europe. The RAL classic list contains 213 different colors and each color is assigned a hue name from yellow, orange, red, violet, blue, green, grey, brown, white, and black. To divide the 30 colors in our color code book, each color was matched to the most similar color in RAL classic list and the hue name of that specific RAL color was used as the hue name of our code book color.

**APPENDIX I: CROSS-VALIDATION TO SELECT REGULARIZATION FACTORS AND NEURAL NETWORK STRUCTURE**

First, Table I1 presents the cross-validation performance (negative log-likelihood value) of the linear utility function in Equation (8) for different combinations of two regularization factors $\lambda$ and $\gamma$. Each row of Table I1 lists the results for a fixed $\lambda$ and each column displays the results for a fixed $\gamma$. From the cross-validation results, we should select $\lambda = \exp(4.4)$ and $\gamma = \exp(3.0)$.

Second, we introduce the detail structure of the neural network $NN(\cdot)$ and the cross-validation process of selecting the right network structure. $NN(\cdot)$ is a feedforward neural network with two hidden layers. Letting $h_j^{(1)}$ denote the values of node $j$ in the first hidden layer (with $H_1$ nodes), it can be calculated by

$$h_j^{(1)} = F\left(\sum_{i=2}^{30} w_{ij}^{(1)} C_i + b_j^{(1)}\right),$$

where $w_{ij}^{(1)}$ and $b_j^{(1)}$ are parameters and $F(\cdot)$ is the logistic activation function $F(x) = 1/(1 + e^{-x})$. The values of node $j$ in hidden layer 2 (with $H_2$ nodes) are based on the value
TABLE G1 Cross-validation performance of different $\lambda$ (on a log scale)

<table>
<thead>
<tr>
<th>log($\lambda$)</th>
<th>1.0</th>
<th>2.0</th>
<th>3.0</th>
<th>3.6</th>
<th>3.8</th>
<th>4.0</th>
<th>4.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perform.</td>
<td>325,560.1</td>
<td>325,478.9</td>
<td>325,383.8</td>
<td>325,343.9</td>
<td>325,336.1</td>
<td>325,331.5</td>
<td>325,330.4</td>
</tr>
<tr>
<td>log($\lambda$)</td>
<td>4.2</td>
<td>4.3</td>
<td>4.4</td>
<td>4.6</td>
<td>5.0</td>
<td>6.0</td>
<td>—</td>
</tr>
<tr>
<td>Perform.</td>
<td>325,330.1</td>
<td>325,330.7</td>
<td>325,332.1</td>
<td>325,337.7</td>
<td>325,360.3</td>
<td>325,503.3</td>
<td>—</td>
</tr>
</tbody>
</table>

TABLE I1 Cross-validation performance of different regularization factor combinations

<table>
<thead>
<tr>
<th>$\log(\lambda)$, $\log(\gamma)$</th>
<th>2.4</th>
<th>2.6</th>
<th>2.8</th>
<th>3.0</th>
<th>3.2</th>
<th>3.4</th>
<th>3.6</th>
<th>4.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.8</td>
<td>324,428</td>
<td>324,420</td>
<td>324,416</td>
<td>324,418</td>
<td>324,425</td>
<td>324,438</td>
<td>324,458</td>
<td>324,519</td>
</tr>
<tr>
<td>4.0</td>
<td>324,421</td>
<td>324,412</td>
<td>324,408</td>
<td>324,409</td>
<td>324,415</td>
<td>324,428</td>
<td>324,448</td>
<td>324,508</td>
</tr>
<tr>
<td>4.2</td>
<td>324,418</td>
<td>324,409</td>
<td>324,404</td>
<td>324,404</td>
<td>324,409</td>
<td>324,410</td>
<td>324,422</td>
<td>324,441</td>
</tr>
<tr>
<td>4.4</td>
<td>324,420</td>
<td>324,410</td>
<td>324,404</td>
<td>324,408</td>
<td>324,410</td>
<td>324,422</td>
<td>324,441</td>
<td>324,450</td>
</tr>
<tr>
<td>4.6</td>
<td>324,427</td>
<td>324,416</td>
<td>324,409</td>
<td>324,407</td>
<td>324,411</td>
<td>324,422</td>
<td>324,444</td>
<td>324,497</td>
</tr>
</tbody>
</table>

TABLE I2 Cross-validation performance of different numbers of nodes in hidden layers

<table>
<thead>
<tr>
<th>$H_1$, $H_2$</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>325,148</td>
<td>324,960</td>
<td>325,269</td>
<td>325,015</td>
</tr>
<tr>
<td>10</td>
<td>324,831</td>
<td>324,972</td>
<td>324,705</td>
<td>324,983</td>
</tr>
<tr>
<td>15</td>
<td>324,825</td>
<td>324,948</td>
<td>325,060</td>
<td>325,012</td>
</tr>
<tr>
<td>20</td>
<td>324,920</td>
<td>324,819</td>
<td>324,745</td>
<td>325,082</td>
</tr>
<tr>
<td>25</td>
<td>325,012</td>
<td>324,877</td>
<td>324,916</td>
<td>325,062</td>
</tr>
</tbody>
</table>

of hidden layer 1 according to

$$h_j^{(2)} = F \left( \sum_{i=1}^{H_1} w_{ij}^{(2)} h_i^{(1)} + b_j^{(2)} \right),$$

where $w_{ij}^{(2)}$ and $b_j^{(2)}$ are parameters. The output utility $U$ of this neural network depends on the values of hidden layer 2 and can be calculated by $U = \sum_{i=1}^{H_2} w_{ij}^{(3)} h_i^{(2)}$.

Table I2 presents the cross-validation performance (negative log-likelihood value) of different $(H_1, H_2)$ combinations. Each row of Table I2 lists the results for a fixed $H_1$ and each column displays the results for a fixed $H_2$. In conclusion, the best $(H_1, H_2)$ combination is (10, 15).

**APPENDIX J: DATA-FILTERING PROCESS IN THE ORIGINAL TEST SET**

Before performance evaluation, the test set data were filtered to make the evaluation process more rigorous. If this filtering process was skipped, our model’s prediction performance could be overestimated, as not all items in the test set were “new” to the market. It also involved some items that hit the market at the end of season $s = 3$ and continued to be sold during season $s = 4$. These items might already be in the training set, and thus should not be used to measure our model’s predictive power for “new items.” We now briefly illustrate the data-filtering process. Assume the test set contains consumer group $G_g$, with conditional choice sets $X_g = (x_{g1}, x_{g2}, x_{g3})$ and consumer choice observations $Q_g = (q_{g1}, q_{g2}, q_{g3})$. If the item corresponding to $x_{g1}$ is also sold during season $s = 3$, then the filtered data are $X_g = (x_{g2}, x_{g3})$ and $Q_g = (q_{g2}, q_{g3})$.

Furthermore, considering that the observed consumer choice proportion $P_g$ contains too much randomness when consumer group $g$ only has a few consumers (i.e., $q_g < 10$), we artificially eliminated those groups with fewer than 10 consumer purchase observations.