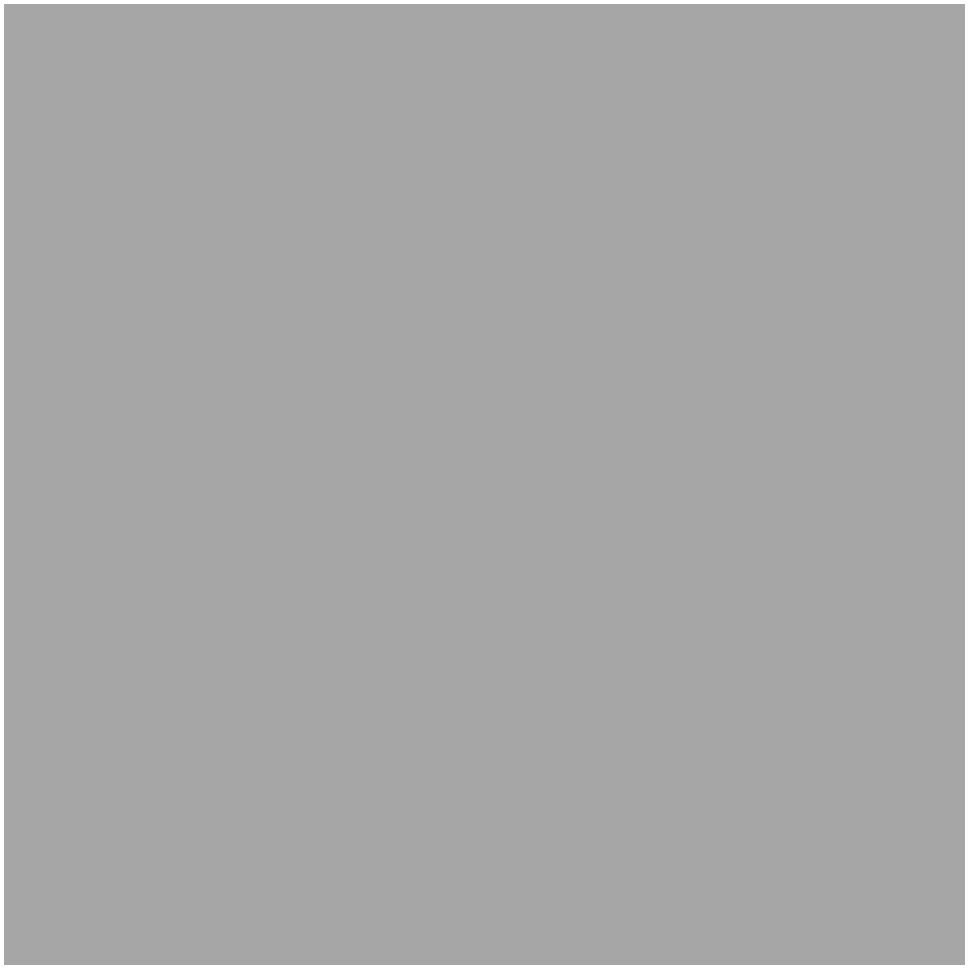


LEVERAGING THE POWER OF IMAGES IN MANAGING PRODUCT RETURN RATES

Daria Dzyabura
Siham El Kihal
John R. Hauser
Marat Ibragimov

Working Paper
No 259

NES Working
Paper series

September
2019

Leveraging the Power of Images in Managing Product Return Rates

Daria Dzyabura*

New Economic School, Moscow, Russia, ddzyabura@nes.ru

Siham El Kihal

Frankfurt School of Finance & Management, Germany, s.elkihal@fs.de

John R. Hauser

MIT Sloan School of Management, USA, hauser@mit.edu

Marat Ibragimov

MIT Sloan School of Management, USA, mibragim@mit.edu

September 3, 2019

*Authors listed alphabetically

This paper has benefited from feedback obtained from Marketing Effectiveness through Customer Journeys and Multichannel Management, Bologna, June 2019, the JAMS Thought Leaders' Conference on Innovating in the Digital Economy, Bocconi University, June 2019, the 11th Triennial Invitational Choice Symposium, Cambridge, MD, June 2019, the 48th Conference of the European Marketing Academy (EMAC), University of Hamburg, June 2019, the Theory + Practice Conference, Columbia University, NY, May 2019, the Annual Marketing Research Camp, Tuck School of Business, Dartmouth, May 2019, the 47th Conference of the European Marketing Academy (EMAC), University of Strathclyde, Glasgow UK, the Special Session on Machine Learning, 40th INFORMS Marketing Science Conference, Fox School of Business, Temple University, USA, the Workshop on Multi-Armed Bandits and Learning Algorithms, Rotterdam School of Management at Erasmus University, Rotterdam, May 2018, the Washington University Foster School of Business Seminar Series, February 2018, and the NYU 2017 Conference on Digital, Mobile Marketing, and Social Media Analytics, NYU Stern School of Business, New York. We also wish to thank Martin Artz, Katrijn Gielens, and Katharina Hombach for comments and suggestions.

Leveraging the Power of Images in Managing Product Return Rates

Abstract

In online channels, products are returned at high rates. Shipping, processing, and refurbishing are so costly that a retailer's profit is extremely sensitive to return rates. In many product categories, such as the $500 billion fashion industry, direct experiments are not feasible because the fashion season is over before sufficient data are observed. We show that predicting return rates prior to product launch enhances profit substantially. Using data from a large European retailer (over 1.5 million transactions for about 4,500 fashion items), we demonstrate that machine-learning methods applied to product images enhance predictive ability relative to the retailer's benchmark (category, seasonality, price, and color labels). Custom image-processing features (RGB color histograms, Gabor filters) capture color and patterns to improve predictions, but deep-learning features improve predictions significantly more. Deep learning appears to capture color-pattern-shape and other intangibles associated with high return rates for apparel. We derive an optimal policy for launch decisions that takes prediction uncertainty into account. The optimal deep-learning-based policy improves profits, achieving 40% of the improvement that would be achievable with perfect information. We show that the retailer could further enhance predictive ability and profits if it could observe the discrepancy in online and offline sales.

Keywords: machine learning, image processing, product returns

1. Introduction

Maria needed a fashionable dress for her son's wedding in Germany. She visited the website of a major fashion retailer and focused on two dresses. One was very simple and she was confident it would meet her needs, if not, there were many places she could wear it. Another dress was colorful with innovative patterns—quite fashionable, exactly why she used the website to find a dress. But she was not sure the fashionable dress would look good on her. She ordered both and returned the second dress. The return was simple and costless to Maria, but very costly to the firm.

Consider the retailer's dilemma. The retailer knows that some apparel items (products) have vastly higher return rates than others in the online channel. In our data, return rates for women's apparel range from 10% to 96% for different items. The retailer wants sales, but would like to avoid costly returns. In an ideal world, the retailer would "test" items in its offline stores where returns are rare and low cost, but in a fashion category, the fashion season would be over before the retailer would have enough offline experience to make decisions. The retailer might exploit historical data from previous years—returns might be more prevalent at various times of year, some categories of apparel might be returned more often, and some colors might be particularly hard to evaluate online. But fashion is more complicated; designers and consumers make holistic judgments. Absent the ability of consumers to experience the physical product online, the retailer may wish to use the high-dimensional information in images to launch items with profitable return rates.

Managing returns depends on predicting return rates. We demonstrate that advanced machine learning, deep learning in particular, helps the retailer use images (of apparel on its website) to manage returns better. Using a large data set from a European apparel retailer (over 1.5 million online and offline transactions involving about 4,500 unique fashion items), we demonstrate that advanced models of color (RGB color histograms) and patterns (Gabor filters) help, but deep learning of "features" is even better. We derive optimal policies that use the (still noisy) predictions to enable the retailer to manage

its online channel for increased profit. The increase in profit is substantial relative to historical category, seasonality, price, and color data and achieves 40% of what the retailer would realize if it had perfect prior information on return rates.

2. Online and Offline Retail and Product Returns

Managing returns is of broad interest beyond our illustration with a European retailer. Online channels have many advantages alone and as complements to offline channels including broader reach, lower travel costs for consumers, and saved costs of renting and operating retail space. However, the cost of processing product returns is a major drain on profitability. The founder of the UK's largest fashion retailer ASOS stated that a 1% drop in the retailer's return rate could increase the retailer's bottom line by 30% (Thomasson 2013). Even large online retailers such as Amazon struggle with managing product returns (Safdar and Stevens 2018).

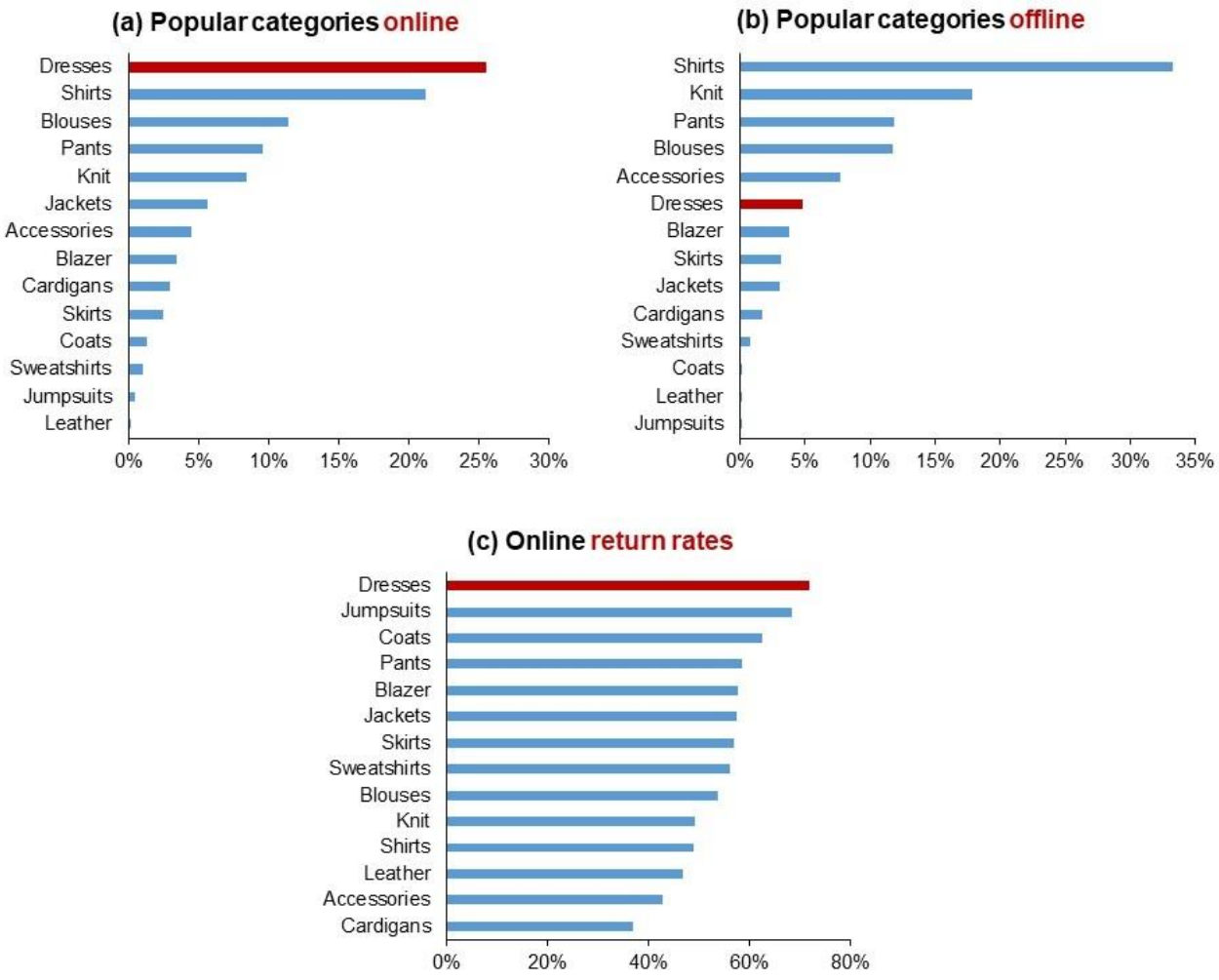
Product returns are vastly more common online than offline. In our data, for the same items, the average return rate per item is 56% online compared to 3% offline. These data should not surprise us. In an offline channel, the consumer can examine an item, feel its texture, see its colors and patterns, and try it on to observe both fit and match. In an online channel, consumers make item choice based primarily of the image.

The retailer's return costs differ dramatically by channel. Offline, the customer brings the item back to the store; a sales associate evaluates its condition and processes the return. Online, the retailer pays shipping and pays an employee to open the box, evaluate the item, and issue a refund. The retailer often needs to refurbish the items and/or send them to be discounted in outlet stores—some are even discarded. The resulting costs range between $6 and $20 per returned item (The Economist 2013). A small difference in return rates can make or break a retailer.

Returns are exacerbated when returns are low cost to consumers, as in the European Union. Consumers may need to evaluate actual colors, patterns, and the touch and feel of the item through

physical inspection (Dzyabura et al. 2019). If these attributes do not match the consumer's taste, the consumer may choose to return the item. We hypothesize that, if an item sells well online (based mostly on its image), but sells poorly offline (based more on its look, feel, and/or fit), then the difference in relative sales online versus offline (online/offline discrepancy) for an item is correlated with return rates. Figure 1 motivates this hypothesis: the best-selling online category is dresses (27% of sales), but dresses is the sixth best offline category (8% of sales). Dresses have the highest online/offline discrepancy and are returned at the highest rate in the online channel (72%). For Figure 1, the correlation is high (0.52, p < 0.05), but our hypothesis is at the item level. The item-level correlation is modest (0.14), but significant (p < 0.05).

Figure 1. Online and Offline Performance (Share of Sales) and Online Return Rates by Category



Online/offline discrepancy is also motivated when online purchases have a greater search component than offline purchases (Anderson et al. 2009). To the extent that consumers treat online purchases as search, it is optimal for consumers to search more if the variance is higher (e.g., Weitzman 1979). Higher (online) variance implies that more items are rejected and returned because the items do not meet expectations. Thus, higher variance implies both higher online sales (search) and higher online returns.

When online/offline discrepancy can often be observed with sufficiently high precision in advance of a retailer's ability to observe return rates, we can use the online/offline discrepancy as a potential early indicator of return rates. In some product categories, online/offline discrepancy may be available well before returns rates are observed. For our retailer, the fashion season is over before the retailer can observe either online/offline discrepancy or return rates with sufficient precision to modify its online offerings. Nonetheless, in §5.4, we assume availability and test whether online/offline discrepancy augments and/or can substitute for image-based predictions.

3. Related Literature

3.1. Leveraging Unstructured Data

In the online channel, especially for our fashion retailer, the primary stimulus is an image of the item. We examine whether information contained in the image improves return management beyond the data that is otherwise available to the retailer. But images are inherently high-dimensional; a three-color image can contain up to millions of pixels depending on the resolution. Because we deal directly with the images, we rely on machine learning methods.

We build on methods now being used successfully in marketing science. For example, Timoshenko and Hauser (2019) use convolutional neural networks to extract customer needs from review texts and Liu and Toubia (2018) infer consumer preferences based on search query texts. Archak et al. (2011), Chevalier and Mayzlin (2006), Lee and Bradlow (2011), Netzer et al. (2012), Onishi and

Manchanda (2012), Tirunillai and Tellis (2012), and Toubia and Netzer (2017) use text-based user-generated content to predict or evaluate demand, financial performance, consumer engagement, market structure, and creative ideas. He and McAuley (2016), Lynch et al. (2015), and McAuley et al. (2015) use images to create recommendations regarding clothing styles, substitutes, and more accurate personalized rankings. Liu et al. (2018) use images posted by consumers to evaluate how brands are portrayed in social media; Zhang and Luo (2018) use images and text from Yelp reviews to predict restaurants' survival; and Zwebner et al. (2017) demonstrate that it is feasible to predict a person's name based on an image of their face.

We adopt best practices in the analysis of unstructured data (images) to predict return rates prior to launch and use these predictions in our policy analysis.

3.2. Product Returns

The literature on product returns focuses either on the impact of return policies such as fees or deadlines on purchases and on return rates (e.g., Shulman et al. 2011; see Janakiraman et al. 2016 for a review) or on understanding and managing the product return behavior of individual customers (El Kihal, Erdem, and Schulze 2019; Narang and Shankar 2019; Nasr-Bechwati and Schreiner 2005). By contrast, our research focuses on the selection of which items to place in the online channel. Our research complements existing research on return policies and on managing customers.

3.3. Online and Offline Channels as Complements

Much of the literature on online/offline retail channels examines spillovers or cannibalization. For example, Pauwels and Neslin (2015) demonstrate that introducing a physical store in a geographic region cannibalizes catalogue sales but has less impact on internet sales. Wang and Goldfarb (2017) find that the presence of a physical store increases customer acquisition in online channels. Ansari et al. (2008) examine customer channel migration and its impact on channel selection and demand. Dzyabura et al. (2019) show large differences between how consumers evaluate the same individual product

online versus offline. We focus on the use of images to manage online returns, and in §5.4 we examine whether early observations of online/offline discrepancy improves the prediction of online returns.

4. Predicting Returns Using Category, Seasonality, Price, and Color

4.1. Data

The retailer, which specializes in women's apparel, has a network of 39 retail stores in Germany complemented by a large online operation that accounts for 22% of its sales. We use data on 1,510,083 transactions, including sales and returns, that occurred in online and offline channels during the observation period (2014–2016). We exclude non-fashion items such as perfume or gift cards and truncate by date to exclude purchases for which the return deadline fell outside our observation period. To ensure reliable estimates of the return rates, we focus on items that were sold at least twenty times. The resulting data set consists of 4,585 distinct items from fourteen different product categories—categories used by the retailer to manage its sales.

The retailer has a lenient return policy: customers can return any purchased item for any reason within 14 days. This is the case for all retailers in the European as mandated by the law. The average return rate for items purchased through the online channel is 56% (ranging from 10–96%). It is only 3% for items purchased through the offline channel. Offline return costs are well below online return costs.

Before we explore the use of images to manage returns, we explore benchmark return-rate predictions that use information routinely collected by the retailer. We then examine whether traditional machine-learning methods with hand-crafted image-based features improve predictions and whether we can improve predictions further with deep-learning features. Armed with prediction models, we explore managerial policy. In §5.3, we examine the robustness of our predictions to modeling decisions.

4.2. Baseline Predictions (Item's Category, Seasonality, and Price)

For each item in its inventory, the retailer observes the time of year (by month), the product category (e.g., dresses), and price. Because this is fashion, seasonality is clearly important. We also

expect return rates to vary by product category. For example, Figure 1 shows that, of the fourteen categories tracked by the retailer, dresses are returned on average about 70% of the time while cardigans are returned about 35% of the time. For the purposes of this analysis, we treat price as exogenous to the decision on whether or not to post an item online. Our data do not contain sufficient information on the demand curve to optimize price. If we can improve profitability when price is exogenous, future research with improved data could include price optimization in policies to improve profitability further.

We can choose a variety of prediction methods with which to predict return rates. These methods vary from simple regression to highly non-linear functions obtained with machine learning. In our data, we obtain the best predictive ability, using gradient boosted regression trees (GBRTs). We report in the appendix prediction results using bagging methods (random forest) and LASSO.(for details , refer to appendix A, Table A.5).

Like a regression tree, GBRT partitions the space of explanatory variables into multiple regions and predicts a value for all points in a region. The advantage of tree-based models is their ability to capture higher-order interactions among features. To avoid exploiting random variation, we regularize the tree by limiting the number of splits. A random forest generalizes regression trees by using a set of regression trees, each trained on a bootstrapped subsample of the original data. A GBRT further generalizes random forests by boosting each tree greedily based on the residual of the current model. We use LightGBM (Guolin et al. 2017) based on its performance in Kaggle machine learning competitions. This algorithm, having comparable accuracy with the alternative algorithms (for example, XGBOOST by Chen and Guestrin (2016)), converges significantly faster. GBRTs have performed well in marketing applications such as predicting clickstreams (Rafieian and Yoganarasimhan 2018) and predicting customer churn (Neslin et al. 2006).

Table 1 reports the predictive ability of the baseline model. We evaluate the predictions with

twenty-fold cross-validation. For each of twenty draws, we randomly select 75% of the data to train the model, validate the model on 20% of the sample (optimized over a set of hyperparameters), and compare predictions on the remaining 5% of the data. Overall performance is the weighted average over all twenty draws. We report the *out-of-sample* R-squared (predictive accuracy) calculated based on all twenty draws (K_{all}):

$$R^2_{model} = 1 - \frac{\sum_{i \in K_{all}} \left(r_i - \hat{r}_i^{model}\right)^2}{\sum_{i \in K_{all}} \left(r_i - \hat{r}_i^{random}\right)^2}, \quad (1)$$

where $\hat{r}_i^{model}$ is the predicted return rate of item *i* according to the *model* and $\hat{r}_i^{random}$ is the prediction we obtain with the *random* model which predicts the average return rate for all items. To address the variance in the estimated R^2_{model} we generated 25 different sets of cross-validation folds; we report the standard deviations of the estimated R^2_{model}.

4.3. Benchmark Predictions (Baseline Plus Color Labels)

In apparel, color clearly matters. Consumers can more easily imagine themselves in common conservative colors such as blacks, blues, and greys, but often want to try fashion colors such as pinks and purples. Indeed, in our data (after controlling for category, seasonality, and price), pinks are returned substantially more often than blacks. The retailer provides color labels (thirteen categories) for each item of apparel. Although the color labels are not perfect, for example there are many shades of pink and some items have multi-colored patterns or highlights, we expect that predictions improve if we include color categories. Table 1 shows that color labels improve predictions slightly. We adopt a model with category, seasonality, price, and color labels as our benchmark because it represents the best set of explanatory variables currently available to the retailer. In this way, we isolate the incremental advantage of using machine learning to extract information directly from images.

Table 1. Baseline and Benchmark Predictions

Model	Product Features Included	Image Features Included	Predictive Accuracy, Out of Sample R^2 (standard dev)	Improvement over the benchmark
Baseline	Category, seasonality, price	None	41.3 (0.18)	-2.8%
Color-based Benchmark	Category, seasonality, price, and color labels	None	42.5 (0.20)	0.0%

5. Incorporating Images

Images are more than just color. Consider the three items in Table 2. The first item, the white top, is easily categorized and a common color; the benchmark model does well. The second item, the top with stripes, is multicolored and hard to categorize by color; the benchmark model does less well. The third item, the dress, is readily categorized as pink, but the benchmark model does not do well, likely because the pink is not a prototypical pink and because the dress's shape does not work well for everyone.

Table 2. Return Rates and Benchmark Predictions for Three Apparel Items

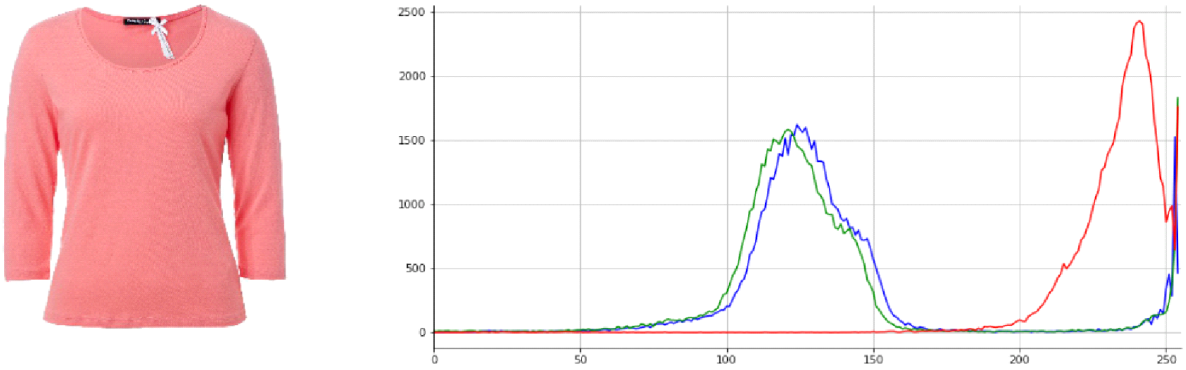
			
Actual Return Rate	50.0%	35.0%	82.2%
Benchmark Prediction	49.0%	47.0%	67.0%

To improve upon the color-based benchmark, we begin with commonly used image-encoding features before turning to deep-learning features. These features are well-suited to represent diverse colors, including multiple colors, and represent periodic patterns such as stripes. We demonstrate in §5.2 that these features improve predictive ability, but that deep-learning encodes images and predicts return rates even better for the hard-to-predict items (multicolored top and pink dress).

5.1. RGB Color Histograms and Gabor Features

RGB color histograms. Images are composed of pixels and pixels are composed of red, green, and blue (RGB) channels, each represented by an integer (0 to 255) to signify the intensity of each color. For example, a standard medium-quality image with 600 x 600 = 360,000 pixels is represented by 1,080,000 numbers. An RGB color histogram counts the number of pixels of a given intensity for each of the RGB channels. Figure 2 provides an example. Because 256 x 256 x 256 ≈ 16 million, there are too many RGB "bins" for a prediction model. For feasibility, we split the three-dimensional histogram into fewer RGB bins—in our case, we use 5 x 5 x 5 = 125 bins. RGB color histograms are a much finer categorization of an item's color than the retailer's thirteen color labels. Hopefully, RGB color histograms capture the more-exact shade of pink of the dress in Table 2.

Figure 2. Example RGB Color Histogram Encoding of an Apparel Item



Gabor-filter features. The multicolored top in the middle of Table 2 contains stripes in three colors. While the three colors might be captured by an RGB color histogram, the horizontal stripes

represent a periodic pattern. Gabor filters provide a proven way to capture patterns by isolating periodicity and the direction of that periodicity (Manjunath and Ma 1996). For example, a set of Gabor filters can capture horizontal stripes, vertical stripes, or even an asymmetric checkered pattern. In particular, Gabor filters use a set of superimposed sinusoidal waves of different frequencies to capture patterns. Similar to Liu et al. (2018), we implement Gabor filters by applying the following transformations to the apparel image. If *x* and *y* are the pixel coordinates of the image, the Gabor filter is a function of frequency (λ) and direction (θ). The Gabor filter includes a Gaussian smoothing filter with parameter (σ).

(2) $$g(x, y \mid \lambda, \theta, \sigma) = e^{-\frac{\tilde{x}^2+\tilde{y}^2}{2\sigma^2}} \cos\left(\frac{2\pi\tilde{x}}{\lambda}\right),$$

$$\begin{pmatrix}\tilde{x}\\ \tilde{y}\end{pmatrix} = \begin{pmatrix}\cos\theta & \sin\theta\\ -\sin\theta & \cos\theta\end{pmatrix}\begin{pmatrix}x\\ y\end{pmatrix}$$

For example, a checkered pattern is represented by two Gabor filters, one with $\theta = 0^o$ and one with $\theta = 90^o$. The frequency parameter(s), λ, determine the width of each dimension of the checkered pattern.

We use B filters where each filter, b, is described by a set of values for $(\lambda_b, \theta_b, \sigma_b)$ and apply these filters to greyscale versions of our images. For each b, the output of the filter is a greyscale image where bright segments correspond to the segments of the original with patterns close to the frequency and orientation parameters of the filter. For each b, we compute the mean and variance of the intensity of the corresponding greyscale image. The set of Gabor features is the set, for all B filters, of these means and variances.

Table 3 reports the incremental predictive ability of models based on RGB color histograms, Gabor filters, and RGB color histograms plus Gabor features. For comparison, we repeat the benchmark model from Table 2. RGB color histograms are a better representation of color than color labels (predictions improve by 3.8%) and modeling patterns with Gabor features improves predictions further

(6.6%; significantly more than the benchmark, $p < .01$).[1]

Table 3. Predictions Using Image-Based Features (RGB/Gabor versus Deep Learning)

Model	Product Features Included	Image Features Included	Predictive Accuracy, Out of Sample R^2 (standard dev)	Improvement over the Benchmark
Color-based Benchmark	Category, seasonality, price, and color labels	None	42.5 (0.20)	0.0%
Color Features	Category, seasonality, price, color labels	RGB	44.1 (0.21)	+3.8%
Color and Patterns	Category, seasonality, price, color labels	RGB + Gabor	45.3 (0.18)	+6.6%
CNN Features	Category, seasonality, price, color labels	Deep-learning	46.9 (0.15)	+10.4%
Test of All Features	Category, seasonality, price, color labels	RGB + Gabor + Deep-learning	46.9 (0.25)	+10.4%

5.2. Deep-Learning Features

We improved predictions with hand-crafted representations of images, but we can do even better with deep-learning features because deep learning is particularly suited to representing images. Images are unstructured and high-dimensional. While color and (periodic) patterns clearly help, apparel images are often more complex. For example, deep-learning features might capture the shape of the dress in Table 2. Other dresses might feature floral patterns or complex geometric shapes that are hard to represent with Gabor filters. Deep-learning algorithms have the advantage that they learn feature representations automatically and can be tuned to particular applications. To explore the potential of deep learning for image-based predictions of apparel return rates, we tune an established convolutional

[1] Gabor filters alone improve predictions relative to the benchmark more than RGB features (4.5%), but in either case, the combination of Gabor filters and RGB features leads to incremental improvement over either Gabor filters or RGB features.

neural network (CNN). Each layer of the CNN transforms the features of the previous layer to obtain a new set of features. Through a series of simple nonlinear transformations, the CNN learns highly complex nonlinear transformations to map an image to a set of features. For greater detail on each transformation and for an application of a CNN to unstructured marketing data, see Zhang and Luo (2018).

Our 4,500+ images are sufficient to tune an established CNN, but not to train a deep CNN from scratch, thus we use the next-to-last pre-output layer of the Residual Neural Network (ResNet; He et al. 2015). ResNet won the 2015 ImageNet Large Scale Visual Recognition Challenge and was trained on the ImageNet data set (1.3 million images in roughly 1,000 categories). The ResNet network has 152 layers, making it the deepest network yet presented on ImageNet. The last layer is the output layer. We use the second-to-last layer of the network, which contains 2,048 features.

Our goals are (1) to demonstrate the power of a deep representation of apparel images, (2) to test whether the deep representation improves on machine-learning color and pattern features, and (3) examine whether predictive ability translates into enhanced profitability. To the extent we succeed with a tuned, pre-trained CNN, we provide a lower bound on what can be achieved with a custom deep learning model.

In Table 3, we see substantial and significant improvement with deep-learning features relative to color and patterns features—both models include benchmark features. We also see that there is no significant improvement when we add color/pattern features to the deep-learning features—the deep-learning features are sufficient. Returning to the images in Table 2, models based on color/pattern features and models based on deep-learning features predict return rates better for the hard-to-predict tops, with deep-learning features predicting best. Deep-learning features predict a return rate of 39.7% for the striped top (vs. a true rate of 35.0%, a color/pattern rate of 45.6%, and a benchmark rate of 48.6%,) and a return rate of 76.6% for the pink dress (vs. a true rate of 82.2%, a color/pattern rate of

72.8%, and a benchmark rate of 69.0%). Deep representations enhance predictive ability and do so above and beyond standard color and pattern machine-learning features. Before we turn to the implications of improving predictions for retailers' profitability, we examine robustness.

5.3. Robustness Tests

The robustness tests suggest that the basic insights of Table 3 are not sensitive to many of our specific modeling decisions. Deep-learning features appear predict best even when we relax assumptions.

Minimum sales. We screened the data to require each item to have been sold at least twenty times. We obtain that same insights if we screen the data to require minimum sales of 10 or 30. See Table A.1 in appendix A.

Alternative "hand-crafted" features. We tested (1) HSV (hue, saturation, value) color histograms and (2) local attributes of images (ORB features, see Rublee et al. 2011). The alternative features improve predictions relative to the benchmark, but not quite as much as RGB color histograms. See Table A.2 in appendix A.

Dimensionality reduction. We tested (nonlinear) principal components analysis (PCA) to reduce the dimensionality of all features. We obtain the same basic insights that RGB features improve on color labels and deep-learning features improve still further. However, the information in Gabor filters does not seem to be compatible with PCA. See Table A.3 in appendix A.

Alternative CNN. To test the robustness of our results, we used the output from the second-to-last layer of an alternative pre-trained CNN, the VGG-19 network (Simonyan and Zisserman 2014). The VGG-19 performance is slightly worse, but not statistically so. See Table A.4 in appendix A.

Alternative predictive models. As discussed in §4.2, the results we achieve using GBRT are better than predictive models such as bagging methods and LASSO for the same features. See Table A.5 in appendix A.

5.4. Online/Offline Discrepancy

In §2, we considered how we might extend an image-based model when the fashion season was sufficiently long to observe online and offline shares of sales with adequate precision. We hypothesized that online/offline discrepancy could be used as an early indicator of item-based return rates (the former can be observed much earlier than the latter). Although the fashion season is not sufficiently long, we can examine this hypothesis using early observations of online/offline discrepancy. (We used two weeks, but the insights are not contingent on the exact number of weeks.)

When we add online/offline discrepancy to our best model (CNN features, color labels, category, seasonality, and price), the out-of-sample R^2 improves significantly from 46.9 to 48.7 raising to 14.6% above the benchmark—an incremental improvement of 4.2%.[2] Thus, there is information in discrepancy above and beyond that in image-based CNN features. We also address whether image-based features add information beyond online/offline discrepancy. They do. The improvement is statistically significant—an additional 5.4%. We provide details in Table A.6 in appendix A.

6. Using Deep-Learning Features to Enhance Profit

6.1. Profit Depends on the Return Rate

The return rate for the pink dress in Table 2 is 82.2%. For that dress, the costs of returns likely exceed any profits earned from the sales of non-returned pink dresses (17.8% of those sold). On the other hand, the striped top is returned 35.0% of the time. Profits earned from the sales of non-returned striped tops (65.0%) are likely to exceed the costs for the 35.0% that are returned. Perhaps it would be best if the retailer had never launched the pink dress in the online channel.

The retailer's costs for returned items are driven by a fixed return cost (shipping and handling, c_{fix}) and by a cost that is proportional to price (items which must be discounted or discarded because

[2] Because average offline sales exceed average online sales by over three to one, our measure of discrepancy is normalized. Alternative combinations, including online sales and offline sales alone, add information relative to image-based features. We explicitly did not search over all possible measures of discrepancy to avoid overfitting the model.

they are damaged or out of season, c_{var}). Let p_i be the price of item i and c_i be the purchase cost of the item. If we had perfect information on return rates, r_i, we would compute profits, π_i, per item i by:

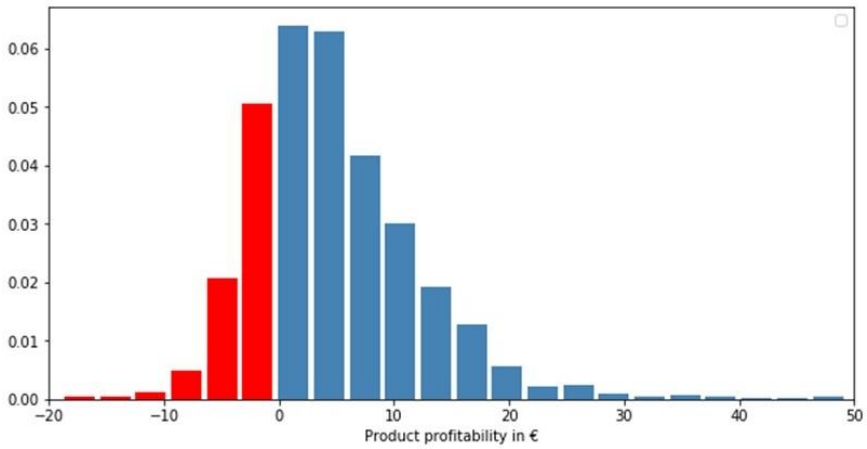
(3) $$\pi_i = (1 - r_i)(p_i - c_i) - r_i(c_{fix} + p_i c_{var}),$$

Equation 3 is a linear function of the return rate, hence, by rearranging the terms, we obtain a simplified expression for the profit per item ($\mathcal{R}_i$and $\mathcal{C}_i$ are defined implicitly by Equation 4):

(4) $$\pi_i = (p_i - c_i) - r_i(p_i - c_i + c_{fix} + p_i c_{var}) \equiv \mathcal{R}_i - r_i \cdot \mathcal{C}_i$$

Figure 3 illustrates the results of applying Equation 3 to our data; 26.4% of the items are unprofitable, although some are more unprofitable than others.[3] If the retailer had perfect predictions of return rates, it would not launch items with negative profitability (the red bars in the distribution). It might ask its designers to create new designs, convert them to images, and retest until only profitable items were launched. (We assume the designers are sufficiently skilled to maintain the right variety of items.)

Figure 3. Distribution of Items' Profitability in Our Data



Item-by-item decisions are feasible in the online channel, but less feasible in the offline channel. Replacing unprofitable items with profitable items is unlikely to affect the overall image presented by the

[1] The retailer's costs are proprietary. To illustrate the analysis, we used a fixed return cost of 5€ (iBusiness 2016) and a variable return cost 13.1% of an item's price (Asdecker 2015).

website. Furthermore, the consumer tends to view one or a few items in detail while scrolling through many items. Interactions in the online channel are much less substantial than in the offline channel.

The offline assortment decision differs dramatically because the offline consumer experiences the entire store layout. Offline stores rely on buyers to arrange colors or items together for an overall shopping experience. Some items are displayed to help the consumer visualize how she would complete an outfit—a manikin might display a red blouse with designer jeans in the hopes of selling the jeans rather than the blouse. Stores are laid out so that the consumer can make decisions on entire ensembles. Thus, while it is reasonable to assume independence among items in an online policy (as in this paper), it would be difficult to do so in an offline policy.

6.2. Optimal Policy

We seek an online policy that is optimal in light of the uncertainty in the predicted return rates. We assume that, based on substantial experience in the apparel industry, the retailer has prior beliefs on the return rate, and implied profitability, for the set of items on its website. We assume the prior beliefs about profit are normally distributed across items: $\pi \sim \mathcal{N}(\mu_o;\ \sigma_o^2)$. For modeling purposes, we assume that our estimate of profitability, $\hat{\pi}$, has a mean equal to the true profit with a variance based on the out-of-sample MSE of the model: $\hat{\pi} \mid \pi \sim \mathcal{N}(\pi;\ \sigma_1^2)$. Let $\mathcal{P}$ be a policy such that the retailer launches the item if $\mathcal{P} = 1$ and does not launch the item if $\mathcal{P} = 0$. Let $\phi \equiv (\hat{\pi}, \mu_0, \sigma_0^2, \sigma_1^2)$, then we solve the following optimization problem:

$$\max_{\mathcal{P}(\phi)\in[0,1]} \mathbb{E}[\mathcal{P}(\phi) * \pi + \big(1 - \mathcal{P}(\phi)\big) * 0], \tag{5}$$

In the appendix, we demonstrate that the solution to the optimization problem in Equation 5 is the following threshold-based policy:

$$\mathcal{P}(\phi) = \begin{cases} 1 & if\ \hat{\pi} \geq -\mu_o \frac{\sigma_1^2}{\sigma_o^2} \\ 0 & if\ \hat{\pi} < -\mu_o \frac{\sigma_1^2}{\sigma_o^2} \end{cases} \tag{6}$$

The policy is intuitive. For example,

- If predictions are perfect, then $\sigma_1^2 = 0$ and the policy reverts to that of perfect prediction; launch those items for which $\hat{\pi} \geq 0$.
- If the model has no predictive ability, then $\sigma_1^2 \to \infty$ and the policy reverts to the prior mean, μ_o; launch all items if and only if the prior mean is positive.
- If there is no uncertainty in the prior, then $\sigma_o^2 \to 0$ and the policy again reverts to the prior mean; launch all items if and only if the prior mean is positive.
- For finite values of σ_1^2 and σ_o^2, the ratio, σ_1^2/σ_o^2, modifies the amount by which the predicted profits must exceed prior beliefs in order to launch.

Assuming that the retailer has positive priors, Figure 4 illustrates the optimal policy. (1) For perfect predictions ($\sigma_1^2 = 0$), launch all items predicted to be profitable. (2) For good predictions (σ_1^2 small), launch most items predicted to be profitable. And (3), when predictions are extremely noisy (σ_1^2 large), launch almost all items. Figure 4 suggests that we are likely to screen out more items for the better-predicting deep-learning-based model than we are for the benchmark model.

6.3 Policy Simulations

Perfect-prediction policy. If there were no uncertainty in predictions, the retailer would launch all profitable items and not launch the 26.4% of items that were not profitable. The perfect-prediction policy increases profits by 21.2%.

Deep-learning-based policy. Our best predictive model is the GBRT analysis applied to deep-learning features, but the R^2 of 46.9% is far from perfect. When we apply the optimal policy, we choose not to launch 8.4% of the items, since these items are the items most likely to be unprofitable. The resulting expected profits increase by 8.5% relative to a policy, which launches all the items. This is 40% of that achievable with perfect prediction (See Table 4). Policies based on RGB-histograms/Gabor-filters and based on benchmark model also do well, but not as well as the best predictive model.

Figure 4. Graphical Illustration of the Optimal Policy

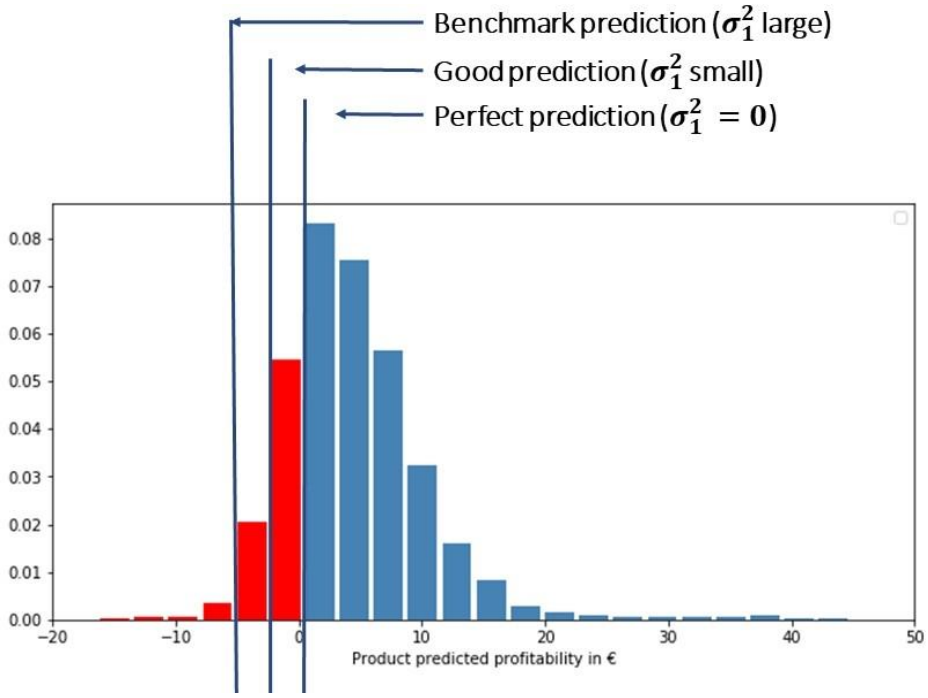


Table 4. Expected Profit Improvement Using Deep-Learning Policy

Policy	Estimate of Return Rate	Percent Items not Launched	Profit Improvement
Deep-learning-based policy	GBRT based on deep-learning features	8.4%	8.5% (0.21)
Perfect-prediction policy	Perfect knowledge of the true return rate	26.4%	21.2% —

To substitute or not. Our results likely apply whether or not the retailer has the option to substitute new items for those items it chooses not to launch. To the extent that the substitute items are available, the retailer's profit would be greater than in Table 4. However, it is reasonable to assume that profits scale proportionally for every predictive model when substitute items are drawn from the same distribution. Under this assumption, the substitute items do not affect relative comparisons. Alternatively, if the retailer cannot launch substitute items, we expect the expected relative profits to be close to those in Table 4. Previous studies suggest that small to moderate assortment reductions have a

positive impact on store choice (Briesch et al. 2009) and increase sales (Broniarczyk et al.1998), especially in categories with wide variety (Boatwright and Nunes 2001) and a large number of products (Zhang and Krishna 2007). Moderate assortment reductions are more likely if we remove 8.4% (deep-learning-based policy) of the items rather than 26.4% of the items (perfect-prediction policy), thus our results are a conservative estimate of the profitability improvement due to deep-learning features. Furthermore, deep learning identifies the items most likely to be returned. Avoiding launching highly-returned items may positively impact consumer brand perceptions and increase satisfaction.

7. Summary and Discussion

Using data from a large European fashion retailer, we demonstrate that an optimal policy based on apparel images posted in an online channel can be used to enhance profits. Images are effective because (1) profit in the online channel is extremely sensitive to return rates and (2) machine-learning extracts sufficient information from images to enhance predictive ability relative to the retailer's benchmark predictions (category, seasonality, price, and color labels). Standard, custom machine-learning features (RGB color histograms, Gabor [pattern] filters) do well, but deep-learning features do better. If a retailer has sufficient time in a fashion season to use early observations of online/offline discrepancy, it could enhance profit still further.

Our data are from the fashion industry. This industry is important by itself, but we believe that many insights generalize. Returns are important in almost all online channels and we hypothesize that product images can be mined with deep learning to enhance profitability in many product categories. We tuned a well-established deep-learning CNN, but other deep-learning methods might do as well or even better. If so, improved models provide even stronger support for using images to manage returns.

Finally, there are at least two complementary research opportunities. First, online/offline discrepancy is a valuable source of information about returns. We tested the value of estimates from the first two weeks, but we might improve that strategy with optimal experimentation (Gittins et al. 2011)

trading off exploration (longer observation period) versus exploitation (longer use of an optimal policy). Second, the fashion retailer might improve the accuracy of the return-rate predictions and/or reduce return rates by optimizing the images to communicate key aspects of each item to consumers using methods such as those studied in Zhang et al. (2017).

Appendix A: Robustness Tests

Table A.1. Improvement in Predictive Accuracy Varying Minimum Threshold on Online Sales

Model	Product Features Included	Image Features Included	Predictive Accuracy, Out of Sample R^2 (standard dev)	Improvement over the benchmark[4]
CNN Features with **10** as threshold for online sales	Category, seasonality, price, color labels	Deep-learned	46.5 (0.26)	+9.3%
CNN Features with **20** as threshold for online sales	Category, seasonality, price, color labels	Deep-learned	46.9 (0.15)	+10.4%
CNN Features with **30** as threshold for online sales	Category, seasonality, price, color labels	Deep-learned	51.2 (0.17)	+9.5%

Table A.2. Improvement in Predictive Accuracy Using Alternative Image-Feature Extraction Methods

Model	Product Features Included	Image Features Included	Predictive Accuracy, Out of Sample R^2 (standard dev)	Improvement over the benchmark
RGB Features	Category, seasonality, price, color labels	RGB	44.1 (0.21)	+3.8%
HSV Features	Category, seasonality, price, color labels	HSV	43.9 (0.20)	+3.3%
ORB Features	Category, seasonality, price, color labels	ORB	43.4 (0.20)	+2.1%

Table A.3. Improvement in Predictive Accuracy Using Nonlinear PCA

Model	Product Features Included	Image Features Included	Predictive Accuracy, Out of Sample R^2 (standard dev)	Improvement over the Benchmark
Color Features	Category, seasonality, price, color labels	RGB	43.5 (0.18)	+2.4%
Color and Patterns	Category, seasonality, price, color labels	Gabor	41.4 (0.33)	-2.6%
CNN Features	Category, seasonality, price, color labels	Deep-learned	46.6 (0.21)	+9.6%

[4] The improvement is calculated for the benchmarks estimated on the corresponding samples

Table A.4. Improvement in Predictive Accuracy Using an Alternative CNN

Model	Product Features Included	Image Features Included	Predictive Accuracy, Out of Sample R^2 (standard dev)	Improvement over the Benchmark
ResNet CNN (this paper)	Category, seasonality, price, color labels	Deep-learned	46.9 (0.15)	+10.4%
VGG-19 CNN	Category, seasonality, price, color labels	Deep-learned	46.8 (0.18)	+10.1%

Table A.5. Improvement in Predictive Accuracy Using Alternative Prediction Models

Model	Product Features Included	Image Features Included	Predictive Accuracy, Out of Sample R^2 (standard dev)	Improvement over the Benchmark
GBRT (CNN Features)	Category, seasonality, price, color labels	Deep-learned	46.9 (0.15)	+10.4%
Bagging Methods (CNN Features)	Category, seasonality, price, color labels	Deep-learned	45.4 (0.20)	+6.8%
LASSO (CNN Features)	Category, seasonality, price, color labels	Deep-learned	44.1 (0.32)	+3.8%

Table A.6. Improvement in Predictive Accuracy Using Online/Offline Discrepancy

Model	Product Features Included	Image Features Included	Predictive Accuracy, Out of Sample R^2 (standard dev)	Improvement over the Benchmark
Color-based Benchmark	Category, seasonality, price, color labels	None	42.5 (0.20)	0.0%
Adding Online/ Offline Discrepancy	Category, seasonality, price, color labels	None	46.4 (0.22)	+9.2 %
Adding CNN Features and Online/Offline Discrepancy	Category, seasonality, price, color labels	Deep-learned	48.7 (0.22)	+14.6%

Appendix B: Proof that the Optimal Policy is a Threshold Policy

Result 1. Suppose (1) the firm's prior on the profitability of an item, π, is normally distributed, $\pi \sim \mathcal{N}(\mu_0;\ \sigma_0^2)$, (2) the firm observes an estimate of profitability $\hat{\pi}|\pi \sim \mathcal{N}(\pi;\ \sigma_1^2)$, and (3) the firm seeks a policy to decide whether to put an item online or not. Then the profit maximizing policy, $\mathcal{P}(\phi)$, is a threshold policy:

$$\text{(A1)} \qquad \mathcal{P}(\phi) = \begin{cases} 1 & if\ \hat{\pi} \geq -\mu_o \frac{\sigma_1^2}{\sigma_o^2} \\ 0 & if\ \hat{\pi} < -\mu_o \frac{\sigma_1^2}{\sigma_o^2} \end{cases}$$

***Proof**:* The firm solves the following optimization problem:

$$\text{(A2)} \qquad \max_{\mathcal{P}(\phi)\in[0,1]} \mathbb{E}[\mathcal{P}(\phi) * \pi + \big(1 - \mathcal{P}(\phi)\big) * 0] = \max_{\mathcal{P}(\phi)\in[0,1]} \mathbb{E}[\mathcal{P}(\phi) * \pi]$$

where $\phi \equiv (\hat{\pi}, \mu_0, \sigma_0^2, \sigma_1^2)$ is the set of all known parameters; $\hat{\pi}|\pi \sim \mathcal{N}(\pi;\ \sigma_1^2)$ and $\pi \sim \mathcal{N}(\mu_0;\ \sigma_0^2)$

Using the law of iterative expectations, we rewrite the initial maximization problem (A2) as:

$$\text{(A3)} \qquad \max_{\mathcal{P}(\phi)\in[0,1]} \mathbb{E}[\mathcal{P}(\phi) * \pi] = \max_{\mathcal{P}(\phi)\in[0,1]} \mathbb{E}\big[\mathcal{P}(\phi) * \mathbb{E}[\pi|\phi]\big] = \max_{\mathcal{P}(\phi)\in[0,1]} \mathbb{E}\big[\mathcal{P}(\phi) * \mathbb{E}[\pi|\hat{\pi}]\big]$$

The last step relies on the assumption that $\sigma_0, \sigma_1, \mu_0$ are observable.

Because $\mathbb{E}[\pi|\phi]$ is a function of observables, ϕ, we can denote $\mathbb{E}[\pi|\phi] = f(\phi)$. Equation (A3) is rewritten as:

$$\text{(A4)} \qquad \max_{\mathcal{P}(\phi)\in[0,1]} \mathbb{E}[\mathcal{P}(\phi) * f(\phi)]$$

Equation (A4) implies that the optimal policy $\mathcal{P}^*(\phi)$ has the following form ($\mathcal{I}(\cdot)$ is an indicator function):

$$\text{(A5)} \qquad \mathcal{P}^*(\phi) = \mathcal{I}(f(\phi) \geq 0) = \mathcal{I}(\mathbb{E}[\pi|\phi]) \geq 0)$$

We show in the following that, for the case of normal priors, this policy would have a threshold form. [Note that the optimal policy in Equation (A5) does not depend on the normality assumption profitability; the policy is easily generalized to other distributions.]

Because $\hat{\pi}$ is normally distributed conditionally on π and since the prior is also normally distributed, the posterior is normally distributed. Using standard formulae, we write:

(A6) $$\pi|\hat{\pi}\sim\mathcal{N}\left(\frac{\hat{\pi}\sigma_0^2+\mu_0\sigma_1^2}{\sigma_0^2+\sigma_1^2};\ \frac{\sigma_0^2\sigma_1^2}{\sigma_0^2+\sigma_1^2}\right) \quad \text{and} \quad \hat{\pi}\sim\mathcal{N}(\mu_0;\ \sigma_0^2+\sigma_1^2)$$

From (A6), it follows that:

(A7) $$\mathbb{E}[\pi|\phi]=\frac{\hat{\pi}\sigma_0^2+\mu_0\sigma_1^2}{\sigma_0^2+\sigma_1^2}\Rightarrow\mathcal{P}^*(\phi)=\mathcal{I}(\mathbb{E}[\pi|\phi]\geq 0)=\mathcal{I}\left(\hat{\pi}\geq-\mu_0*\frac{\sigma_1^2}{\sigma_0^2}\right)$$

Which is the threshold policy.

Result 2. Under the assumptions of Result 1, the optimal expected profit is:

(A8) $$\Pi^*=\left(1-\Phi\left(-\frac{\mu_0}{\sigma_\nu}\right)\right)*\mu_0+\sigma_\nu*\varphi\left(-\frac{\mu_0}{\sigma_\nu}\right)$$

Where $\Phi(\cdot)$ and $\varphi(\cdot)$ are the standard normal CDF and PDF respectively, and $\sigma_\nu=\frac{\sigma_0^2}{\sqrt{\sigma_0^2+\sigma_1^2}}$.

***Proof**: By* substituting the optimal policy from (A7) and conditional expectation from (A8) to (A2), we rewrite the expected optimal profit as:

(A9) $$\Pi^*=\mathbb{E}\left[\mathcal{I}\left(\hat{\pi}\geq-\mu_0*\frac{\sigma_1^2}{\sigma_0^2}\right)*\left(\frac{\hat{\pi}\sigma_0^2+\mu_0\sigma_1^2}{\sigma_0^2+\sigma_1^2}\right)\right]=\mathbb{E}[\mathcal{I}(\nu\geq 0)*\nu]=\mathbb{P}[\nu\geq 0]\mathbb{E}[\nu|\nu\geq 0]$$

where $\nu=\frac{\hat{\pi}\sigma_0^2+\mu_0\sigma_1^2}{\sigma_0^2+\sigma_1^2}\sim\mathcal{N}\left(\frac{\hat{\pi}\sigma_0^2+\mu_0\sigma_1^2}{\sigma_0^2+\sigma_1^2};\frac{\sigma_0^4}{(\sigma_0^2+\sigma_1^2)^2}(\sigma_0^2+\sigma_1^2)\right)\sim\mathcal{N}\left(\mu_0;\frac{\sigma_0^4}{\sigma_0^2+\sigma_1^2}\right)\sim\mathcal{N}(\mu_0;\sigma_\nu^2)$

Because ν is normally distributed, (A9) can be rewritten using the formula for the expectation of the truncated normal distribution:

(A10) $$\Pi^*=\left(1-\Phi\left(-\frac{\mu_0}{\sigma_\nu}\right)\right)*\mu_0+\sigma_\nu*\varphi\left(-\frac{\mu_0}{\sigma_\nu}\right)$$

Result 3. The expected profit under the optimal policy is a decreasing function of σ_1^2.

***Proof**:* Taking the derivative of (A10) with respect to σ_1^2:

(A11) $$-\mu_0*\varphi\left(-\frac{\mu_0}{\sigma_\nu}\right)\left(-\frac{\mu_0}{2\sigma_0^2(\sigma_0^2+\sigma_1^2)^{\frac{1}{2}}}\right)-\frac{\sigma_0^2}{2(\sigma_0^2+\sigma_1^2)^{\frac{3}{2}}}\varphi\left(-\frac{\mu_0}{\sigma_\nu}\right)+$$

$$\frac{\sigma_0^2}{(\sigma_0^2+\sigma_1^2)^{\frac{1}{2}}}\varphi'\left(-\frac{\mu_0}{\sigma_\nu}\right)\left(-\frac{\mu_0}{2\sigma_0^2(\sigma_0^2+\sigma_1^2)^{\frac{1}{2}}}\right)=\left(\frac{\mu_0^2}{2\sigma_0^2(\sigma_0^2+\sigma_1^2)^{\frac{1}{2}}}-\frac{\sigma_0^2}{2(\sigma_0^2+\sigma_1^2)^{\frac{3}{2}}}+\right.$$

$$\frac{\sigma_0^2}{(\sigma_0^2+\sigma_1^2)^{\frac{1}{2}}}\left(\frac{\mu_0(\sigma_0^2+\sigma_1^2)^{\frac{1}{2}}}{\sigma_0^2}\right)\left(-\frac{\mu_0}{2\sigma_0^2(\sigma_0^2+\sigma_1^2)^{\frac{1}{2}}}\right)\varphi\left(-\frac{\mu_0}{\sigma_v}\right)=\left(\frac{\mu_0^2}{2\sigma_0^2(\sigma_0^2+\sigma_1^2)^{\frac{1}{2}}}-\frac{\sigma_0^2}{2(\sigma_0^2+\sigma_1^2)^{\frac{3}{2}}}+\right.$$

$$\left.\left(-\frac{\mu_0^2}{2\sigma_0^2(\sigma_0^2+\sigma_1^2)^{\frac{1}{2}}}\right)\right)\varphi\left(-\frac{\mu_0}{\sigma_v}\right)=-\frac{\sigma_0^2}{2(\sigma_0^2+\sigma_1^2)^{\frac{3}{2}}}\varphi\left(-\frac{\mu_0}{\sigma_v}\right)$$

Because $\varphi(\cdot)>0$ and $-\frac{\sigma_0^2}{2(\sigma_0^2+\sigma_1^2)^{\frac{3}{2}}}<0$, the expected profitability is decreasing function of σ_1^2 and therefore an increasing function of model accuracy.

References

Asdecker B (2015). Returning mail-order goods: analyzing the relationship between the rate of returns and the associated costs. *Logistics Research*. 8(1):1–12.

Anderson ET, Hansen K, Simester D (2009) The option value of returns: Theory and empirical evidence. *Marketing Science. 28*(3): 405–423.

Ansari A, Mela CF, Neslin SA (2008) Customer channel migration. *Journal of Marketing Research*. *45*(1):60–76.

Archak N, Ghose A, Ipeirotis, PG (2011) Deriving the pricing power of product features by mining consumer reviews. *Management science. 57*(8):1485–1509.

Boatwright P, Nunes JC (2001), Reducing assortment: An attribute-based approach. *Journal of Marketing*. 65 (3):50–63.

Briesch RA, Chintagunta PK, Fox EJ (2009) How does assortment affect grocery store choice?. *Journal of Marketing Research* 46 (2): 176–189.

Broniarczyk SM, Hoyer WD, McAlister L (1998) Consumers' perceptions of the assortment offered in a grocery category: The Impact of Item Reduction. *Journal of Marketing Research.* 35 (2):166–176.

Chevalier JA, Mayzlin D (2006) The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, *43*(3):345–354.

Chen T, Guestrin C (2016) XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

Dzyabura D, Jagabathula S, Muller E (2019) Accounting for discrepancies between online and offline shopping behavior. *Marketing Science*. 38(1):88–106.

El Kihal S, Erdem T, Schulze C (2019) Is how you start how you finish? Customer return rate evolution over time. Working Paper.

Gittins J, Glazebrook K, Weber R (2011) *Multi-arm bandit allocation indices*, John Wiley & Sons, Ltd:

Chichester United Kingdom.

Guolin K, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liw T (2017) LightGMB: A Highly Efficient Gradient Boosting Decision Tree. *Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS), Long Beach, CA*.

He R, McAuley J (2016) VBPR: Visual Bayesian personalized ranking from implicit feedback. *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*.

He K, Zhang X, Ren S, Sun J (2015) Deep residual learning for image recognition. arXiv preprint https://arxiv.org/abs/1512.03385.

Janakiraman N, Syrdal HA, Freling R (2016) The effect of return policy leniency on consumer purchase and return decisions: A meta-analytic Review. *Journal of Retailing*. 92(2):226–235.

iBusiness (2016), "Wie Shopbetreiber das Retourenproblem wirklich lösen" [How Online Retailers are Solving the Problem with Product Returns], http://www.ibusiness.de/aktuell/db/858729veg.html

Luo L (2011) Product line design for consumer durables: An integrated marketing and engineering approach. *Journal of Marketing Research.* 48(1):128–139.

Lee TY, Bradlow ET (2011) Automated marketing research using online customer reviews. *Journal of Marketing Research.* 48(5):881–894.

Liu J, Toubia O (2018) A semantic approach for estimating consumer content preferences from online search queries. *Marketing Science.* 37(6):930–952.

Liu L, Dzyabura D, Mizik NV (2018) Visual listening in: Extracting brand image portrayed on social media. Working Paper.

Lynch C, Aryafar K, Attenberg J (2015) Images don't lie: Transferring deep visual semantic features to large-scale multimodal learning to rank. arXiv preprint arXiv:1511.06746.

Manjunath BS, Ma WY (1996) Texture features for browsing and retrieval of image data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 18(8): 837–842.

McAuley J, Targett C, Shi Q, van den Hengel A (2015) Image-based recommendations on styles and substitutes. arXiv preprint arXiv:1506.04757

Nasr-Bechwati N, Schneier Siegal W (2005) The impact of the prechoice process on product returns. *Journal of Marketing Research.* 42 (3):358–67.

Narang U, Shankar V (2019) Mobile App Introduction and Online and Offline Purchases and Product Returns. *Marketing Science*. Forthcoming.

Neslin S, Gupta S, Kamakura W, Lu J, Mason C (2006) Defection detection: Measuring and understanding the predictive accuracy of customer churn models. *Journal of Marketing Research.* 43(2):204–211.

Netzer O, Feldman R, Goldenberg J, Fresko M (2012) Mine your own business: Market-structure surveillance through text mining. *Marketing Science.* 31(3):521–543.

Onishi H, Manchanda P (2012) Marketing activity, blogging and sales. *International Journal of Research in Marketing.* 29(3):221–234.

Pauwels K, Neslin S (2015) Building with bricks and mortar: The revenue impact of opening physical stores in a multichannel environment. *Journal of Retailing.* 91(2):182–197.

Rafieian O, Yoganarasimhan H (2018) Targeting and privacy in mobile advertising. Working Paper.

Rublee E, Rabaud V, Konolige K, Bradski G (2011) ORB: An efficient alternative to SIFT or SURF. *International Conference on Computer Vision, Barcelona*, 2011, pp. 2564–2571

The Economist (2013) Return to Santa – E-commerce firms have a hard core of costly, impossible-to-please customers. *The Economist*. (December 21), https://www.economist.com/news/business/21591874-e-commerce-firms-have-hard-core-costly-impossible-please-customers-return-santa

Thomasson E (2013) Online retailers go Hi-Tech to size up shoppers and cut returns. Reuters (October 2), http://www.reuters.com/article/net-us-retail-online-returns-idUSBRE98Q0GS20131002

Timoshenko A, Hauser JR (2019) Identifying customer needs from user-generated content. *Marketing*

Science.38(1): 1–20.

Tirunillai S, Tellis GJ (2012) Does chatter really matter? Dynamics of user-generated content and stock performance. *Marketing Science.* 31(2):198–215.

Toubia O, Netzer O (2017) Idea generation, creativity, and prototypicality. *Marketing Science.* 36(1):1–20.

Safdar K, Stevens S (2018) Banned from Amazon: The shoppers who make too many returns. *Wall Street Journal* (June 11), https://www.wsj.com/articles/banned-from-amazon-the-shoppers-who-make-too-many-returns-1526981401

Shulman JD, Coughlan AT, Savaskan RC (2011) Managing consumer returns in a competitive environment. *Marketing Science.* 57(2): 347–362

Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. *Proceedings of International Conference on Learning Representations*.

Striapunina K (2019) Fashion ecommerce report 2019, Statista (July 2), https://www.statista.com/study/38340/ecommerce-report-fashion/

Wang K, Goldfarb A (2017) Can offline stores drive online sales? *Journal of Marketing Research. 54* (5):706–719.

Weitzman, ML (1979) Optimal search for the best alternative. *Econometrica.* 47(3):641–654.

Zhang M, Luo L (2018) Can user generated content predict restaurant survival: Deep learning of Yelp photos and reviews. Working Paper.

Zhang J, Krishna A (2007) Brand-level effects of stock keeping unit reductions. *Journal of Marketing Research.* 44 (4):545–559.

Zhang S, Lee D, Singh PV, Srinivasan K (2017) How much is an image worth? An empirical analysis of property's image aesthetic quality on demand at AirBNB, (May 25, 2017). Available at SSRN: https://ssrn.com/abstract=2976021

Zwebner Y, Rosenfeld N, Sellier A, Goldenberg J, Mayo R (2017) We look like our names: The manifestation of name stereotypes in facial appearance*. Journal of Personality and Social Psychology.* 112 (4): 527–554.