Demand Estimation with Text and Image Data*

Giovanni Compiani
University of Chicago

Ilya Morozov
Northwestern University

Stephan Seiler
Imperial College London
& CEPR

We propose a demand estimation method that leverages unstructured text and image data to infer substitution patterns. Using pre-trained deep learning models, we extract embeddings from product images and textual descriptions and incorporate them into a random coefficients logit model. This approach enables researchers to estimate demand even when they lack data on product attributes or when consumers value hard-to-quantify attributes, such as visual design or functional benefits. Using data from a choice experiment, we show that our approach outperforms standard attribute-based models in counterfactual predictions of consumers' second choices. We also apply it across 40 product categories on Amazon.com and consistently find that text and image data help identify close substitutes within each category.

Keywords: Demand Estimation, Unstructured Data, Deep Learning

*We thank Steve Berry, Alex Burnap, Matt Gentzkow, Joonhwi Joo, Philip Haile, Yufeng Huang, Ayush Kanodia, Carl Mela, Sanjog Misra, Olivia Natan, Matt Osborne, Katja Seim, Adam Smith, Yuyan Wang, Kevin Williams and seminar participants at Bocconi University, Chicago Booth, London Business School, Stockholm School of Economics, University of Cologne, Nova SBE, Marketing Science Conference 2022, SICS Conference 2024, AI/ML Yale Conference, and SED Conference for helpful comments. We also thank Celina Park, Tanner Parsons, James Ryan, Janani Sekar, Andrew Sharng, Franklin She, Adam Sy, and Vitalii Tubdenov for outstanding research assistance. Contact information: Giovanni Compiani (gio1compiani@gmail.com), Ilya Morozov (ilya.morozov@kellogg.northwestern.edu), Stephan Seiler (stephan.a.seiler@gmail.com). Giovanni Compiani gratefully acknowledges support from the Center for Applied AI at Booth.

1 Introduction

Many problems in economics and marketing—such as merger analysis, tariff evaluation, and optimal pricing—require researchers to estimate demand for differentiated products. A standard approach to these problems has been to estimate demand models that capture substitution through the similarity of product attributes. While common, this approach faces two practical challenges. First, researchers rarely observe all attributes that differentiate products. Instead, they rely on third-party data, where attributes are chosen based on unknown criteria, or gather their own data, subjectively selecting which attributes to collect. The collected attributes may not align with those most relevant to consumer choices. Second, consumers often consider visual design and functional benefits of products—dimensions that are difficult to capture through observed attributes.[1]

In this paper, we show how researchers can incorporate text and image data in demand estimation to recover substitution patterns. Our approach uses product images, titles, descriptions, and customer reviews—unstructured data that are widely available even in markets where collecting product attributes is challenging. Using pre-trained deep learning models, we extract low-dimensional features from images and texts and include them in a standard logit demand model, interacting them with random coefficients to capture substitution patterns. This approach enables researchers to incorporate hard-to-quantify product attributes, such as visual design from images and functional benefits from text, while also circumventing the need to subjectively choose which observed attributes to collect.

To validate our approach, we run an experiment that elicits consumers' first and second choices. We estimate our model on first choices and show that it outperforms standard attribute-based demand models at counterfactually predicting second choices—a key empirical measure of substitution. Our findings suggest that researchers can use text and images instead of traditional attribute data and that these unstructured data may even capture substitution beyond what is reflected in observed product attributes. Lastly, we apply our approach to e-commerce data and consistently find that text and images effectively identify close substitutes across many product categories.

[1] This criticism of attribute-based models has a long history in the economics literature. For example, Hausman (1994, p.229) skeptically remarks that applying such models to French champagne choices would require researchers to somehow quantify the bubble content.

Proposed Approach and Validation To extract information from unstructured data, we use pre-trained deep learning models, transforming images and texts into vector representations—*embeddings*. We then apply principal component analysis (PCA) to these embeddings to further reduce dimensionality and capture the main dimensions of product differentiation. This process generates separate embeddings and principal components for each data source: product images, titles, descriptions, and customer reviews. We incorporate these principal components into a mixed logit demand model (Berry et al., 1995; Nevo, 2001), interacting them with random coefficients similar to how researchers usually treat observed product attributes in demand estimation. To identify the principal components that best capture substitution in the data, we select the specification that minimizes the Akaike Information Criterion (AIC).

Just as standard demand models capture substitution through the similarity of observed attributes, our approach captures it by looking at how close products are in the space of texts and images. This proximity will reflect substitution if consumers consider how a product looks and how it is described by sellers and other consumers. Alternatively, visual and textual descriptions may serve as proxies for product features and functional benefits. Take tablets, for example: product titles may reveal brand and screen size, descriptions may highlight whether the tablet is suitable for drawing or gaming, reviews may mention that a tablet is durable and child-friendly, and photos may showcase design features like color and casing style. By extracting embeddings from texts and images, we identify these dimensions of differentiation.

To evaluate our approach, we run an experiment where participants choose a book from a list of options. We randomize both the prices and rankings of books to generate the variation needed to identify substitution patterns (Berry and Haile, 2014). We elicit participants' first and second choices, estimate each demand model using only first choices, and then counterfactually predict second choices. Because second choices reveal which books consumers substitute to when their first choice is unavailable, predicting them well requires a demand model that accurately captures substitution patterns.

We find that incorporating either text or images in demand estimation improves both in-sample fit and counterfactual predictions relative to the plain logit model without random coefficients. Images capture substitution because book covers visually convey genre and distinguish between fiction and non-fiction. Text performs even better, especially when we use more advanced models and richer text data. This is likely because descriptions and reviews capture nuanced details about book plots, helping identify similar books even within the same genre. Crucially, our best-fitting specification—based on customer reviews—outperforms standard attribute-based demand models in counterfactual predictions. These findings suggest that text and images can replace standard attribute data and may even capture substitution patterns beyond what is reflected in observed attributes.

Beyond improving fit and counterfactual second-choice predictions, our approach accurately identifies the closest substitutes among products. We show that for some books our method correctly recovers the five closest substitutes, whereas alternative models misidentify them. Additionally, we simulate hypothetical mergers of book publishers and find that demand models diverge in their predictions of expected price increases. This divergence has significant implications for downstream decisions. For instance, if our model captures substitution accurately, an antitrust agency relying on a plain logit model or an attribute-based mixed logit model will incorrectly approve mergers that should be challenged and block mergers that should be approved.

Lastly, we apply our approach to 40 product categories on Amazon.com, including groceries, pet food, office supplies, beauty products, electronics, video games, and clothing. We combine data on online purchases from the Comscore Web Behavior panel with product images, titles, descriptions, and reviews collected from Amazon's product detail pages. Across all categories, we consistently find that text and images contain useful information about substitution patterns. This additional information helps identify close substitutes within each category, leading to substantial deviations from restrictive substitution patterns in the plain logit model. These results confirm that our approach performs consistently well across a wide range of products, even in observational data that has less price variation than our experiment.

Contributions Our main innovation is to show how researchers can incorporate unstructured text and image data in demand estimation to better recover substitution patterns. Standard demand estimation methods assume a pre-determined nesting structure (McFadden, 1974a), account for preference heterogeneity over observed product attributes (Berry et al., 1995; McFadden and Train, 2000; Berry et al., 2004), or parameterize utility covariances based on distances between products in attribute space (Bresnahan et al., 1997; Pinkse and Slade, 2004; Dotson et al., 2018; Chintagunta and Li, 2024).[2] Building on this body of work, we show how researchers can incorporate product images and textual product descriptions into standard demand models.[3]

Given the prevalence of unstructured data, we see our approach as a valuable addition

[2]The idea of treating products as bundles of characteristics dates back to Lancaster (1966) and McFadden (1974b).

[3]In concurrent work, Sisodia et al. (2024) extract interpretable product attributes from images and use them as inputs to a conjoint analysis.

to the toolbox of empirical researchers. To make it easier for others to apply our approach, we provide a publicly available Python package, *DeepLogit*, alongside this paper.[4]

Extracting information from images can be especially valuable in markets where visual design and style matter, such as shoes, clothing, luxury watches, handbags, and art. Similarly, textual descriptions and reviews can capture hard-to-quantify attributes, such as network reliability and customer service for telecom services or usability, customizability, and developer support for software products. Even when product attributes are available, including information from texts and images may still improve demand estimates by capturing additional dimensions of product differentiation.

A natural consequence of using text and images is that our approach scales well across product categories. In contrast, capturing substitution through product attributes across many categories can be challenging. For example, Döpper et al. (2024) remark that doing so "would be difficult to implement at scale because it would require category by-category assessments about which characteristics are appropriate to include and whether or not relevant data are available." Our approach circumvents the need to collect distinct attributes for each category, making it easier to estimate demand across many categories. This scalability may also be valuable for retailers and platforms seeking to estimate demand across diverse categories and optimize prices at scale.

We validate our approach using both experimental and observational data. Our experiment design lets us directly measure second-choice *diversion ratios*, allowing us to assess how well an estimated demand model predicts them counterfactually. Diversion ratios reflect substitutability and thus the intensity of price competition between two products. They are crucial for antitrust authorities trying to predict price increases from horizontal mergers (Shapiro, 1995; Conlon and Mortimer, 2021; Conlon et al., 2022), and for managers optimizing assortments and prices. We show that incorporating texts and images into demand models improves diversion ratio estimates, highlighting the value of our approach. Additionally, we use observational data to show that our method performs well across various product categories. We consider this extensive validation with both experimental and observational data an important contribution of the paper.

Several other papers have proposed leveraging text and image data in demand estimation.[5] Quan and Williams (2019) incorporate product image embeddings as shifters

[4]The package is available on PyPI and in the public GitHub repository: github.com/deep-logit-demand/deeplogit. Our package heavily borrows from the existing package *Xlogit* for GPU-accelerated estimation of mixed logit models, developed by Arteaga et al. (2022).

[5]Netzer et al. (2012) use data on the co-occurrence of products mentioned in online discussion forums to generate a visual representation of competing products, but they do not incorporate these measures into demand estimation.

of mean utilities in a demand model, showing their approach produces reasonable price elasticity estimates. Similarly, Lee (2024) employs large-language models to predict utility intercepts of new products based on their textual descriptions. These methods use texts or images to predict mean utilities but are not designed to estimate utility covariances, which play a crucial role in shaping substitution.[6] In contrast, our approach uses the same types of unstructured data to model utility covariances. This enables us to capture flexible substitution patterns—a key input for many empirical applications of interest.

A few papers propose estimating substitution patterns using survey data, a possible alternative to our use of unstructured data. Dotson et al. (2019) have survey participants rate each product image and incorporate rating correlations as utility correlation shifters in their demand model. Similarly, Magnolfi et al. (2022) ask survey participants to rank products relative to each other (e.g., "Product A is closer to B than to C"). They generate low-dimensional product embeddings from these rankings and use them as attributes in a random coefficients logit model. Like our paper, Magnolfi et al.'s approach provides a way to crowdsource product attributes when collecting such data is challenging. While our approach can complement these survey-based methods, it has the advantage of using only widely available text and image data, eliminating the need to conduct costly category-specific surveys.[7]

Our approach is well-suited for several empirical applications. Researchers can use it to address a wide range of economics and marketing questions that require accurate and flexible estimates of product substitution. This includes analyzing how horizontal mergers (Nevo, 2000; Federal Trade Commission, 2022), new product launches (Hausman, 1994; Petrin, 2002), corrective taxes (Allcott et al., 2019; Seiler et al., 2021), and trade restrictions (Goldberg, 1995; Berry et al., 1999) influence consumers' choices and welfare through changing assortments and prices. Additionally, marketers can apply this approach to help multi-product firms optimize prices and promotions (Hoch et al., 1995; Hitsch et al., 2021).

[6]For example, Lee (2024) inverts a plain logit model without random coefficients to extract product fixed effects for the prediction part of his analysis. Thus, his model restricts substitution patterns due to the IIA property.

[7]Dew (2024) also uses survey data to estimate consumer preferences for text and image embeddings. However, his focus is on predicting individual choices rather than estimating demand and modeling substitution.

2 Proposed Approach

Our approach involves three steps: (1) extracting embeddings from images and texts, (2) reducing the dimensionality of these embeddings using PCA, and (3) including the resulting principal components in a random coefficients logit model. In this section, we focus on the first two steps, as the mixed logit model is standard and does not require a detailed explanation. Section 3.2 describes the exact model specification used in our empirical applications.

2.1 Extracting Embeddings from Texts and Images

We extract embeddings from texts and images using pre-trained deep-learning models. This enables us to leverage models trained on large, general-purpose datasets, which can efficiently capture differences in product images and descriptions. Further, using pre-trained models reduces the computational burden of estimation.

2.1.1 Image Embeddings

We employ deep learning models that were originally trained for object detection and classification tasks. Specifically, we use four pre-trained convolutional neural networks available in the Keras Deep Learning Library: VGG19, ResNet50, InceptionV3, Xception.[8] All four models achieve high predictive accuracy on the ImageNet validation dataset, and are widely known for their strong performance in image classification tasks.[9]

We extract image embeddings from product photos, which researchers can easily obtain via online search or by scraping product pages from e-commerce platforms. Each model transforms the original image into a vector representation – an *embedding* – and classifies an image from its embedding by predicting the object it contains. These models were originally trained to classify images into labeled categories (e.g., "cup," "book," or "sofa"). However, since our goal is not to label products but to measure visual features that distinguish them from each other, we remove the classification layer from these models and work directly with embeddings. Because these models perform well at distinguishing similar objects, we expect the embeddings to capture key visual features that differentiate products.

[8] VGG19, a deep convolutional neural network with 19 layers (Simonyan and Zisserman, 2015). ResNet50, is a 50-layer convolutional neural network (He et al., 2016). InceptionV3 and Xception are convolutional neural networks with 48 and 71 layers (Szegedy et al., 2016; Chollet, 2017).

[9] The ImageNet data (https://www.image-net.org/) is used for the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), which evaluates algorithms for object detection and image classification, a benchmark in object category classification (Russakovsky et al., 2015).

We do not commit to a single model and intentionally use multiple deep learning models with different architectures. As we explain in Section 3.2, we select the one that best explains substitution patterns in the data.

2.1.2 Text Embeddings

Next, we extract information from text data, including product titles, descriptions, and customer reviews. We use two simple but widely used "count" models as well as pre-trained models with varying levels of sophistication, generating a separate set of embeddings for each model and type of text data. Before extracting embeddings, we pre-process text data as described in Appendix A.

First, we use a bag-of-words count model, which transforms text into fixed-length vectors by counting word occurrences. This model does not consider word order and only measures how often words appear in a document. Second, we apply a bag-of-words model with a TF-IDF vectorizer. Although similar to the first method, TF-IDF assigns more weight to words that are frequent in a specific text but rare across others, thus emphasizing unique words. This makes it more likely that text embeddings capture distinctive words that differentiate products from one another.

Next, we use the Universal Sentence Encoder (USE), a transformer-based language model (Cer et al., 2018), and the Sentence Transformer (ST) model, a pre-trained Sentence-BERT model designed to produce semantically meaningful sentence embeddings (Reimers and Gurevych, 2019).[10] Unlike the bag-of-words models, both USE and ST account for word order and the context in which words appear within sentences. Embeddings from these models are commonly used for text classification, semantic similarity analysis, clustering, and other natural language processing tasks. Both models achieve excellent performance on semantic textual similarity benchmarks (Cer et al., 2017).

While count models are relatively simple, they can detect individual attributes mentioned in titles, descriptions, and reviews. In contrast, the more advanced ST and USE models are better at capturing nuanced similarities in how sellers and consumers describe products. For example, they can identify when sellers or consumers describe the same functional benefits using similar language, even if the words are not exactly the same. Thus, we view count models as a useful benchmark that tells us whether our method merely extracts attributes mentioned in texts or captures more subtle cues about product substitution.

[10] The model trained by Reimers and Gurevych (2019) is a more efficient version of the widely used BERT network (Devlin et al., 2018).

2.2 Generating Principal Components

Embeddings provide a lower-dimensional representation of the original texts and images. However, they remain high-dimensional compared to the number of attributes researchers typically include in demand models. For example, it would be impractical to incorporate 512-dimensional VGG19 embeddings into a random coefficients logit model, as it would require computationally intensive numerical integration over all dimensions. For this reason, we further reduce the dimensionality by applying Principal Component Analysis (PCA) to the embeddings.

For a given category, we extract principal components from each set of embeddings corresponding to a given model and data type (e.g., USE embeddings from customer reviews) and incorporate them into a standard mixed logit model (Berry et al., 1995, 2004), estimating a separate random coefficient for each component. Principal components are particularly appealing because they are orthogonal to each other, which mitigates multicollinearity and simplifies the estimation of random coefficients in the mixed logit model (Backus et al., 2021). We perform model selection by choosing the combination of principal components that minimizes the AIC (see Section 3.2.1).

Because we apply PCA separately within each category, the extracted principal components capture the key dimensions that differentiate products within that category. By contrast, raw embeddings come from models trained on large, general-purpose datasets. For instance, image embeddings come from models trained to classify images into broad categories, such as "tablet," "laptop," or "cellphone," and thus contain features that distinguish images across these categories. However, to estimate demand, we need to analyze consumer choices *within a given category*, where products often share visual features due to similar designs. PCA helps us filter out the variation in embeddings necessary for sorting products into categories, allowing us to focus on dimensions most relevant for analyzing substitution patterns within each category.

2.3 Interpretation of Principal Components

Standard mixed logit models capture substitution by measuring how similar products are in terms of their observed attributes. Similarly, our approach captures substitution by measuring how similar products are in terms of their visual designs and textual descriptions.

There are several reasons why this proximity may be linked to substitution. Consumers may choose based on how a product looks and how it is described by sellers or other consumers. For instance, a consumer who finds one product visually appealing may also be drawn to similar-looking products, and they might expect products with similar descriptions to meet their needs in a similar way.

Alternatively, even when consumers do not directly consider visual or textual descriptions, these descriptions may reflect key product features and benefits. Take tablets, for example: product titles may reveal brand, screen size, and camera resolution (e.g., "Apple iPad 10.2-inch 12MP camera"); seller descriptions may highlight whether the tablet is suitable for drawing or gaming; and consumer reviews may mention that a tablet is durable and child-friendly. Such texts may convey functional benefits that are hard to quantify. Similarly, photos may showcase design features like color, casing style, or the presence of a keyboard. By extracting principal components from texts and images, our approach identifies these dimensions of product differentiation and uses them to capture which products are close substitutes.

Given this interpretation, researchers should not treat the coefficients on the principal components as causal in the sense that changing product images or textual descriptions does not necessarily cause a change in substitution patterns. For example, if the principal components serve as proxies for choice-relevant product attributes, these attributes need not change if the seller modifies the product's description or visual design. Hence, the principal components should be held fixed in counterfactual analyses. For this reason, our approach is not suited for studying optimal product design and positioning, or any other questions that require causal estimates of the effects of non-price characteristics on choices.

In some markets, texts and images might not convey information relevant to consumer choices, in which case the principal components would fail to predict substitution patterns. However, we consider this scenario unlikely, as sellers have strong incentives to highlight important product attributes in titles and descriptions, and online platforms may encourage them to do so. Similarly, when writing reviews, consumers often mention product aspects they think others will find useful. Consistent with this idea, we find that in all product categories we study, at least some image- or text-based principal components predict substitution patterns.

Similar to observed product attributes in standard mixed logit demand models, we treat principal components as fixed over time. This assumption is reasonable because principal components capture time-invariant product attributes, functional benefits, and visual design features. While sellers could theoretically modify images and textual descriptions to reposition their products, we find that in practice, product images, titles, and descriptions rarely change in our empirical context (see Appendix C for details).

2.4 Advantages of Our Approach

To understand how our approach relates to existing methods, we compare it to the standard approach commonly used in demand estimation, which relies on observed product attributes. These attributes are usually incorporated into a discrete choice demand model, such as a random coefficients logit model, where substitution patterns are captured through the interaction between product attributes and preference heterogeneity.

While common, this approach presents two practical challenges. First, researchers rarely observe all attributes that differentiate products. Instead, they often rely on third-party datasets, where data providers report a limited set of attributes selected based on unknown criteria. Others gather their own data using a combination of web scraping and human coders—–a costly and time-consuming process that typically limits researchers to collecting only easily quantifiable attributes. Since attributes are selected subjectively in both cases, they may not align with those most relevant to consumer choices. Our approach avoids collecting data on product attributes and instead extracts information about product substitution from text and image data.

The second challenge is that consumers often consider visual design and functional benefits—dimensions that are difficult to capture through observed attributes. For instance, in categories like clothing or home decor, product aesthetics can strongly influence choices. A shopper selecting a dress might care about overall style, design pattern, or neckline type—details visible in photos but difficult to quantify without training a custom classification model. Similarly, textual descriptions might reveal attributes that are hard to quantify. Customer reviews, for instance, might highlight that certain headphones are ideal for listening to audiobooks, revealing a benefit not captured by observed attributes. Our approach allows researchers to estimate substitution patterns along these otherwise unobserved dimensions.

A natural consequence of using text and images is that our approach scales more easily across categories than standard attribute-based models. For instance, Döpper et al. (2024) estimate demand estimation across a large set of categories and note that incorporating product attributes would be difficult because it would require them to decide which attributes are relevant in each category and collect data on those attributes. Our approach circumvents the need to collect category-specific attributes, making it easier to estimate demand across many categories. This scalability may also be valuable for online retailers like Amazon or Alibaba looking to optimize prices across hundreds of categories. In Section 4, we further illustrate this point by applying our approach across 40 Amazon categories.

3 Choice Experiment

We begin by applying our method to data from a choice experiment. In this experiment, we randomize product prices and positions, generating clean variation that helps estimate substitution patterns. We also collect data on consumers' second choices, which are rarely available in observational datasets. These data allow us to assess how well our model recovers substitution patterns by testing its ability to predict second choices counterfactually relative to alternative models.

3.1 Experiment Design

We recruited participants from the online platform Prolific and asked each participant to complete two choice tasks via a Qualtrics survey.

In the first choice task, we asked each participant to choose one book from a set of ten alternatives (Figure 1). These ten books were chosen from Amazon's bestseller lists (Figure 2). We chose them from three different genres—Mystery, Fantasy, and Self-Help—anticipating that genre would predict substitution patterns and could thus be incorporated in attribute-based demand models. We instructed each participant to choose the book they would purchase if faced with this selection in a real bookstore.

Figure 1 shows an example of a choice task, displaying the top portion of the page as it appeared to participants. For each book, we showed participants the cover image, title, author, publication year, genre, number of pages, description, and five customer reviews, all collected from Amazon product pages. We presented books in a random order to each participant and drew each book's price from a discrete uniform distribution ranging from $3 to $7.

After a participant selected a book from the list of ten options, we removed this book from the list. The participant then proceeded to a second choice task, where they chose from the remaining nine books. The positions and prices of the books remained the same as in the first choice task, with the only difference being that the participant's first choice was no longer available.

Based on power calculations from a pilot experiment, we estimated that we needed around 10,000 participants to determine whether our model outperforms alternative models in predicting second choices. We pre-registered this target sample size.[11] In total, we recruited 10,775 participants between June 14 and June 27, 2024. Following the pre-registered sample selection criteria, we excluded around 14% participants who failed

[11] The pre-registration document for this study can be accessed at https://tinyurl.com/ekjp4pfp.

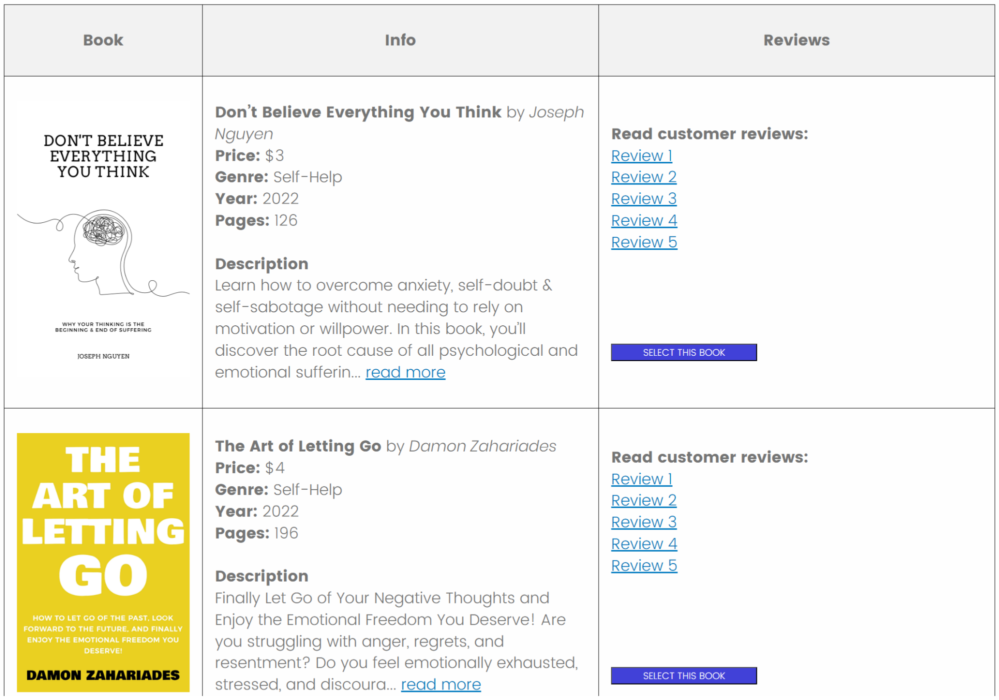

Figure 1: **Example of a choice task in our experiment.** The screenshot displays the top portion of the page as it appeared to participants.

comprehension questions, did not complete the survey, or spent less than one minute on the entire study. After applying these criteria, our final sample consists of 9,265 participants.

Because choice tasks in our experiment were hypothetical and not incentivized, it is important to verify that participants made meaningful selections. In Appendix B, we show that participants did not rush through the survey, responded to changes in book rankings and prices, and made choices consistent with their self-reported genre preferences.

3.2 Model Specifications

We compare our approach against two benchmark models: the plain logit model and a standard random coefficients model that leverages observed product attributes but uses no texts or images. All three models fall into the following framework, where the indirect utility of consumer i from purchasing book j is given by

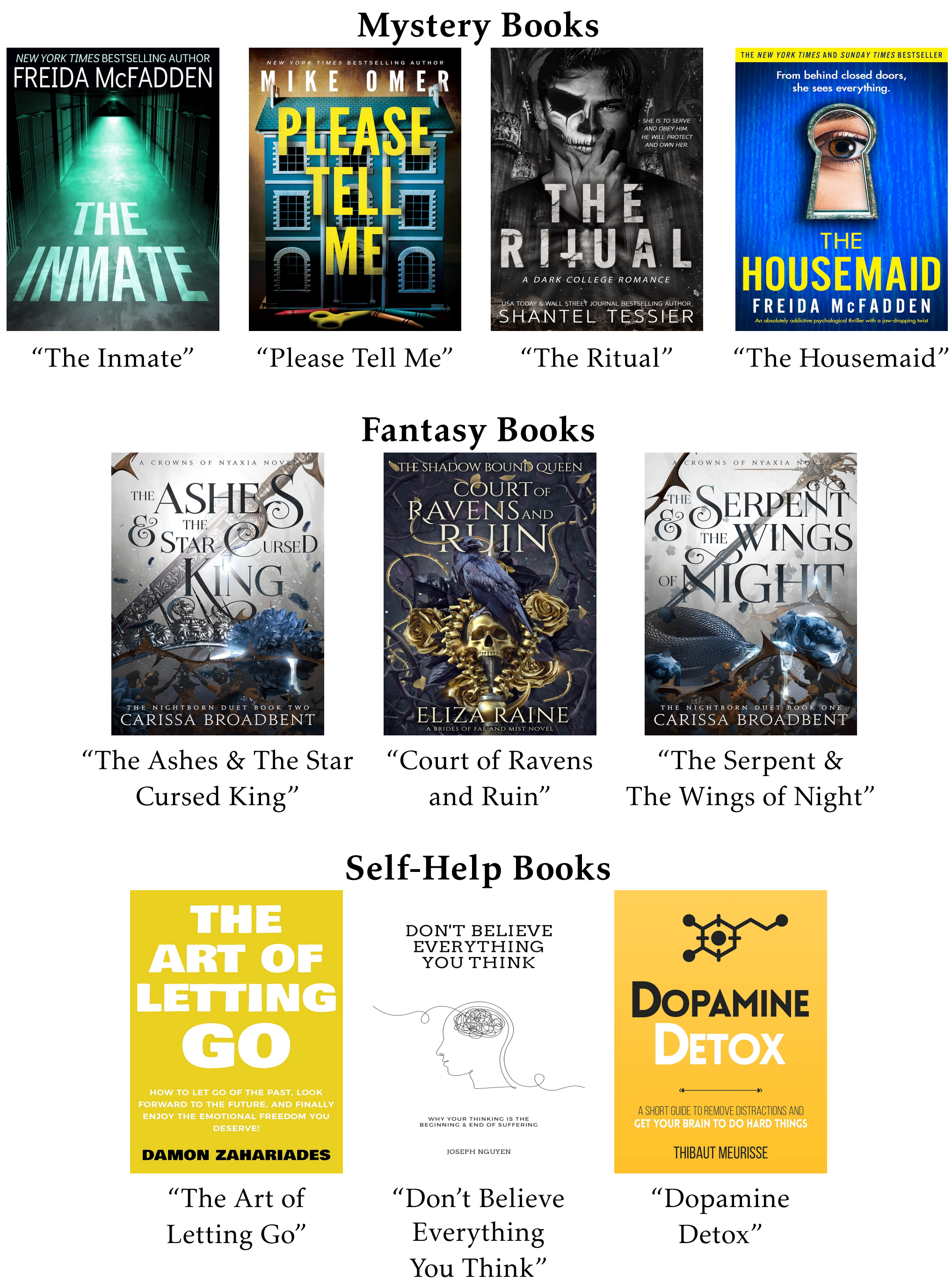

Figure 2: **Ten books used in our experiment.**

$$u_{ij} = \beta_i' x_j + \theta_i' PC_j - \gamma \cdot \text{rank}_{ij} - \alpha_i \cdot \text{price}_{ij} + \delta_j + \varepsilon_{ij}. \tag{1}$$

Here x_j is a vector of observed book attributes, PC_j are principal components extracted from text and image embeddings, price_{ij} and rank_{ij} are the price and position of book j in consumer i's choice tasks, δ_j are product fixed effects, and ε_{ij} are i.i.d. taste shocks following a Type I Extreme Value distribution. We include rank_{ij} in the model because our experiment induces random variation in the placement of books. This variation gives us an additional exogenous shifter of choices, which helps us identify substitution patterns in a manner similar to how price variation does. When specifying observed attributes x_j, we include genre dummies for Mystery and Sci-Fi, publication year, and length in pages—all attributes participants observe in the choice tasks.

We estimate the following three models within this framework:[12]

1. **Mixed Logit with Principal Components.** We include principal components PC_j, interacting them with random coefficients $\theta_i \sim N(0, \Sigma_\theta)$, where Σ_θ is diagonal. Since the principal components do not vary for a given product over time, the mean of the random coefficients is absorbed by the product fixed effects δ_j. In some specifications, we also estimate a random coefficient on price, assuming $\alpha_i \sim N(\bar{\alpha}, \sigma_\alpha)$. We do not include observed attributes, setting $\beta_i = 0$. This model mimics a scenario where a researcher has access to unstructured data but does not observe attributes other than price.

2. **Mixed Logit with Observed Attributes.** We include observed attributes x_j, interacting them with random coefficients $\beta_i \sim N(0, \Sigma_\beta)$, where Σ_β is diagonal, and again, the mean is absorbed by product fixed effects. In some specifications, we also estimate a random coefficient on price, assuming $\alpha_i \sim N(\bar{\alpha}, \sigma_\alpha)$. We do not include principal components PC_j, setting $\theta_i = 0$. This model reflects the standard practice of estimating substitution patterns through an interaction of observed attributes with consumer heterogeneity.

3. **Plain Logit.** We include neither attributes x_j not principal components PC_j, estimating only the uniform price coefficient $\alpha_i = \alpha$, the rank coefficient γ and product fixed effects δ_j. This model provides a baseline where product substitution is solely driven by the average utilities of books and not their correlations.

[12] We do not include a random coefficient on rank_{ij} in any model specification because our focus is on estimating substitution patterns related to product features, captured through observed characteristics or principal components from text and image data.

Comparing plain logit to mixed logit with observed attributes tells us whether adding random coefficients on book attributes improves the estimation of substitution patterns. Similarly, comparing mixed logit with attributes to mixed logit with principal components tells us whether incorporating texts and images into demand estimation enables us to recover substitution patterns as well as, or better than, using product attributes.[13]

3.2.1 Model Selection

Within each of the three model classes listed in Section 3.2, we perform model selection based on AIC.[14] In the mixed logit model with observed attributes, we consider all possible combinations of random coefficients across four book attributes: price, publication year, length in pages, and genre. Genre is represented by two dummy variables, one each for Mystery and one for Self-Help, making Sci-Fi the baseline genre. We estimate all 16 specifications and select the one with the lowest AIC.

In the mixed logit model with principal components, we take the first three principal components from each set of embeddings and evaluate all possible combinations of random coefficients on the three components and price, selecting the specification with the lowest AIC.[15] We repeat this selection process for every text and image model and data type (e.g., USE embeddings from reviews). Lastly, among the selected specifications, we choose the one that minimizes AIC.

We do not combine principal components extracted from different embeddings (e.g, mixing texts with images) in our main analysis. As we discuss later in Section 3.4.4, we tested mixing data types in our experimental data but found no improvement in model fit.

3.3 Validation Approach

Prior research often evaluated demand models based on how intuitive the estimated cross-price elasticities look, given researchers' prior knowledge of the market (Berry et al., 1995; Nevo, 2001). We follow a different approach by leveraging second choice

[13]To ensure a fair comparison between models, for each model we start optimization from 100 different initial points covering a wide range of parameter values. This helps us maximize the probability of finding the global optimum in each model.

[14]An advantage of using AIC is that it can be interpreted as an estimator of the expected relative Kullback-Leibler (KL) Information based on the maximized log-likelihood function, corrected for asymptotic bias (Akaike, 1998; Burnham and Anderson, 2004).

[15]Increasing the number of principal components beyond three has only a marginal impact on our results and does not affect the in-sample fit or counterfactual performance of our selected model.

data to compare our approach against alternatives. We estimate each model using only first choices and evaluate how well it predicts second choices.[16]

Since the second-choice data are not used in estimation, our validation approach forces alternative models to confront a difficult counterfactual prediction problem. Second choices reveal which books consumers substitute to when their most preferred option becomes unavailable. Therefore, a model can only predict second choices well if it has accurately captured the true substitution patterns. Predicting second choices is also directly relevant in some applications. For instance, antitrust authorities can use them in merger analysis to calculate *diversion ratios*—a measure reflecting substitutability and thus the intensity of price competition between products (Conlon and Mortimer, 2021). We discuss this further in Section 3.6.

We evaluate each model as follows. First, we obtain MLE estimates by maximizing the total likelihood of participants' first choices. Given these point estimates $\hat{\theta}$, we then compute the counterfactual second choice probabilities for each participant conditional on their first choice. Averaging these probabilities across participants, we construct the matrix $\mathbb{P}$, where each element (j,k) represents the probability that a participant whose first choice is j selects $k \neq j$ as their second choice. To assess each model, we compute the $RMSE$ based on element-wise differences between the matrix $\hat{\mathbb{P}}$ predicted by the model and $\mathbb{P}$ computed from empirical probabilities in the experimental data. A lower $RMSE$ indicates better model performance in predicting counterfactual second choices.

3.4 Validation Results

We compare the results of three models described in Section 3.2: (1) the plain logit model, (2) the mixed logit model with observed attributes, and (3) the mixed logit model with principal components. We evaluate these models based on the $RMSE$ of their counterfactual second-choice predictions. Figures 3 and 4 summarize the validation results, whereas Appendix Table A1 reports the AIC and $RMSE$ values for all specifications considered within each model class.

3.4.1 Plain Logit

We first report the results from the plain logit model. As shown in Appendix Table A1, this model achieves a counterfactual $RMSE$ of 0.089. Because this model assumes inde-

[16] Conceptually, our validation approach reverses the logic of Berry et al. (2004), who show that second choice data are useful for estimating substitution patterns. By contrast, we omit second choice data from estimation and use them only for model validation.

pendence of irrelevant alternatives (IIA), it cannot capture which close substitutes participants switch to when their first choices become unavailable. As a result, the model predicts relatively uniform second-choice probabilities that do not match the second choice data.

3.4.2 Mixed Logit with Principal Components

Next, Figure 3 presents results from the mixed logit model with principal components. The first panel shows $RMSE$ from the plain logit model, indicated by a vertical dashed line, while other panels display results from our proposed approach, which incorporates principal components obtained from different types of unstructured data and machine learning models. Each marker reflects the second-choice $RMSE$ from the best configuration of random coefficients on the principal components and price chosen by our model selection algorithm (see Section 3.2.1).

Performance varies widely across specifications. While some models perform similarly to the plain logit, others deliver significantly better counterfactual predictions. These differences highlight that researchers should always perform model selection to find the specification that best captures substitution patterns from text and images. In practice, researchers can rely on AIC as the main selection criterion. This works well in our experimental data, as AIC selects the model that also achieves the lowest counterfactual $RMSE$ (see Appendix Table A1).

Images Figure 3 shows that image-based mixed logit specifications perform well. All models in this class outperform the plain logit model, and the lowest-AIC model, InceptionV3, reduces $RMSE$ by 9.8% relative to the plain logit model (from 0.089 to 0.080).

To understand why images predict substitution, consider the book covers used in our study (see Figure 2). Within the same genre, book covers often share similar design elements. For example, the covers of all three fantasy books use dark, muted color palettes with metallic accents, include symbolic elements such as skulls and swords that convey danger or peril, feature natural objects like twisted vines and golden roses, and have similar fonts and title placements. Two of these books are by the same author and have nearly identical covers. Similarly, the self-help books have minimalistic layouts and consistent color schemes, such as black-and-white text on yellow backgrounds. Given these visual similarities, we expect image embeddings in our models to capture genre information, which correlates with substitution patterns.

However, we emphasize that image embeddings might be informative about substitution beyond simply revealing the book's genre. For example, second-choice data shows

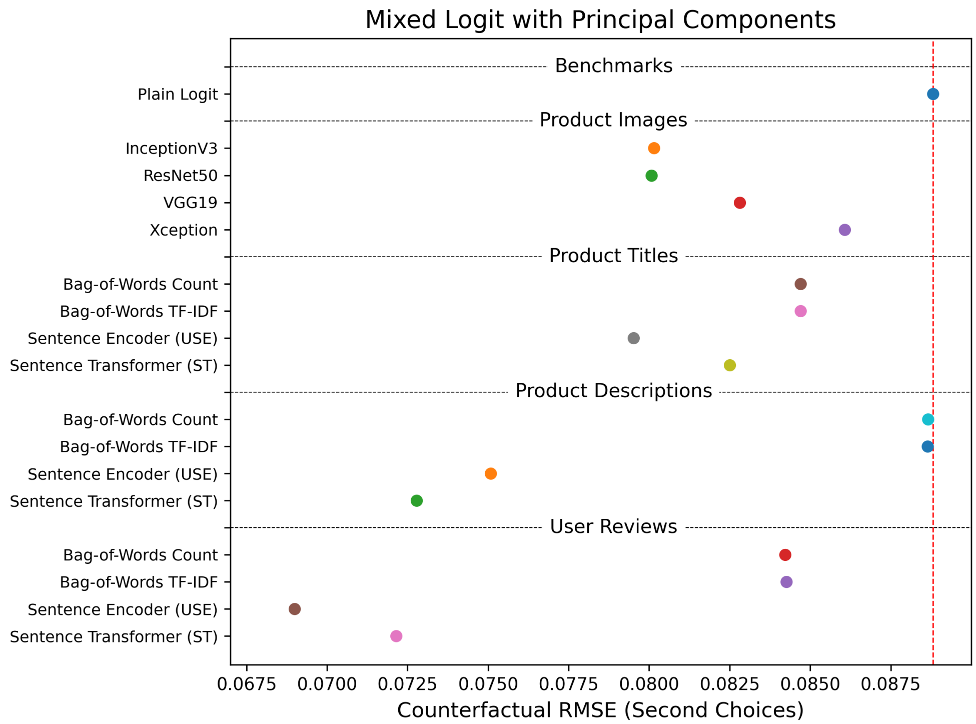

Figure 3: **Counterfactual performance of mixed logit models with principal components in terms of predicting second choices.**

that participants perceive mystery books as closer substitutes to fantasy books than self-help titles (Appendix Figure A3). Our image embeddings capture this pattern because mystery and fantasy book covers tend to be similar. Both use dark or muted tones with dramatic highlights (e.g., gold, silver, teal, or yellow) to create contrast, which is captured by our image embeddings.

Texts The remaining panels of Figure 3 show the performance of text-based models. We find that counterfactual second-choice predictions improve when we either use more advanced text models or incorporate richer text data.

First, counterfactual performance improves substantially when we move from simple bag-of-words models to the more advanced USE and ST models. While bag-of-words models provide little to no improvement over the plain logit, both USE and ST consistently outperform the plain logit across all types of text data. This may be because textual descriptions contain detailed information about book plots but do not always use the exact same words or phrases. Therefore, extracting substitution patterns from text requires a natural language model that can accurately measure semantic similarity.

Second, counterfactual performance improves as we include more text data into the

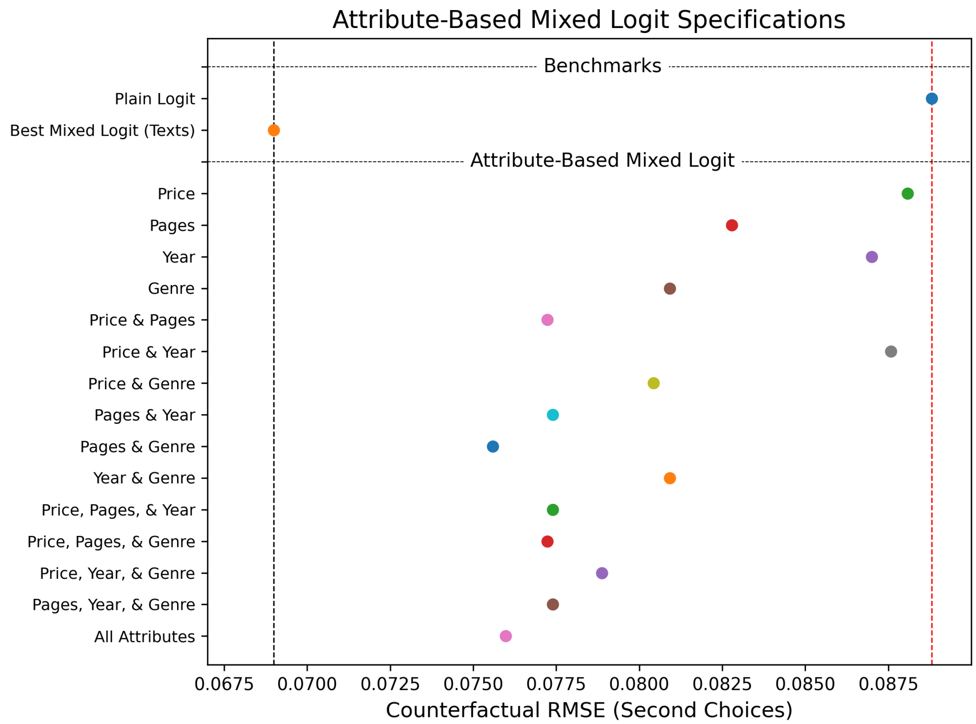

Figure 4: **Counterfactual performance of attribute-based mixed logit models in terms of predicting second choices.**

demand model. For instance, the USE model achieves a second-choice $RMSE$ of 0.080 with book titles, 0.075 with descriptions, and 0.069 with reviews. This indicates that customer reviews provide more information about substitution patterns than descriptions, which, in turn, are more informative than titles.

These results make intuitive sense. Although book titles are brief, they often contain subtle genre cues. For example, mystery titles frequently hint at secrets (*Please Tell Me*) or characters in confined situations (*Housemaid*, *Inmate*), whereas fantasy titles often include words that evoke a gothic atmosphere (*Serpent*, *Wings*, *Ravens*, *Ruin*). Book descriptions provide even clearer genre signals, emphasizing unexpected twists such as hidden identities, past relationships, and betrayals. Reviews go a step further: consumers summarize plots while also expressing opinions. When reviewing mystery novels, some readers praise cliffhangers and intriguing twists while others critique pacing issues like slow starts or rushed endings.

As with images, texts can provide information about substitution beyond simply revealing genres. For example, book descriptions and reviews make it clear that the self-help books *Dopamine Detox* and *Don't Believe Everything You Think* both discuss how modern distractions and mental habits negatively affect well-being. In contrast, another self-help book, *The Art of Letting Go*, focuses on letting go of painful memories and emotions. These differences may explain why our text-based models correctly identify the first two books as much closer substitutes, something alternative models fail to detect (see Section 3.5 for details).

The USE specification with reviews delivers the best counterfactual performance among all tested models, reducing $RMSE$ by 22.3% compared to the plain logit model (from 0.089 to 0.069). This model reduces the AIC by 30.8 and rejects the plain logit with p-value < 0.0001, indicating that the first-choice data provide significantly stronger support for the "USE Review" model than for the plain logit model.

3.4.3 Mixed Logit with Observed Attributes

Figure 4 presents the results for the mixed logit model with observed attributes. We compare it against two benchmarks: the plain logit model and the lowest-AIC text-based model identified earlier. Several specifications significantly outperform the plain logit model. The lowest-AIC specification is a mixed logit model with random coefficients on pages, year, and genre.[17] By capturing similarities in observed attributes, this model breaks the IIA pattern and generates more accurate second-choice predictions. Consequently, it reduces $RMSE$ by 12.9% compared to plain logit (from 0.089 to 0.077).

3.4.4 Summary

Taken together, these findings illustrate the value of our approach, showcasing its ability to extract information about substitution patterns from text and images. We deliberately chose books for this study, thinking that observed attributes like genre would strongly predict substitution. Yet, we find that many text-based models achieve comparable $RMSE$ to that of attribute-based models. Moreover, the best-fitting text-based model reduces $RMSE$ by an additional 10% relative to the best attribute-based model. Thus, incorporating text improves counterfactual predictions beyond what is achievable with observed attributes.

One might wonder whether texts, images, and attributes provide distinct information about substitution patterns. If so, the best approach to estimating demand would be to incorporate all three data types. To explore this, we start with our selected review-based specification and examine whether model fit can be improved by adding (1) any combination of observed product attributes or (2) any combination of principal components

[17] To be precise, this model achieves the same AIC as two other mixed logit models (i.e., "Price, Pages & Year" and "Pages & Year"), but all three specifications generate identical second-choice predictions, and thus, achieve the same $RMSE$.

from ResNet, the lowest-AIC image model. In both cases, we find no fit improvements. This result suggests information across the three data types is highly correlated, with text providing the strongest signal of substitution.

3.5 Implications for Product Substitution

So far, we have shown that our approach improves counterfactual predictions of second choices. One might wonder whether these $RMSE$ improvements are substantial and if they fundamentally change our understanding of product substitution in this market. To explore this further, in Table 1 we compare predicted second-choice probabilities with their counterparts observed in the experimental data.

Panel A examines substitution patterns for the self-help book *Dopamine Detox*. The data shows that the two other self-help books are, by far, its closest substitutes. The plain logit model misidentifies these substitutes, incorrectly predicting that people would switch to books *Please Tell Me* and *The Inmate*—two popular books with the largest market shares. The attribute-based mixed logit correctly identifies *Don't Believe Everything You Think* as the closest substitute but mispredicts the second one, likely due to its over-reliance on estimated fixed effects. By contrast, our review-based model is the only one that correctly predicts all five closest substitutes in the correct order.

Panel B shows a similar example with the substitutes for the mystery book *Please Tell Me*. The plain logit model mispredicts second-choice probabilities, incorrectly suggesting that, in the absence of this book, participants would mainly switch to self-help books. By contrast, the attribute-based mixed logit model captures strong within-genre substitution, correctly identifying the top three closest substitutes. Our review-based model also correctly predicts the identities of these substitutes but swaps the first two. However, the review-based model recognizes that *The Housemaid* and *The Inmate* have significantly higher second-choice probabilities than the third-closest substitute—an insight the attribute-based model misses.

These examples illustrate that, beyond reducing $RMSE$ and predicting second-choice probabilities more accurately on average, our approach can learn *which* products are the closest substitutes.

Despite these favorable examples, our approach does not always accurately capture substitution. Panel C shows an example where the review-based model misidentifies the closest substitutes for the fantasy book *The Ashes & The Star-Cursed King*. In fact, all three models fail, incorrectly predicting mystery and self-help books as the closest substitutes. These deviations from observed second choices suggest that there is not enough variation

Panel A. Predicted Second-Choice Probabilities when First Choice is *Dopamine Detox (S)*

Experimental Data		Plain Logit		Attribute-Based Mixed Logit		Review-Based Mixed Logit	
Book	Prob.	Book	Prob.	Book	Prob.	Book	Prob.
Don't Believe (S)	0.353	Please Tell Me (M)	0.182	Don't Believe (S)	0.221	Don't Believe (S)	0.259
Art of Letting Go (S)	0.249	The Inmate (M)	0.152	The Inmate (M)	0.153	Art of Letting Go (S)	0.176
Please Tell Me (M)	0.112	The Housemaid (M)	0.141	Please Tell Me (M)	0.152	Please Tell Me (M)	0.129
The Inmate (M)	0.094	Don't Believe (S)	0.137	Art of Letting Go (S)	0.135	The Inmate (M)	0.125
The Housemaid (M)	0.057	Serpent & Wings (F)	0.095	The Housemaid (M)	0.132	The Housemaid (M)	0.100
Serpent & Wings (F)	0.042	Art of Letting Go (S)	0.094	Court of Ravens (F)	0.101	Court of Ravens (F)	0.074
Court of Ravens (F)	0.034	Court of Ravens (F)	0.080	Serpent & Wings (F)	0.057	The Ritual (M)	0.049
The Ritual (M)	0.031	Ashes & Star (F)	0.065	The Ritual (M)	0.028	Serpent & Wings (F)	0.044
Ashes & Star (F)	0.030	The Ritual (M)	0.054	Ashes & Star (F)	0.022	Ashes & Star (F)	0.043

Panel B. Predicted Second-Choice Probabilities when First Choice is *Please Tell Me (M)*

Experimental Data		Plain Logit		Attribute-Based Mixed Logit		Review-Based Mixed Logit	
Book	Prob.	Book	Prob.	Book	Prob.	Book	Prob.
The Inmate (M)	0.325	Don't Believe (S)	0.164	The Inmate (M)	0.163	The Housemaid (M)	0.170
The Housemaid (M)	0.250	Dopamine Detox (S)	0.149	The Housemaid (M)	0.149	The Inmate (M)	0.158
Don't Believe (S)	0.088	The Inmate (M)	0.147	Don't Believe (S)	0.145	Don't Believe (S)	0.119
Art of Letting Go (S)	0.066	The Housemaid (M)	0.145	Dopamine Detox (S)	0.117	Serpent & Wings (F)	0.110
Serpent & Wings (F)	0.066	Art of Letting Go (S)	0.119	Art of Letting Go (S)	0.109	Art of Letting Go (S)	0.108
Dopamine Detox (S)	0.063	Serpent & Wings (F)	0.091	Court of Ravens (F)	0.099	Court of Ravens (F)	0.104
Court of Ravens (F)	0.058	Court of Ravens (F)	0.087	Serpent & Wings (F)	0.097	Dopamine Detox (S)	0.096
The Ritual (M)	0.056	The Ritual (M)	0.057	Ashes & Star (F)	0.070	Ashes & Star (F)	0.072
Ashes & Star (F)	0.029	Ashes & Star (F)	0.041	The Ritual (M)	0.050	The Ritual (M)	0.064

Panel C. Predicted Second-Choice Probabilities when First Choice is *Ashes & Star (F)*

Experimental Data		Plain Logit		Attribute-Based Mixed Logit		Review-Based Mixed Logit	
Book	Prob.	Book	Prob.	Book	Prob.	Book	Prob.
Serpent & Wings (F)	0.275	The Housemaid (M)	0.155	Please Tell Me (M)	0.178	The Inmate (M)	0.184
Court of Ravens (F)	0.243	Don't Believe (S)	0.155	The Inmate (M)	0.157	Please Tell Me (M)	0.166
The Inmate (M)	0.105	The Inmate (M)	0.153	The Housemaid (M)	0.134	The Housemaid (M)	0.123
The Ritual (M)	0.082	Please Tell Me (M)	0.139	Serpent & Wings (F)	0.116	Don't Believe (S)	0.110
Please Tell Me (M)	0.080	Dopamine Detox (S)	0.116	The Ritual (M)	0.105	Court of Ravens (F)	0.108
The Housemaid (M)	0.070	Art of Letting Go (S)	0.100	Don't Believe (S)	0.093	Dopamine Detox (S)	0.092
Don't Believe (S)	0.052	Court of Ravens (F)	0.080	Dopamine Detox (S)	0.075	Serpent & Wings (F)	0.090
Art of Letting Go (S)	0.048	Serpent & Wings (F)	0.062	Art of Letting Go (S)	0.075	The Ritual (M)	0.065
Dopamine Detox (S)	0.045	The Ritual (M)	0.041	Court of Ravens (F)	0.066	Art of Letting Go (S)	0.063

Table 1: **Predicted second-choice probabilities and their data counterparts.** Letters in parentheses indicate book genres: F=Fantasy, M=Mystery, and S=Self-Help.

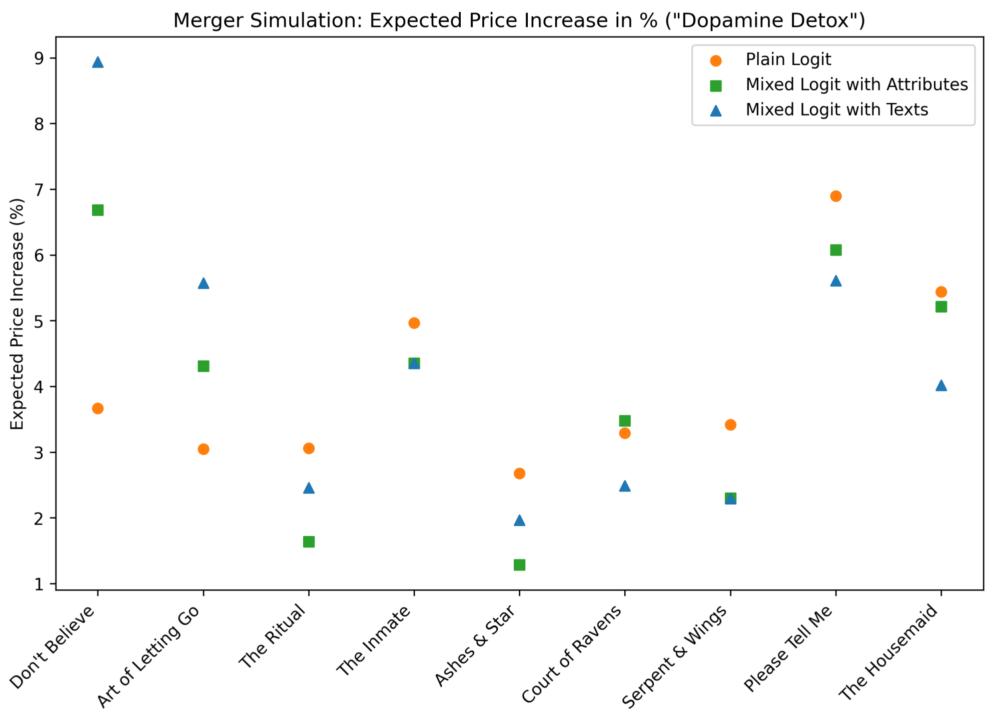

Figure 5: **Results of hypothetical merger simulations.** For each simulated merger of *Dopamine Detox* with one other book, the figure shows the average price increase among the two merging books.

in the first-choice data to reliably estimate substitution patterns for some alternatives.

3.6 Implications for Pricing: Merger Simulations

To illustrate how estimated substitution patterns can influence counterfactuals of interest, we conduct simulations of horizontal mergers—a natural application of our approach. Antitrust agencies routinely use demand models to assess whether a hypothetical merger would lead to a significant price increase (Federal Trade Commission, 2022). If the merging firms' products are close substitutes, the merger creates strong upward pricing pressure, as the firm can "recapture" some consumers after raising prices. Therefore, predicting how a merged firm would set prices requires accurate estimates of substitution patterns.

For each pair of books, we compute their prices under two scenarios: (a) when the books are owned by separate publishers competing in a Bertrand-Nash equilibrium, and (b) when both books are owned by the same publisher, who sets prices by solving a joint first-order condition. In both cases, publishers take the prices of the other eight books as given, fixed at $5.[18]

We recognize that limiting the choice set to ten books makes our analysis somewhat artificial, as major publishers typically manage vast assortments with hundreds of thousands of titles. However, our simulations offer a clear way to evaluate how estimated substitution patterns translate into pricing decisions. This application is also policy-relevant given the high-profile mergers among major book publishers in the past few decades, which have drawn attention and regulatory scrutiny.[19]

Figure 5 illustrates the predicted price increase resulting from the joint ownership of *Dopamine Detox* with each of the other books. We select *Dopamine Detox* as an example where our approach outperforms alternative models in capturing substitution patterns (see Table 1). For each simulated merger, Figure 5 reports the average price increase across the two merging books.

As discussed in Section 3.5, the plain logit model fails to capture strong within-genre substitution. Consequently, it incorrectly identifies as the closest substitutes for *Dopamine Detox* the three mystery books, rather than other self-help ones. This misclassification affects the predicted price changes in merger simulations. Specifically, the plain logit model overstates the price increase when *Dopamine Detox* is merged with the mystery book *Please Tell Me* or *The Housemaid*, while understating the price increase when merged with its true closest substitutes, *The Art of Letting Go* or *Don't Believe Everything You Think*. These discrepancies are substantial: for within-genre mergers, the plain logit predicts modest price increases of 3% and 3.7%, whereas our review-based model estimates them to be 5.6% and 9%—approximately 2 to 2.5 times higher.

The attribute-based mixed logit model produces predictions more aligned with the review-based model. In most cases, both models deviate from the plain logit in the same direction, yielding similar or even identical predicted price increases. However, notable discrepancies remain—for instance, in the case of the self-help books *The Art of Letting Go* or *Don't Believe Everything You Think*, the attribute-based model underestimates the price increase by approximately 25% compared to our selected text-based model.

These divergent predictions could lead decision-makers to different conclusions. As an example, consider an antitrust agency that applies a heuristic rule, challenging all

[18]We fix the prices of the other books at $5, the average value in the experimental dataset. We do not optimize prices of other books because the model does not include an outside option. Consumer choices remain unchanged if all prices increase by the same amount, thus leading to multiple equilibria.

[19]In 2022, a federal judge blocked the proposed merger between *Penguin Random House* and *Simon & Schuster*, citing concerns that it would stifle competition (U.S. Department of Justice, 2022). This merger would have reduced the "Big Five" publishers to the "Big Four" and significantly increased the concentration of the publishing market. Prior to that, *Penguin Books* and *Random House* merged in 2013 to form *Penguin Random House*—the largest trade book publisher globally.

mergers expected to increase prices by more than 5% (Bhattacharya et al., 2023). Compared to decisions made using the review-based model, which best captures substitution patterns, a decision-maker relying on the plain logit model would approve two mergers that should be challenged (*Don't Believe Everything You Think*, *The Art of Letting Go*) and challenge one that should be approved (*The Housemaid*). Similarly, a decision-maker using the attribute-based mixed logit model would approve a merger that should be challenged (*The Art of Letting Go*) and challenge one that should be approved (*The Housemaid*).

4 Application to Online Retail Data

Next, we apply our approach to choice data from several online markets. We pursue two goals with this application. First, we aim to assess the external validity of our experimental results. Although our estimation approach works for books, this does not mean it would capture substitution patterns well in other product categories. By applying this approach across multiple categories, we also show that it can be applied broadly without being tailored to specific markets. Second, by exploring which text or image data best predict substitution patterns in various categories, we can offer practical guidance to empirical researchers on what unstructured data to collect for demand estimation.

4.1 Data

Online Purchases. We use purchase data from the 2019-2020 Comscore Web Behavior Panel, which includes a sample of around two million U.S. households. We use the dataset constructed by Greminger et al. (2023), who classify over 12 million unique products from Amazon.com—browsed or purchased by Comscore panelists—into narrowly defined categories. This dataset also matches purchases with daily product price histories obtained from the third-party database Keepa.com.[20] We focus on Amazon.com due to the high volume of Amazon purchases in the Comscore dataset.

We apply our method to 40 product categories. To construct this sample, we first select the 15 most-purchased products in each category and drop all other products. We then retain only the categories where these 15 products collectively account for at least 2,000 purchases. This selection criterion ensures that we observe a sufficient number of purchases to estimate both product-fixed effects and substitution patterns. Within each category, we treat each purchase as a choice instance and assume that consumers' choice

[20]We do not observe product rankings, so we do not include them in our demand models.

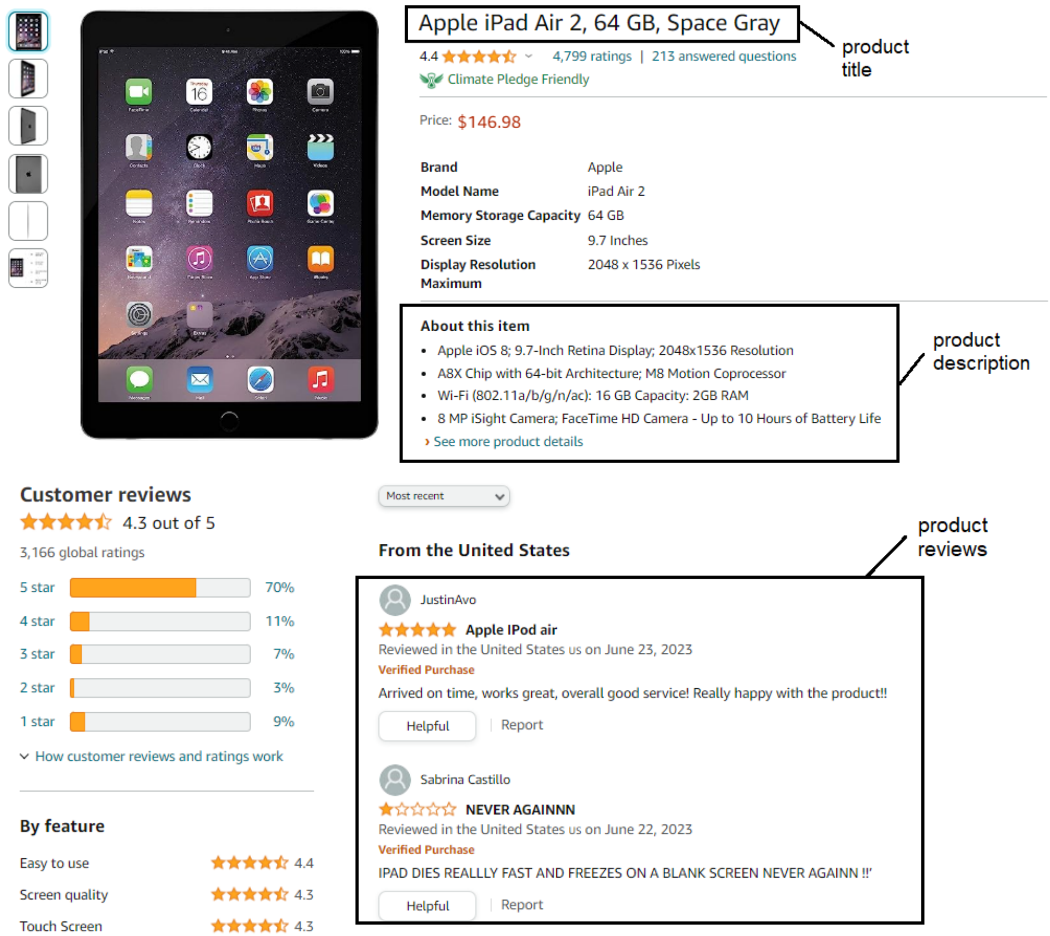

Figure 6: **Example: image and text data collected from product detail pages.**

sets consist of the 15 selected products.[21]

The chosen product categories cover a diverse range of items such as clothing (shirts, blouses, underwear, sleepwear), household goods (paper towels, trash bags, batteries), office supplies (pens, markers, printing paper), groceries (tea, coffee, bottled water), pet food (wet food, treats), electronics (tablets, monitors, headphones, memory cards, media players), and video games (PC, Nintendo, Xbox, PlayStation). The first column of Table 2 lists all included categories. On average, each category contains 662 purchases, with an average product price of $43.

Images and Texts. We collect image and text data directly from product detail pages on Amazon.com. Figure 6 shows a sample product page. For image embeddings, we use the default product photo displayed at the top of the product detail page. For text embed-

[21] We omit products purchased fewer than 10 times and those with missing price data, which means some categories in our sample include fewer than 15 products.

dings, we gather information from several sections: product titles at the top, bullet points describing the product's attributes (which we refer to as the "product description"), and the texts of the 100 most recent reviews for each product.

4.2 Estimation and Results

In each product category, we follow the same approach as in the experiment, extracting embeddings, reducing their dimensionality through PCA, and including them in the random coefficients logit model.

We do not observe product attributes in these categories. Collecting such data is challenging because each category has its own set of choice-relevant attributes. Further, we would need to subjectively decide which attributes to collect in each category—a challenge that originally motivated us to look at images and texts. For these reasons, we focus on comparing our estimates to those from the plain logit model. In other words, we focus on showing that texts and images contain information about product substitution and that our approach can reliably extract this information across many categories.

4.2.1 AIC Improvements

Table 2 summarizes the estimation results, showing the selected model and data type for each category, along with the corresponding AIC improvement relative to the plain logit model. Figure 7 further plots the distribution of AIC improvements across all categories.

As a rule of thumb, we consider that a model has strong support compared to a simpler model if it lowers AIC by at least 2.0, and very strong support if it lowers AIC by at least 5.0. When a simpler model is nested within a more complex one with only one additional parameter, applying these AIC thresholds is approximately equivalent to conducting a likelihood ratio test at the 5% and 1% significance levels.[22]

As Figure 7 shows, we achieve, on average, an AIC reduction of 28.0, with improvements reaching as high as 121.8 in some categories. These results confirm that our approach performs consistently well across a wide range of categories, which span both durable and perishable goods.

[22]When nested models differ by one parameter, the reduction in AIC is given by $\Delta AIC = 2 - \lambda_{LR}$, where λ_{LR} is the likelihood ratio statistic. Therefore, our ΔAIC thresholds of 2.0 and 5.0 correspond to the likelihood ratio thresholds 4.0 and 7.0 (p-values 0.0455 and 0.008).

Category	Data Type	Model Type	Δ AIC	Δ HHI
1. Clothing Active	Descriptions	ST	-30.4	129.0%
2. Clothing Shirts	Reviews	USE	-19.1	116.7%
3. Clothing Underwear	Images	Resnet50	-18.5	88.2%
4. Clothing Sleep	Reviews	ST	-18.5	93.6%
5. Clothing Tops & Blouses	Descriptions	USE	-18.5	94.6%
6. Electronics Cables	Titles	USE	-18.4	240.0%
7. Electronics Accessories	Reviews	TFIDF	-19.8	10.7%
8. Electronics Keyboards	Images	VGG19	-20.0	88.2%
9. Electronics Memory Cards	Images	VGG16	-109.7	103.5%
10. Electronics Tablets	Images	Resnet50	-121.8	120.9%
11. Electronics Monitors	Reviews	USE	-31.6	197.3%
12. Electronics Headphones	Descriptions	COUNT	-26.0	1.2%
13. Electronics Media Players	Reviews	TFIDF	-66.9	42.7%
14. Groceries Water	Descriptions	COUNT	-22.5	123.0%
15. Groceries Coffee	Images	VGG16	-17.2	45.6%
16. Groceries Tea	Titles	ST	-20.0	145.2%
17. Groceries Chips	Reviews	ST	-13.5	61.9%
18. Household Aromatherapy	Titles	COUNT	-19.7	190.7%
19. Household Batteries	Reviews	TFIDF	-20.1	4.8%
20. Household Trash Bags	Titles	ST	-24.6	93.5%
21. Household Paper Towels	Titles	TFIDF	-28.9	188.4%
22. Home Sheets & Pillowcases	Descriptions	USE	-22.3	121.2%
23. Bedroom Beds	Reviews	USE	-17.1	69.4%
24. Bedroom Mattresses	Titles	USE	-17.4	66.3%
25. Kitchen Food Storage	Images	Resnet50	-27.7	133.2%
26. Office Folders	Titles	USE	-29.7	199.7%
27. Office Paper	Images	Resnet50	-19.4	11.4%
28. Office Markers	Images	VGG19	-22.3	228.7%
29. Office Pens	Titles	ST	-14.4	241.0%
30. Office Printer Supplies	Titles	TFIDF	-23.2	247.7%
31. Pet Cat Food	Titles	ST	-16.9	130.6%
32. Pet Cat Litter	Titles	ST	-21.9	126.9%
33. Pet Cat Snacks	Images	Xception	-19.5	142.3%
34. Pet Dog Food	Titles	COUNT	-19.2	65.6%
35. Pet Dog Treats	Reviews	ST	-22.9	163.9%
36. Game Consoles Nintendo	Titles	COUNT	-13.1	118.3%
37. Video Games Nintendo	Images	Resnet50	-36.7	26.2%
38. Video Games PC	Images	VGG19	-18.6	43.9%
39. Video Games PS4	Titles	TFIDF	-32.1	159.6%
40. Video Games Xbox	Reviews	USE	-40.5	120.6%

Table 2: **Estimation results across 40 categories in Comscore data.** For each category, the table shows the selected model and data type yielding the lowest AIC, the corresponding AIC improvement relative to the plain logit model, and the relative change in the HHI of the predicted diversion ratios relative to the plain logit model.

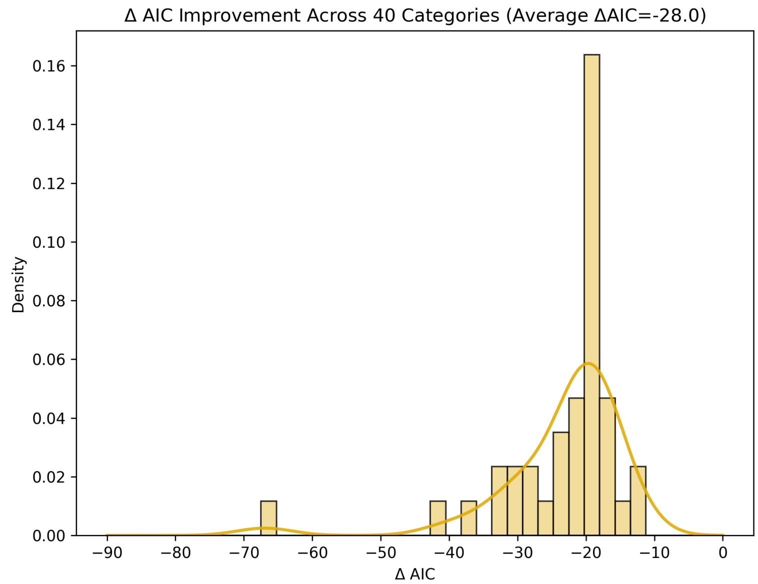

Figure 7: **Fit improvements relative to plain logit across 40 product categories.** For each category, we compute ΔAIC as the AIC of the lowest-AIC specification that includes texts or images minus that of the plain logit model.

4.2.2 Implications for Substitution

Next, we examine how incorporating unstructured data into demand estimation affects our understanding of product substitution. While we do not observe the "ground truth" diversion ratios as we did in the experiment, we can still assess how diversion ratios estimated by our approach deviate from the restrictive patterns imposed by the plain logit model.

The assumption of independent utilities in the plain logit model implies that diversion ratios depend only on market shares and not on product similarities. Consequently, this model often fails to identify close substitutes, leading to overly flat diversion ratios (Conlon et al., 2023). By contrast, if our approach better identifies close substitutes, it should produce more unequal diversion ratios. To quantify this, we compute diversion ratios using the definition from Section 3 and calculate the Herfindahl-Hirschman Index (HHI) of predicted diversion ratios.[23] We expect our approach to yield a higher HHI than the plain logit model.

The last column in Table 2 presents the results.[24] Across categories, our approach

[23] Specifically, we compute the HHI across $(J-1)$ conditional second choice probabilities for each first choice and then average the resulting HHI measure across all J possible first choices.

[24] In some categories, our estimates of the price coefficient α are positive, likely due to correlation between

more than doubles the HHI of diversion ratios, with an average increase of 115%. Additionally, HHI rises in all 40 product categories, with some increases approaching 250%. These findings confirm that our approach produces significantly more variable diversion ratios across the board.

4.2.3 Relevance of Different Data Types

Lastly, we analyze which types of unstructured data yield the largest improvement in fit in each category. It would be misleading to report only the best-performing model in each category, as multiple text or image models may achieve similar AIC improvements over plain logit. We therefore follow the statistical literature on model selection and report Akaike weights for each data type (Burnham and Anderson, 2004). Formally, we define $\Delta_i = AIC_i - AIC_{min}$ for each estimated specification i where AIC_{min} is the lowest AIC across *all* estimated specifications in that product category. We then compute the Akaike weight for a data type $d \in \{\text{images}, \text{titles}, \text{descriptions}, \text{reviews}\}$ as follows:

$$w_d = \frac{\sum_{\{r:d(r)=d\}} \exp(-\Delta_r/2)}{\sum_i \exp(-\Delta_i/2)} \tag{2}$$

The numerator sums over specifications using data type d, while the denominator sums over all estimated specifications. We interpret w_d heuristically as the posterior probability that the mixed logit model based on data type d is the best model given the data.[25]

Figure 8 shows the estimated Akaike weights by category and data type. We find considerable variation across categories in terms of which types of data are most important for predicting substitution patterns.

Importantly, many of these results are difficult to predict ex ante. For example, while we might expect visual product features to be most relevant in clothing categories, where consumers care about visual design, only one of the four clothing categories ("Clothing Underwear") has an image weight exceeding those of texts. By contrast, estimating demand for "Tops & Blouses" requires product descriptions, for "Shirts" relies on customer reviews, and for "Activewear" it works best with either descriptions or reviews. Similarly, while we may expect descriptions and reviews of video games to be as informative

prices and unobserved demand shocks, even after accounting for product fixed effects. While instrumental variables could address this, we do not pursue this approach, as addressing price endogeneity across many product categories is orthogonal to the contribution of our paper. Instead, we focus on counterfactuals, such as the diversion ratios from removing a product, which remain well-defined even when the price coefficient is positive.

[25] More precisely, in large samples, w_d reflects the probability that this class of models is, in fact, the best model for the data in the sense of Kullback-Leibler information (Burnham and Anderson, 2004, p.272).

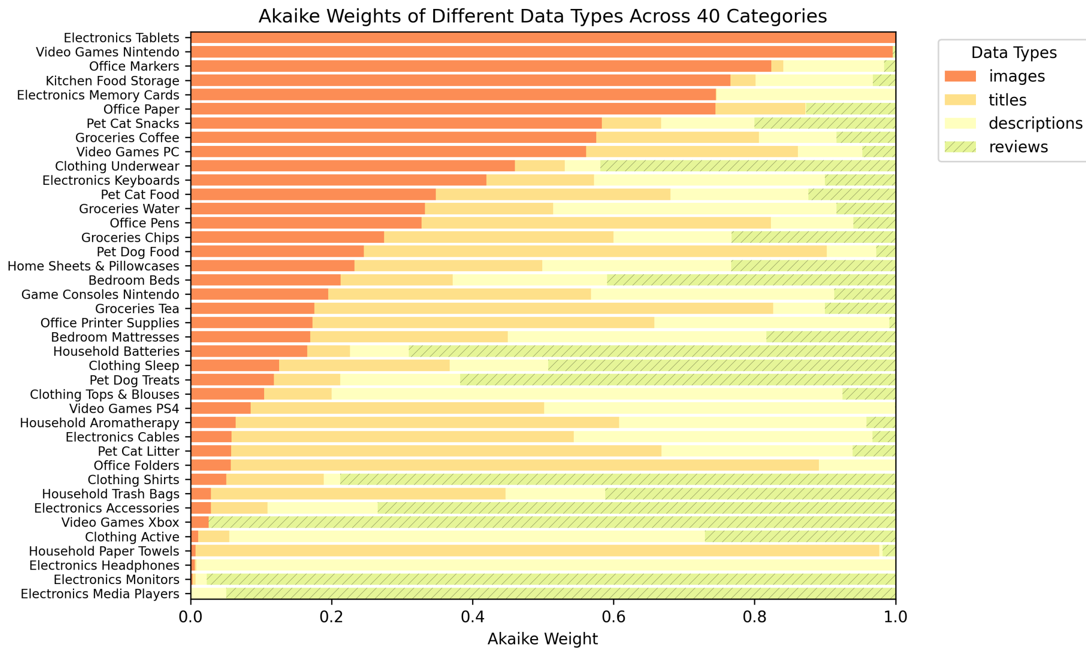

Figure 8: **Akaike Weights of different types of unstructured data across 40 product categories.** These weights reflect the relative importance of data types for predicting substitution patterns in the data.

as those for books, the data strongly suggests that in categories of Nintendo and PC video games, images contain significantly more information about substitution than text.

These results highlight an important lesson: researchers may not be able to predict *ex ante* which data types will best capture substitution.

5 Discussion

5.1 A Practitioner's Guide for Using Text and Images

In this section, we summarize our key results and offer guidelines for researchers on how to incorporate unstructured data into demand estimation.

The main takeaway from our analyses is that text and image data contain information about substitution patterns, meaning these unstructured data can be useful for demand estimation. At the same time, our cross-category analysis in Section 4 highlights that researchers may not be able to predict *ex ante* which data type will best capture substitution in a given product category. Therefore, we recommend that researchers collect various types of unstructured data and perform model selection as we did in Sections 3 and 4. Specifically, researchers can first identify the lowest-AIC set of principal components for each combination of data type and embedding model, and then determine which of these selected specifications minimizes AIC across all data-model combinations. This selection rule performed well in our experiment, where it identified the model with the best counterfactual second-choice predictions.

In our experimental data in Section 3, we found that a text-based model produces the best counterfactual predictions, and that extending this model with observed attributes or image data did not significantly improve model fit. Nevertheless, in other contexts, researchers may still want to check whether incorporating observed product attributes can capture additional dimensions of product differentiation. This can be done by searching for a set of observed attributes that, when added to the selected model, further reduce AIC. A similar search can be conducted over additional text- and image-based principal components to determine whether including them improves model fit.

To conduct inference based on the estimates from our approach, researchers must account for model selection to obtain valid confidence intervals. One potential solution is to perform sample-splitting: researchers can use one randomly selected subset of the data for model selection and the other for estimating the covariance matrix of the estimates.[26]

We mostly abstracted away from the issue of price endogeneity. Endogeneity concerns do not arise in our experiment because we randomize prices, whereas in the analysis of Amazon purchases, we rely on product fixed effects to capture product-level unobservables (e.g., unobserved quality) that may correlate with prices. In practice, researchers are often concerned that prices may be correlated with unobserved demand shocks that vary across markets and over time. To address this, researchers can combine our approach with existing methods for handling price endogeneity based on instrumental variables.

5.2 Future Research Directions

While we are not the first to use text and image data in demand estimation, our paper proposes and validates a concrete method for leveraging such data *to flexibly estimate substitution patterns* – a critical input for answering many empirical questions of interest. We do not claim that ours is the only or the best way to use unstructured data for this purpose. In fact, several questions remain open.

First, other demand modeling approaches might be able to better take advantage of text and image data than the random coefficients logit model. In an earlier version of this paper, we tested a pairwise combinatorial logit model from Koppelman and Wen

[26] For a detailed discussion of sample-splitting, see Wasserman and Roeder (2009). Taylor and Tibshirani (2015) provide examples of how researchers can construct corrected p-values and confidence intervals for specific estimation problems.

(2000), which generalizes the logit model by allowing each product pair to have its own correlation in utilities. We allowed these correlations to depend on the distance between products in the space of text and image embeddings. This model performed significantly worse than our current approach of incorporating principal components into the standard random coefficients logit model, which highlights how the functional form assumptions of the demand model can significantly impact counterfactual performance.[27]

Second, our approach can be easily extended to newer, more advanced ML models that might better extract substitution patterns from texts and images. The approach we propose in this paper is modular in the sense that researchers can easily try alternative embeddings and test whether they improve counterfactual second-choice predictions. To facilitate such tests, we make our experimental dataset and estimation code publicly available.[28]

Third, researchers could improve upon our approach by constructing "optimal" embeddings specifically designed to improve the counterfactual predictions of interest.[29] Our current method relies on text and image models that were trained on datasets unrelated to our specific application, meaning they may not capture the most relevant dimensions of product differentiation. This suggests potential for improvement: researchers could fine-tune existing deep learning models or train custom ones to optimize for a specific objective, such as improving counterfactual predictions. It would be interesting to assess how much these tailored demand models improve upon pre-trained models and whether they can do so without significantly increasing the computational burden.

Lastly, our approach could extend to other forms of unstructured data, such as music or videos, potentially opening doors to studying product differentiation and competition in markets where this would otherwise be challenging. One could also leverage tools from the explainable AI literature to construct interpretable embeddings, which would help tackle questions such as optimal product design.[30]

[27] Researchers can also try relaxing the parametric assumptions of standard demand models. For example, with small number of products, one could incorporate principal components into a nonparametric demand model (Compiani, 2022).

[28] Replication codes and experimental data are available in our public repository: github.com/ilyamorozov/DeepLogitReplication. Separately, the Python package for implementing our method is available on PyPI and at: github.com/deep-logit-demand/deeplogit.

[29] For example, Lee (2024) shows that fine-tuning LLMs helps researchers to better capture brand intercepts.

[30] See Sisodia et al. (2024) for a detailed discussion of related ideas.

6 Conclusion

In this paper, we demonstrated how researchers can incorporate unstructured text and image data in demand estimation to better recover substitution patterns. Our proposed estimation approach extracts low-dimensional features from product images and textual descriptions, integrating them into a standard discrete choice model. Using experimental data, we showed that our model outperforms standard attribute-based models in counterfactual predictions of second choices. We further validated our method with e-commerce data across dozens of categories and found that text and image data consistently help identify close substitutes within each category.

References

Akaike, H. (1998): "Information theory and an extension of the maximum likelihood principle," in *Selected papers of hirotugu akaike*, Springer, 199–213.

Allcott, H., B. B. Lockwood, and D. Taubinsky (2019): "Regressive sin taxes, with an application to the optimal soda tax," *The Quarterly Journal of Economics*, 134, 1557–1626.

Arteaga, C., J. Park, P. B. Beeramoole, and A. Paz (2022): "xlogit: An open-source Python package for GPU-accelerated estimation of Mixed Logit models," *Journal of Choice Modelling*, 42, 100339.

Backus, M., C. Conlon, and M. Sinkinson (2021): "Common Ownership and Competition in the Ready-To-Eat Cereal Industr," *Working Paper*.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841–890.

——— (1999): "Voluntary export restraints on automobiles: Evaluating a trade policy," *American Economic Review*, 89, 400–431.

——— (2004): "Differentiated Products Demand System from a Combination of Micro and Macro Data: The New Car Market," *Journal of Political Economy*, 112, 68–105.

Berry, S. T. and P. A. Haile (2014): "Identification in differentiated products markets using market level data," *Econometrica*, 82, 1749–1797.

Bhattacharya, V., G. Illanes, and D. Stillerman (2023): "Merger Effects and Antitrust Enforcement: Evidence from US Consumer Packaged Goods," Tech. rep., National Bureau of Economic Research.

Bresnahan, T. F., S. Stern, and M. Trajtenberg (1997): "Market Segmentation and the Sources of Rents from Innovation: Personal Computers inthe Late 1980s," *RAND Journal of Economics*, 28, S17–S44.

Burnham, K. P. and D. R. Anderson (2004): "Multimodel inference: understanding AIC and BIC in model selection," *Sociological methods & research*, 33, 261–304.

Cer, D., M. Diab, E. Agirre, I. Lopez-Gazpio, and L. Specia (2017): "Semeval-2017 task 1: Semantic textual similarity-multilingual and cross-lingual focused evaluation," *arXiv preprint arXiv:1708.00055*.

Cer, D., Y. Yang, S. yi Kong, N. Hua, N. Limtiaco, R. S. John, N. Constant, M. Guajardo-Cespedes, S. Yuan, C. Tar, Y.-H. Sung, B. Strope, and R. Kurzweil (2018): "Universal Sentence Encoder," .

Chintagunta, P. and W. Li (2024): "Product Attributes, Cross Elasticities and Dynamic Market Structure in a Category with Many Products," *Working Paper*.

Chollet, F. (2017): "Xception: Deep Learning With Depthwise Separable Convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1251–1258.

Compiani, G. (2022): "Market counterfactuals and the specification of multiproduct demand: A nonparametric approach," *Quantitative Economics*, 13, 545–591.

Conlon, C., J. Mortimer, and P. Sarkis (2022): "Estimating preferences and substitution patterns from second choice data alone," Tech. rep., working paper.

——— (2023): "Estimating preferences and substitution patterns from second choice data alone," Tech. rep., working paper.

Conlon, C. and J. H. Mortimer (2021): "Empirical properties of diversion ratios," *The RAND Journal of Economics*, 52, 693–726.

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova (2018): "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*.

Dew, R. (2024): "Adaptive preference measurement with unstructured data," *Management Science*.

Döpper, H., A. MacKay, N. Miller, and J. Stiebale (2024): "Rising markups and the role of consumer preferences," Tech. rep., National Bureau of Economic Research.

Dotson, J. P., M. A. Beltramo, E. M. Feit, and R. C. Smith (2019): "Modeling the Effect of Images on Product Choices," Working Paper.

Dotson, J. P., J. R. Howell, J. D. Brazell, T. Otter, P. J. Lenk, S. MacEachern, and G. M. Allenby (2018): "A Probit Model with Structured Covariance for Similarity Effects and Source of Volume Calculations," *Journal of Marketing Research*, 55, 35–47.

Federal Trade Commission (2022): "Demand System Estimation and Its Application to Horizontal Merger Analysis," Tech. rep., Federal Trade Commission, accessed: February 12, 2025.

Goldberg, P. K. (1995): "Product differentiation and oligopoly in international markets: The case of the US automobile industry," *Econometrica: Journal of the Econometric Society*, 891–951.

Greminger, R., Y. Huang, and I. Morozov (2023): "Make Every Second Count: Time Allocation in Online Shopping," *Working Paper*.

Hausman, J. A. (1994): *Valuation of new goods under perfect and imperfect competition*, National Bureau of Economic Research Cambridge, Mass., USA.

He, K., X. Zhang, S. Ren, and J. Sun (2016): "Deep Residual Learning for Image Recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778.

Hitsch, G. J., A. Hortacsu, and X. Lin (2021): "Prices and promotions in US retail markets," *Quantitative Marketing and Economics*, 19, 289–368.

Hoch, S. J., B.-D. Kim, A. L. Montgomery, and P. E. Rossi (1995): "Determinants of store-level price elasticity," *Journal of marketing Research*, 32, 17–29.

Koppelman, F. S. and C.-H. Wen (2000): "The paired combinatorial logit model: properties, estimation and application," *Transportation Research Part B: Methodological*, 34, 75–89.

Lancaster, K. J. (1966): "A new approach to consumer theory," *Journal of political economy*, 74, 132–157.

Lee, K. (2024): "Generative Brand Choice," .

Magnolfi, L., J. McClure, and A. Sorensen (2022): "Triplet Embeddings for Demand Estimation," Working Paper.

McFadden, D. (1974a): "Conditional Logit Analysis of Qualitative Choice Behavior," *Frontiers in Econometrics*.

——— (1974b): "The measurement of urban travel demand," *Journal of public economics*, 3, 303–328.

McFadden, D. and K. Train (2000): "Mixed MNL models for discrete response," *Journal of applied Econometrics*, 15, 447–470.

Netzer, O., R. Feldman, J. Goldenberg, and M. Fresko (2012): "Mine Your Own Business: Market-Structure Surveillance Through Text Mining," *Marketing Science*, 31, 521–543.

Nevo, A. (2000): "Mergers with differentiated products: The case of the ready-to-eat cereal industry," *The RAND Journal of Economics*, 395–421.

——— (2001): "Measuring market power in the ready-to-eat cereal industry," *Econometrica*, 69, 307–342.

Petrin, A. (2002): "Quantifying the benefits of new products: The case of the minivan," *Journal of political Economy*, 110, 705–729.

Pinkse, J. and M. E. Slade (2004): "Mergers, brand competition, and the price of a pint," *European Economic Review*, 48, 617–643.

Quan, T. W. and K. R. Williams (2019): "Extracting Characteristics from Product Images and its Application to Demand Estimation," *University of Georgia, Department of Economics*.

Reimers, N. and I. Gurevych (2019): "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics.

Russakovsky, O., J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei (2015): "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision*, 115, 211–252.

Seiler, S., A. Tuchman, and S. Yao (2021): "The impact of soda taxes: Pass-through, tax avoidance, and nutritional effects," *Journal of Marketing Research*, 58, 22–49.

Shapiro, C. (1995): "Mergers with differentiated products," *Antitrust*, 10, 23.

Simonyan, K. and A. Zisserman (2015): "Very Deep Convolutional Networks for Large-Scale Image Recognition." in *International Conference on Learning Representations*.

Sisodia, A., A. Burnap, and V. Kumar (2024): "Generative Interpretable Visual Design: Using Disentanglement for Visual Conjoint Analysis," *Journal of Marketing Research*, forthcoming.

Szegedy, C., V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna (2016): "Rethinking the Inception Architecture for Computer Vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2818–2826.

Taylor, J. and R. J. Tibshirani (2015): "Statistical learning and selective inference," *Proceedings of the National Academy of Sciences*, 112, 7629–7634.

U.S. Department of Justice (2022): "Justice Department Obtains Permanent Injunction Blocking Penguin Random House's Proposed Acquisition of Simon & Schuster," Accessed: February 9, 2025.

Wasserman, L. and K. Roeder (2009): "High dimensional variable selection," *Annals of statistics*, 37, 2178.

Online Appendix

A Text Processing Steps

Before applying our text models, listed in Section 2.1.2, we pre-process our text data as follows. Working with product titles is straightforward because each title is a short text that typically includes only 10-15 words. When assessing similarity based on product descriptions, we merge the text from all bullet points, and apply our models to the merged text. For customer reviews, we transform the text of each review into a separate vector of word occurrences or an embedding, and we average these vectors or embeddings across all reviews of a given product.

For both bag-of-words approaches, we further pre-process text data by removing stopwords and lemmatizing words. We remove stopwords using the standard dictionary of common English words in the NLTK package. We lemmatize words using the WordNet Lemmatizer from the same package, NLTK. Then, we convert each pre-processed text into a vector of word occurrences (weighted word occurrences for TF-IDF). The other two models, USE and ST, have a built-in text pre-processing step. We therefore apply these models directly to the unprocessed text data.

B Choice Survey: Sanity Checks

Because choices in our experiment were hypothetical and not incentivized, we want to verify that participants made meaningful choices. Several summary statistics suggest that participants took the choice tasks seriously. First, only 50 of the initially recruited participants (less than 1%) completed the entire study in less than one minute and thus had to be dropped from our sample. The remaining participants spent, on average, 7 minutes on the survey overall and 1.3 minutes on the choice tasks, indicating they took their time to make careful selections and did not mindlessly click through the survey (Figure A1).

Second, in the choice tasks, participants were disproportionately more likely to select books of the genre they reported to be their favorite in a questionnaire before the choice tasks (Figure A2), suggesting they considered the book attributes when making their choices.

Finally, participants' choices were consistent across the two choice tasks (Figure A3). For example, over 60% of participants who selected a mystery book in the first task chose another mystery book in the second choice task. This observation suggests participants considered their genre preferences when making both choices.

C Changes in Text and Images Over Time

As discussed in Section 2.3, we treat text and image embeddings, as well as the principal components extracted from them, as being fixed over time for a given product. To validate this assumption, we examine whether text and images change over time.

Since we lack data on these changes for 2019-2020, we construct a separate sample by repeatedly collecting unstructured data from Amazon's product detail pages daily from January 23 to March 4, 2025. To keep data collection manageable, we do not gather customer reviews and select a subset of 11 out of 40 categories, ensuring they cover all of Amazon's departments (e.g., "Clothing," "Food," "Electronics") represented in the full dataset.[31] The selected categories are: "Shirts," "Coffee," "Aromatherapy," "Mattresses," "Markers," "Pet Litters," "Nintendo Games," "Tablets," "Memory Cards," "Monitors," and "Earbuds." In each category, we collect data for the products used in our estimation in Section 4 that were not discontinued, totaling 136 products.

We find that product images do not change over time. Titles change for only six products (4%), mostly by adding or removing specific attributes or functional benefits. Similarly, descriptions change for just 21 products (15%). Most changes do not affect which product features are revealed, but they alter which ones are immediately visible versus being revealed only after clicking on "See more product details." Thus, we conclude that changes in text and images are not a significant concern for the products in our empirical application.

[31] Recall that we average over embeddings extracted from the 100 most recent reviews. Even though customer reviews accumulate over time as consumers continue to write them, the average embeddings extracted from these reviews may remain approximately constant if consumers consistently discuss the same attributes. While we do not have review data to verify this claim, future research should examine whether review embeddings are stable in online markets.

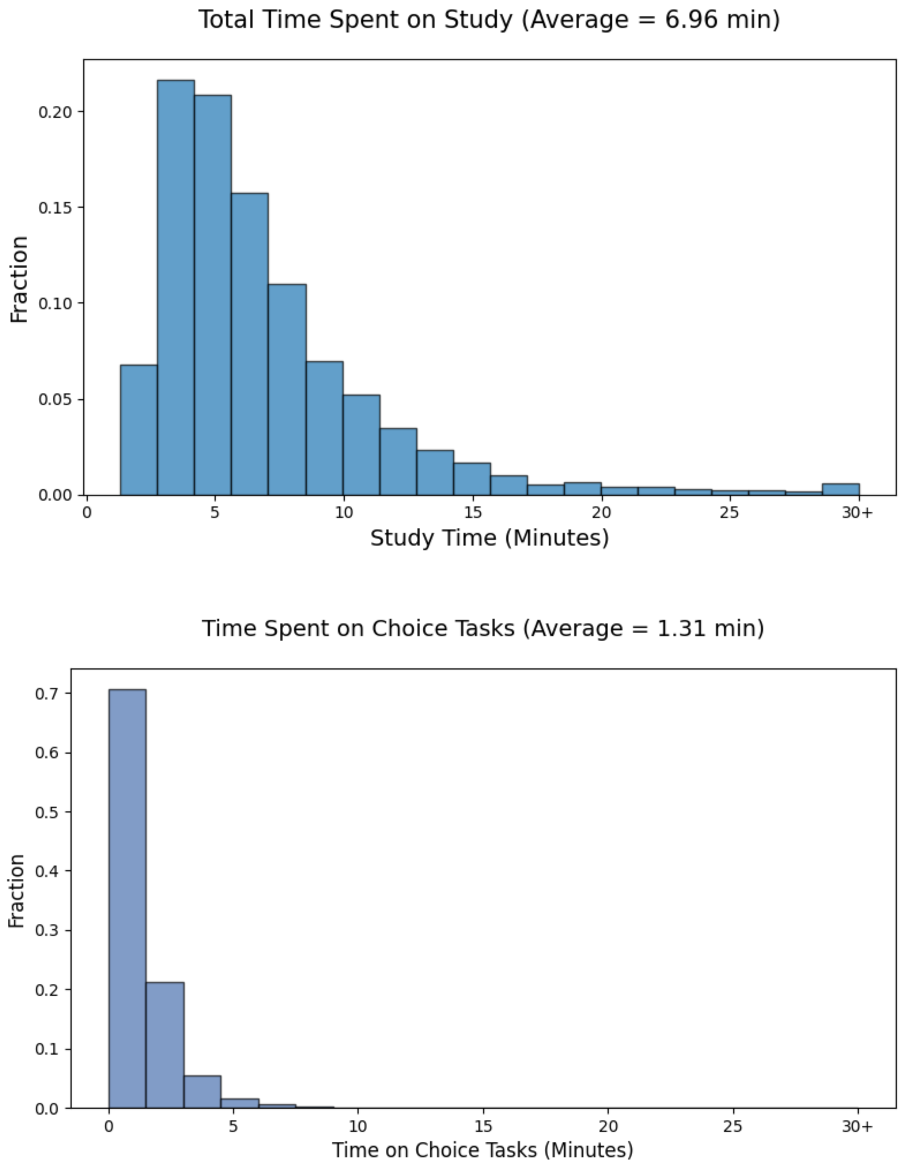

Figure A1: **Time spent by participants on choice tasks in the experiment.**

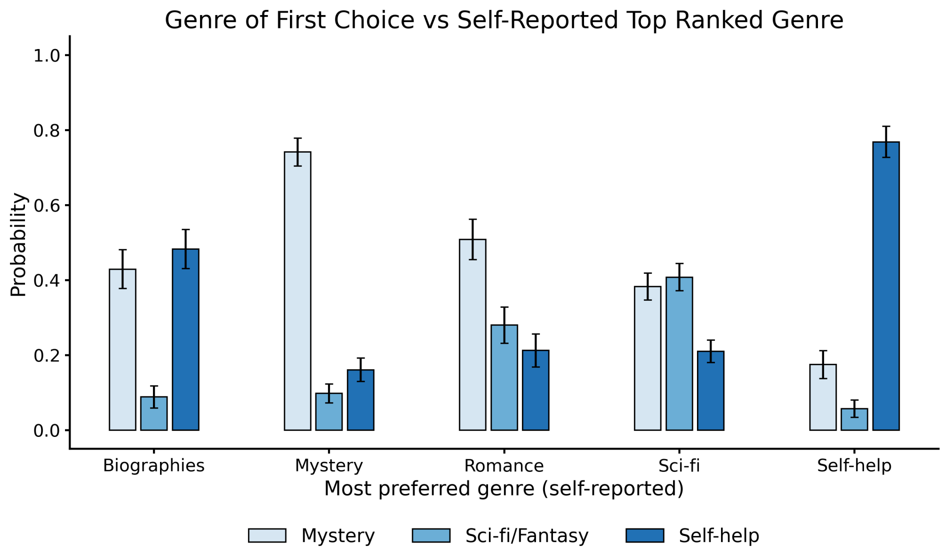

Figure A2: **Selected genres and self-reported genre preferences in the experiment.**

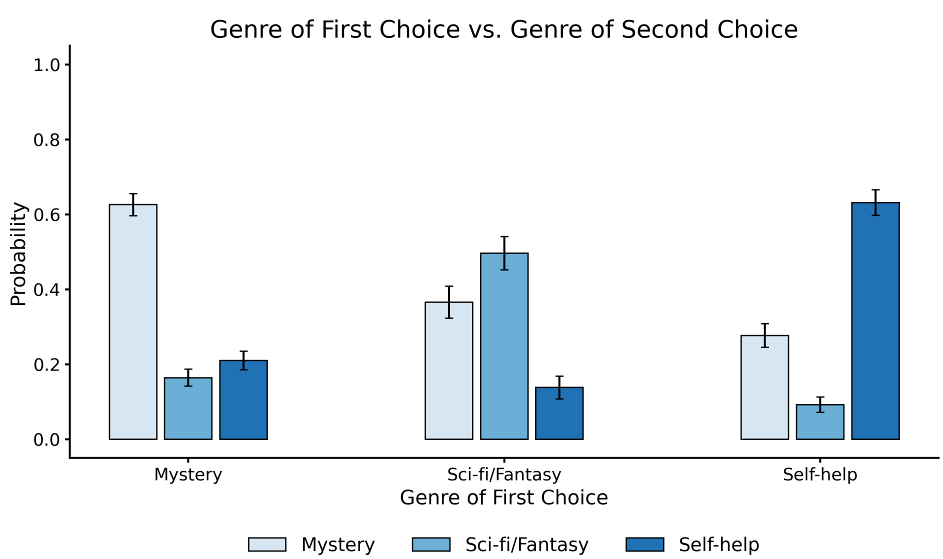

Figure A3: **Genres of participants' first and second choices in the experiment.**

Model	First Choices (Data)		Second Choices (Counterf.)		
	$\log L$	*AIC*	*RMSE*	Rel. to Plain Logit Δ	%
Panel A. Benchmark Model					
Plain Logit	-20488.4	41006.7	0.089		
Panel B. Mixed Logit Models with Principal Components					
Images: InceptionV3	-20475.8	40985.6	0.080	-0.009	-9.8%
Images: ResNet50	-20479.9	40991.8	0.080	-0.009	-9.8%
Images: VGG19	-20481.2	40994.4	0.083	-0.006	-6.8%
Images: Xception	-20479.1	40990.3	0.086	-0.003	-3.1%
Product Titles: Bag-of-Words Count	-20479.3	40992.7	0.085	-0.004	-4.6%
Product Titles: Bag-of-Words TF-IDF	-20479.3	40992.7	0.085	-0.004	-4.6%
Product Titles: Sentence Encoder (USE)	-20479.1	40992.2	0.080	-0.009	-10.5%
Product Titles: Sentence Transformer (ST)	-20477.2	40986.5	0.082	-0.006	-7.1%
Descriptions: Bag-of-Words Count	-20480.1	40992.2	0.089	-0.000	-0.2%
Descriptions: Bag-of-Words TF-IDF	-20480.2	40992.3	0.089	-0.000	-0.2%
Descriptions: Sentence Encoder (USE)	-20476.2	40984.4	0.075	-0.014	-15.5%
Descriptions: Sentence Transformer (ST)	-20475.8	40983.6	0.073	-0.016	-18.0%
Reviews: Bag-of-Words Count	-20481.0	40995.9	0.084	-0.005	-5.2%
Reviews: Bag-of-Words TF-IDF	-20480.9	40995.7	0.084	-0.005	-5.1%
Reviews: Sentence Encoder (USE)	-20472.0	40975.9	0.069	-0.020	-22.3%
Reviews: Sentence Transformer (ST)	-20473.3	40978.6	0.072	-0.017	-18.8%
Panel C. Attribute-Based Mixed Logit Models					
Price	-20485.9	41003.9	0.088	-0.001	-0.8%
Pages	-20484.9	41001.7	0.083	-0.006	-6.8%
Year	-20484.7	41001.4	0.087	-0.002	-2.0%
Genre	-20483.3	41000.7	0.081	-0.008	-8.9%
Price & Pages	-20478.9	40991.9	0.077	-0.012	-13.0%
Price & Year	-20484.1	41002.2	0.088	-0.001	-1.4%
Price & Genre	-20483.2	41002.5	0.080	-0.008	-9.4%
Pages & Year	-20478.3	40990.7	0.077	-0.011	-12.9%
Pages & Genre	-20481.6	40999.2	0.076	-0.013	-14.9%
Year & Genre	-20483.3	41000.7	0.081	-0.008	-8.9%
Price, Pages, & Year	-20478.3	40990.7	0.077	-0.011	-12.9%
Price, Pages, & Genre	-20478.9	40991.9	0.077	-0.012	-13.0%
Price, Year, & Genre	-20482.2	41002.4	0.079	-0.010	-11.2%
Pages, Year, & Genre	-20478.3	40990.7	0.077	-0.011	-12.9%
Price, Pages, Year, & Genre (All Attr.)	-20476.9	40993.9	0.076	-0.013	-14.4%

Table A1: **Model validation results.** The table shows in-sample fit on first-choice data and counterfactual performance on second-choice data for all specifications considered in Figures 3 and 4 (see Section 3.2 for detailed description of these models).