

Demand Estimation with Text and Image Data

Giovanni Compiani (University of Chicago)

Ilya Morozov (Northwestern University)

Stephan Seiler (Imperial College London)

June 2025

Demand Estimation

- ▶ **Demand estimation** is a key building block in many social sciences
 - A key goal: estimate **substitution patterns in response to price changes**
- ▶ For example, take tablets
- ▶ How does the demand for iPad 9 change as the price of Fire 7 varies?
- ▶ This is needed for (among others):
 - tariff evaluation
 - merger analysis
 - markup estimation
 - optimal pricing
- ▶ Important for both equilibrium prices and welfare

Characteristics-Based Approach

► Traditional approach (BLP, 1995):

- Utility from product j : $u_{ij} = -\alpha_i p_j + x_j' \beta_i + \xi_j + \varepsilon_{ij}$
- Substitution patterns are driven by heterogeneity (α_i, β_i) in preferences for p and x
- *Example: If price goes up for Fire 7, consumers who prefer this brand will substitute to other Fire tablets*

► Challenges:

1. Hard to quantify some product features (e.g., visual appearance, user-friendliness)
2. Subjective choice of which and how many attributes to include
3. Need to choose and collect data on attributes for each category under consideration

Challenge with Scalability

"Including more product characteristics would allow for the inclusion of random coefficients on these characteristics and yield more flexible cross-price elasticities. 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."

Rising Markups and the Role of Consumer Preferences

Döpper, MacKay, Miller and Stiebale, JPE (forthcoming)

Texts and Images as Alternative Data



Newshows Women's 2024 Summer Short Sleeve Casual Midi Dresses V Neck Floral Sundress Faux Wrap Dress

Product description:

- Soft & Stretchy: This dress is made of soft and stretchy material. The floral print of this summer dress is vibrant.
- Summer Dresses For Women 2024: Featuring A-line silhouette, knee length dress, faux wrap, V-neck, floral sundresses for women, casual swing dress.
- Has 2 side pockets that are enough to hold essentials
- Features: Short sleeves dress, summer dress with pockets, midi wrap dress, a line dresses, v neck dress, midi dresses for women. You can pair this floral dress with high heels or sandals.
- Suitable For Various Occasions: Shopping, travel, vacation outfits for women, summer beach dress, cocktail party, casual dress, banquet, dating, prom.



OUGES Women's Summer Short Sleeve V-Neck Floral Short Party Dress

Product description:

- Short sleeve, V-neck, A-line, floral patterned with both side pockets design, breathable, skin-touch, makes you feeling relaxed and comfortable.
- This V-neck pattern dress cut simply and knee length, with simple wrap waist design, and useful for nearly any occasion.
- Match tips: Shoulder Bag; High Heels; Different types of shoes
- The vivid pattern of this V-neck pattern dress is minimal to call attention to the wearer without distracting.
- The V-neck pattern dress will great to worn with pumps for daytime business or with more ornate jewelry and accessories for holiday party.

Texts and Images as Alternative Data



Newshows Women's 2024 Summer Short Sleeve Casual Midi Dresses V Neck Floral Sundress Faux Wrap Dress

Product description:

- Soft & Stretchy: This dress is made of soft and stretchy material. The floral print of this summer dress is vibrant.
- Summer Dresses For Women 2024: Featuring A-line silhouette, knee length dress, faux wrap, V-neck, floral sundresses for women, casual swing dress.
- Has 2 side pockets that are enough to hold essentials
- Features: Short sleeves dress, summer dress with pockets, midi wrap dress, a line dresses, v neck dress, midi dresses for women. You can pair this floral dress with high heels or sandals.
- Suitable For Various Occasions: Shopping, travel, vacation outfits for women, summer beach dress, cocktail party, casual dress, banquet, dating, prom.



OUGES Women's Summer Short Sleeve V-Neck Floral Short Party Dress

Product description:

- Short sleeve, V-neck, A-line, floral patterned with both side pockets design, breathable, skin-touch, makes you feeling relaxed and comfortable.
- This V-neck pattern dress cut simply and knee length, with simple wrap waist design, and useful for nearly any occasion.
- Match tips: Shoulder Bag; High Heels; Different types of shoes
- The vivid pattern of this V-neck pattern dress is minimal to call attention to the wearer without distracting.
- The V-neck pattern dress will great to worn with pumps for daytime business or with more ornate jewelry and accessories for holiday party.

Texts and Images as Alternative Data



Newshows Women's 2024 Summer Short Sleeve Casual Midi Dresses V Neck Floral Sundress Faux Wrap Dress

Product description:

- Soft & Stretchy: This dress is made of soft and stretchy material. The floral print of this summer dress is vibrant.
- Summer Dresses For Women 2024: Featuring A-line silhouette, knee length dress, faux wrap, V-neck, floral sundresses for women, casual swing dress.
- Has 2 side pockets that are enough to hold essentials
- Features: Short sleeves dress, summer dress with pockets, midi wrap dress, a line dresses, v neck dress, midi dresses for women. You can pair this floral dress with high heels or sandals.
- Suitable For Various Occasions: Shopping, travel, vacation outfits for women, summer beach dress, cocktail party, casual dress, banquet, dating, prom.



OUGES Women's Summer Short Sleeve V-Neck Floral Short Party Dress

Product description:

- Short sleeve, V-neck, A-line, floral patterned with both side pockets design, breathable, skin-touch, makes you feeling relaxed and comfortable.
- This V-neck pattern dress cut simply and knee length, with simple wrap waist design, and useful for nearly any occasion.
- Match tips: Shoulder Bag; High Heels; Different types of shoes
- The vivid pattern of this V-neck pattern dress is minimal to call attention to the wearer without distracting.
- The V-neck pattern dress will great to worn with pumps for daytime business or with more ornate jewelry and accessories for holiday party.

Texts and Images as Alternative Data



Newshows Women's 2024 Summer Short Sleeve Casual Midi Dresses V Neck Floral Sundress Faux Wrap Dress

Product description:

- Soft & Stretchy: This dress is made of soft and stretchy material. The floral print of this summer dress is vibrant.
- Summer Dresses For Women 2024: Featuring A-line silhouette, knee length dress, faux wrap, V-neck, floral sundresses for women, casual swing dress.
- Has 2 side pockets that are enough to hold essentials
- Features: Short sleeves dress, summer dress with pockets, midi wrap dress, a line dresses, v neck dress, midi dresses for women. You can pair this floral dress with high heels or sandals.
- Suitable For Various Occasions: Shopping, travel, vacation outfits for women, summer beach dress, cocktail party, casual dress, banquet, dating, prom.



OUGES Women's Summer Short Sleeve V-Neck Floral Short Party Dress

Product description:

- Short sleeve, V-neck, A-line, floral patterned with both side pockets design, breathable, skin-touch, makes you feeling relaxed and comfortable.
- This V-neck pattern dress cut simply and knee length, with simple wrap waist design, and useful for nearly any occasion.
- Match tips: Shoulder Bag; High Heels; Different types of shoes
- The vivid pattern of this V-neck pattern dress is minimal to call attention to the wearer without distracting.
- The V-neck pattern dress will great to worn with pumps for daytime business or with more ornate jewelry and accessories for holiday party.

Texts and Images as Alternative Data



Newshows Women's 2024 Summer Short Sleeve Casual Midi Dresses V Neck Floral Sundress Faux Wrap Dress

Product description:

- Soft & Stretchy: This dress is made of soft and stretchy material. The floral print of this summer dress is vibrant.
- Summer Dresses For Women 2024: Featuring A-line silhouette, knee length dress, faux wrap, V-neck, floral sundresses for women, casual swing dress.
- **Has 2 side pockets that are enough to hold essentials**
- Features: Short sleeves dress, summer dress with pockets, midi wrap dress, a line dresses, v neck dress, midi dresses for women. **You can pair this floral dress with high heels or sandals.**
- Suitable For Various Occasions: Shopping, travel, vacation outfits for women, summer beach dress, cocktail party, casual dress, banquet, dating, prom.



OUGES Women's Summer Short Sleeve V-Neck Floral Short Party Dress

Product description:

- Short sleeve, V-neck, A-line, floral patterned **with both side pockets design**, breathable, skin-touch, makes you feeling relaxed and comfortable.
- This V-neck pattern dress cut simply and knee length, with simple wrap waist design, and useful for nearly any occasion.
- **Match tips: Shoulder Bag; High Heels; Sandals.**
- The vivid pattern of this V-neck pattern dress is minimal to call attention to the wearer without distracting.
- The V-neck pattern dress will great to worn with pumps for daytime business or with more ornate jewelry and accessories for holiday party.

This Paper

1. **Method.** Demand estimation approach that incorporates texts and images
 - Extract embeddings from text and images via off-the-shelf deep learning models
 - Incorporate embeddings into a micro-founded demand model
 - Python package *DeepLogit*
2. **Validation.** Experiment to validate our approach
 - Survey with choice tasks that measure both first and second choices
 - Show that unstructured data predicts second choices better than standard attributes
3. **Application.** Real-world application to 40 categories on Amazon.com
 - Show that the model predicts intuitive substitution patterns
 - Illustrate scalability of the approach across categories

Empirical Framework

Our Proposal

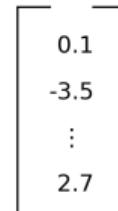
About this item

- Apple iOS 8; 9.7-Inch Retina Display; 2048x1536 Resolution
 - A8X Chip with 64-bit Architecture; M8 Motion Coprocessor
 - Wi-Fi (802.11a/b/g/n/ac); 16 GB Capacity; 2GB RAM
 - 8 MP iSight Camera; FaceTime HD Camera - Up to 10 Hours of Battery Life
- [See more product details](#)



Text & Images

DEEP
LEARNING



Embeddings

ECONOMIC
MODELS



Choices

Data Sources

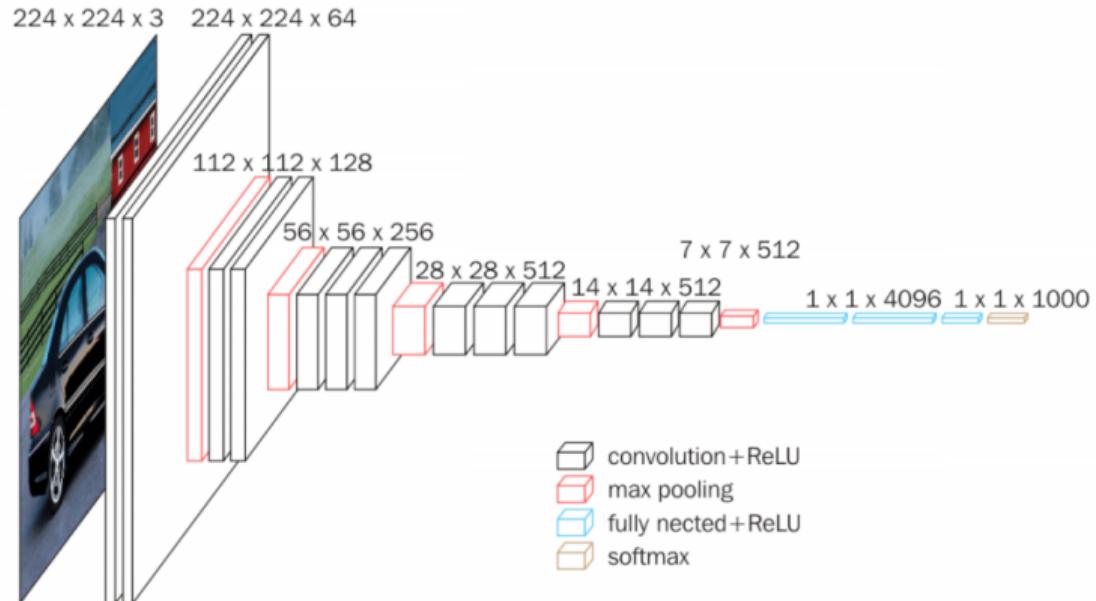
1. **Images:** do products look similar?
 - Aesthetics
2. **Titles and Descriptions:** do sellers describe products similarly?
 - Standard quantifiable attributes
3. **Reviews:** do consumers talk about products in similar ways?
 - User-friendliness, quality, feel, etc

Use Deep Learning to Extract Information from Product Images

- ▶ **Models:** VGG16, VGG19, ResNet50, InceptionV3, Xception
 - Models from *Keras Deep Learning Library*
 - Pre-trained deep neural nets with different depths and architectures
 - Perform well in ImageNet Large Scale Visual Recognition Challenge
- ▶ How we process the data:
 - Models transform images into lower dimensional embeddings
 - Embeddings are then used for classification (to label images)
 - We remove the last step ("classification layer")
 - We work with extracted embeddings e_j

*Similar to how these pre-trained models are used in image-based recommendation algorithms
(Chen et al. 2018, 2019; Deldjoo et al. 2020)*

Extracting embeddings



Extracting embeddings

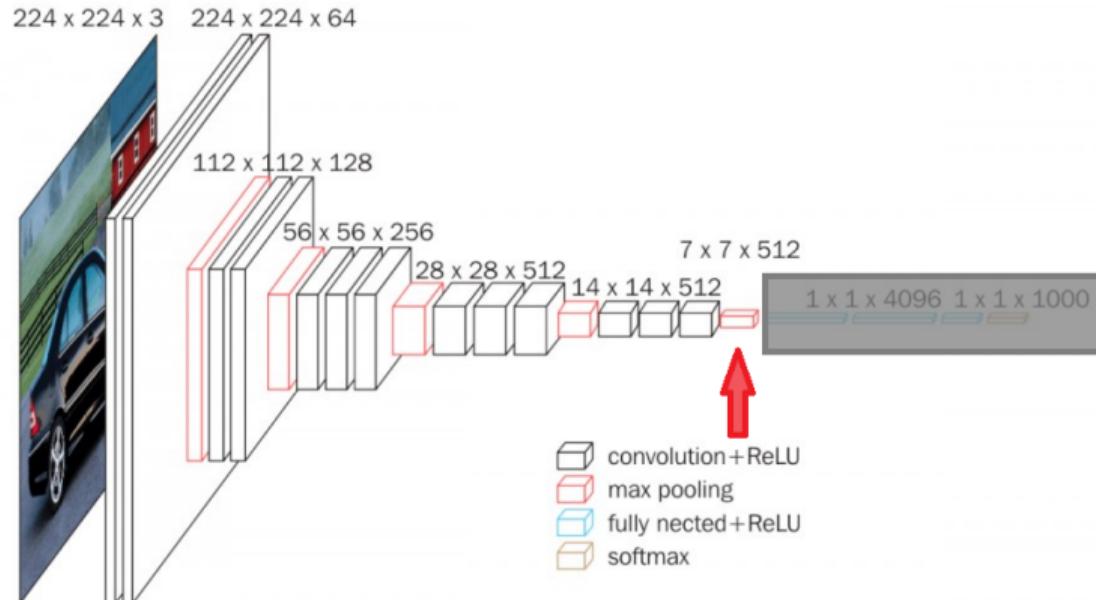


Image Distances: Illustration

Model: VGG19



Extracting Information from Texts

- ▶ **Data sources:** Titles, Descriptions, Customer Reviews
- ▶ **Models:**
 1. Word count model (bag-of-words)
 2. TFIDF count (frequency/uniqueness of words)
 3. Universal Sentence Encoder (USE)
 4. Sentence-BERT Transformer (ST)
- ▶ How we proceed:
 - Pre-process texts
 - Extract embeddings e_j from each text

How To Include Embeddings into Demand Estimation

$$u_{ij} = \underbrace{-\alpha_i p_j}_{\text{price}} + \underbrace{\tilde{e}_j' \beta_i}_{\text{principal components}} + \underbrace{\delta_j}_{\text{product FEs}} + \varepsilon_{ij}$$

Mixed Logit

- ▶ Reduce dimensionality of embeddings e_j via Principal Component Analysis (PCA)
- ▶ Include principal components \tilde{e}_j as “observed attributes” x_j
- ▶ Interact principal components with random coefficients $\beta_i \sim N(\bar{\beta}, \Sigma)$

Model selection

- ▶ Pick the data/ML model combination (e.g., Reviews USE) with lowest in-sample AIC

Python package: *DeepLogit*

How Do We Know This Works?

Experimental Validation

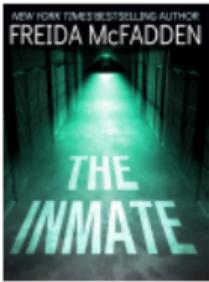
Experiment Design

- ▶ Recruited $N = 9,265$ participants on Prolific
- ▶ Choice task: select one of 10 books
 - Books from Amazon's bestseller list
 - Random prices $p_{ij} \sim U[3, 7]$
 - Random display order
- ▶ Each person performs two choice tasks
 - **Task 1:** unconstrained choice (10 books)
 - **Task 2:** first choice from Task 1 is removed (9 books)
- ▶ We use Task 1 data for estimation and Task 2 data for validation

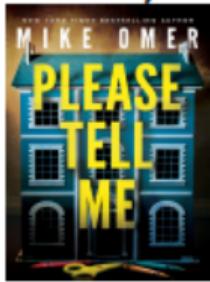


Experiment Design: Overall Choice Set

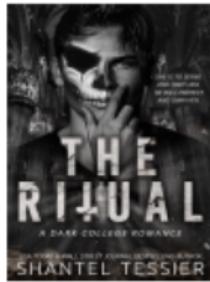
Mystery Books



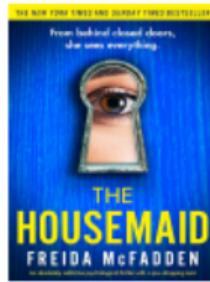
"The Inmate"



"Please Tell Me"

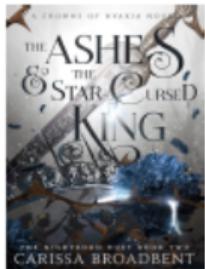


"The Ritual"



"The Housemaid"

Fantasy Books



"The Ashes & the Star
Cursed King"



"Court of Ravens
and Ruin"



"The Serpent &
The Wings of Night"

Self-Help Books



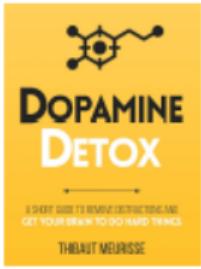
DAMON ZAHARIADES

"The Art of
Letting Go"



SHANTI MEISTER

"Don't Believe
Everything
You Think"



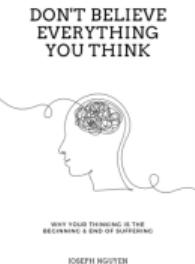
THALITHA酉

"Dopamine
Detox"

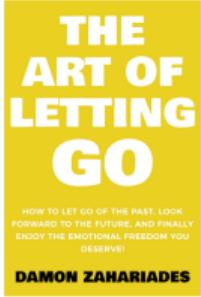
► Self-reported

► First vs second

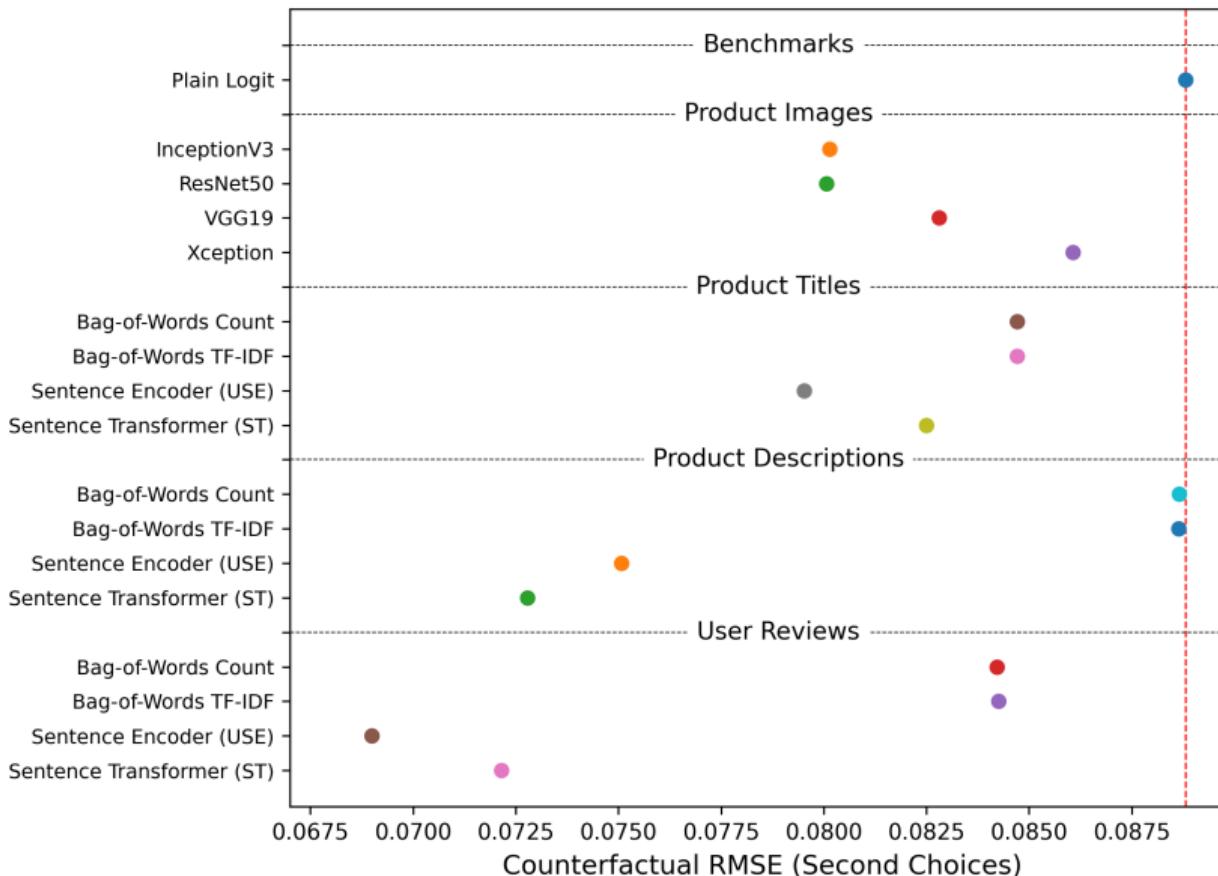
Experiment Design: First Choice Task

Book	Info	Reviews
	<p>Don't Believe Everything You Think by Joseph Nguyen Price: \$3 Genre: Self-Help Year: 2022 Pages: 126</p> <p>Description Learn how to overcome anxiety, self-doubt & self-sabotage without needing to rely on motivation or willpower. In this book, you'll discover the root cause of all psychological and emotional sufferin... read more</p>	<p>Read customer reviews:</p> <p>Review 1 Review 2 Review 3 Review 4 Review 5</p> <p>SELECT THIS BOOK</p>
	<p>The Art of Letting Go by Damon Zahariades Price: \$4 Genre: Self-Help Year: 2022 Pages: 196</p> <p>Description Finally Let Go of Your Negative Thoughts and Enjoy the Emotional Freedom You Deserve! Are you struggling with anger, regrets, and resentment? Do you feel emotionally exhausted, stressed, and discoura... read more</p>	<p>Read customer reviews:</p> <p>Review 1 Review 2 Review 3 Review 4 Review 5</p> <p>SELECT THIS BOOK</p>

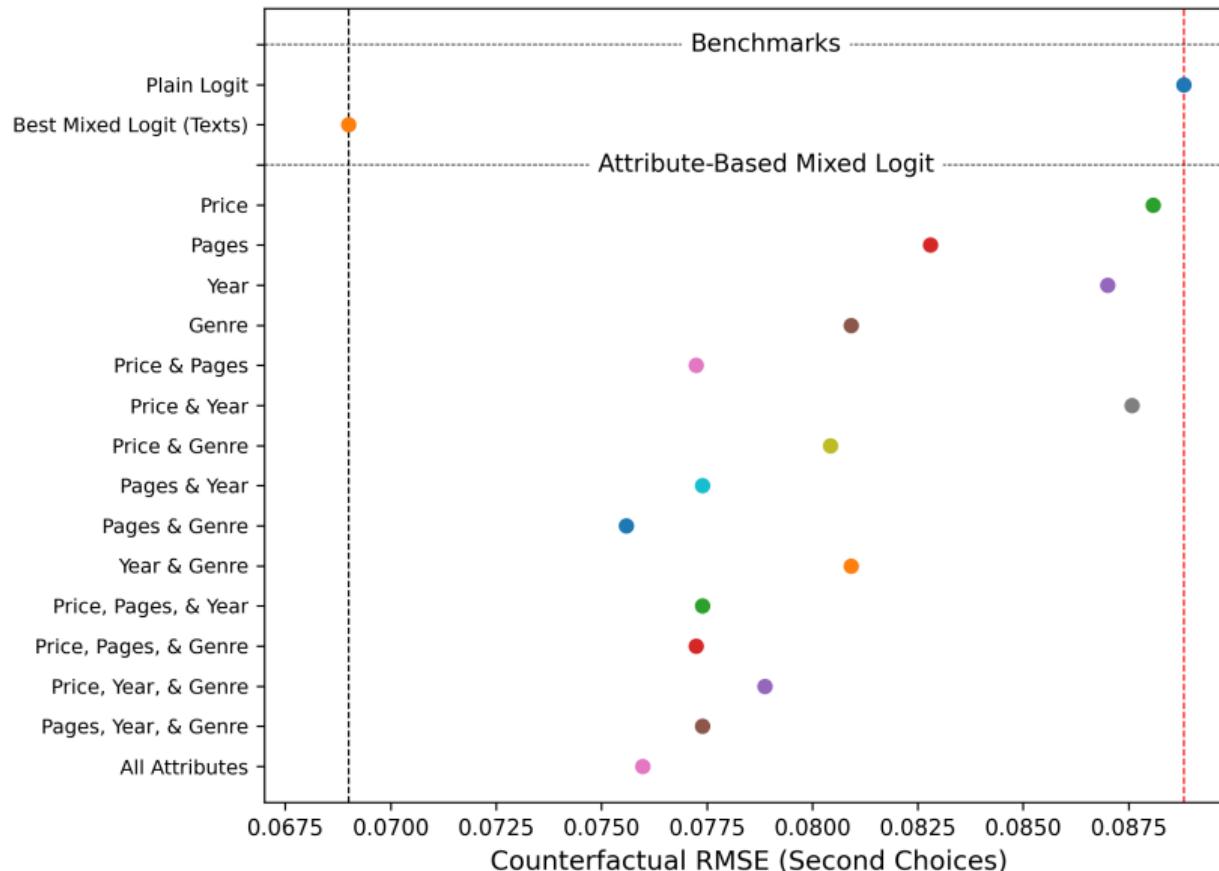
Experiment Design: Second Choice Task

Book	Info	Reviews
	<p>The Art of Letting Go by Damon Zahariades Price: \$4 Genre: Self-Help Year: 2022 Pages: 196</p> <p>Description Finally Let Go of Your Negative Thoughts and Enjoy the Emotional Freedom You Deserve! Are you struggling with anger, regrets, and resentment? Do you feel emotionally exhausted, stressed, and discour... read more</p>	<p>Read customer reviews:</p> <p>Review 1 Review 2 Review 3 Review 4 Review 5</p> <p>SELECT THIS BOOK</p>
	<p>The Ashes & The Star Cursed King by Carissa Broadbent Price: \$5 Genre: Science Fiction & Fantasy Year: 2023 Pages: 626</p> <p>Description Love is a sacrifice at the altar of power. In the wake of the Kejari, everything Oraya once thought to be true has been destroyed. A</p>	<p>Read customer reviews:</p> <p>Review 1 Review 2 Review 3 Review 4 Review 5</p>

Our Approach Improves Over Plain Logit (by 22%)



Our Approach Improves Over Attribute-Based Mixed Logit (by 10%)



Example of Substitution Patterns

First Choice



Dopamine Detox

Mixed Logit

Data



Don't Believe



Art of Letting Go

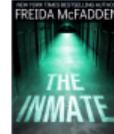


Please Tell Me

Plain Logit



Please Tell Me



The Inmate

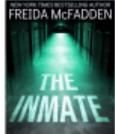


The Housemaid

Attribute-Based



Don't Believe



The Inmate



Please Tell Me

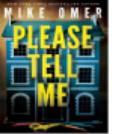
Review-Based



Don't Believe



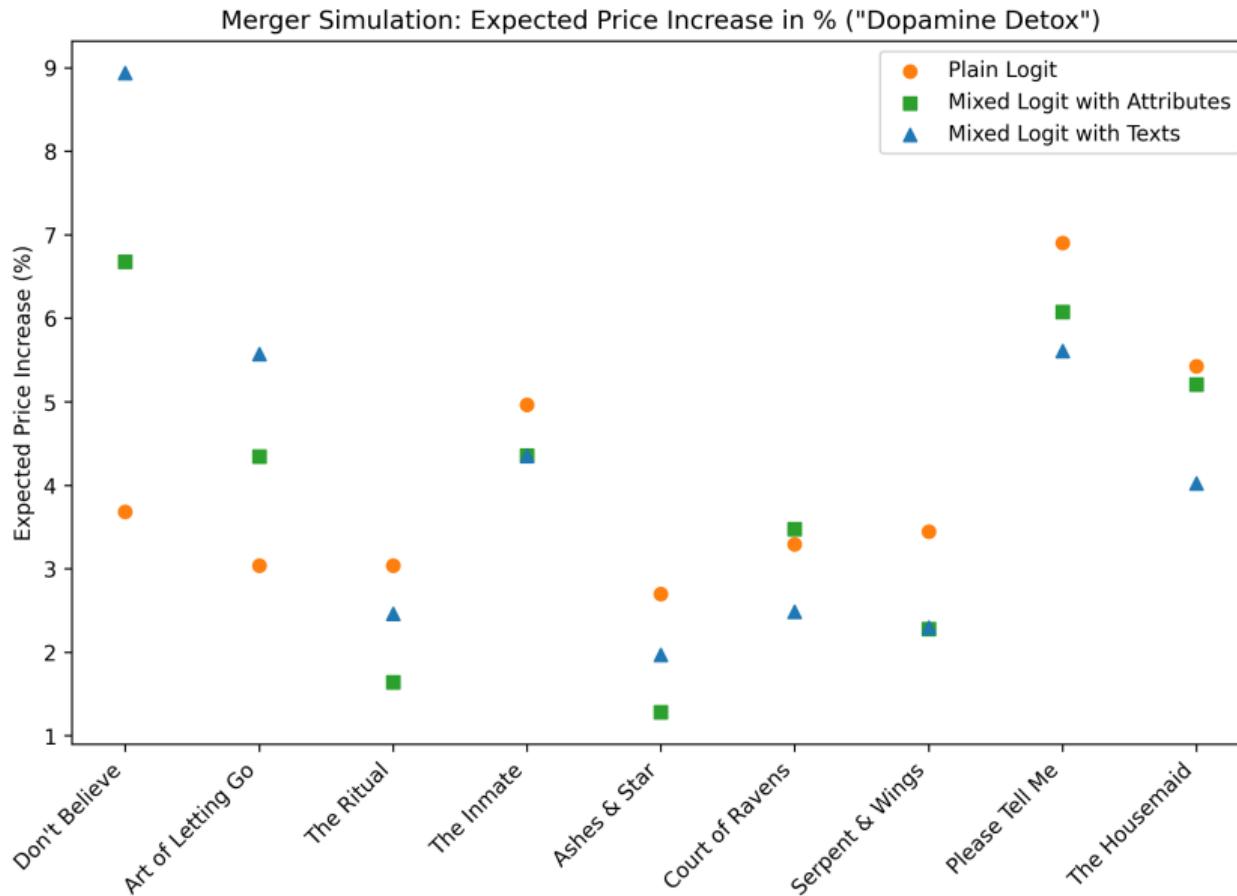
Art of Letting Go



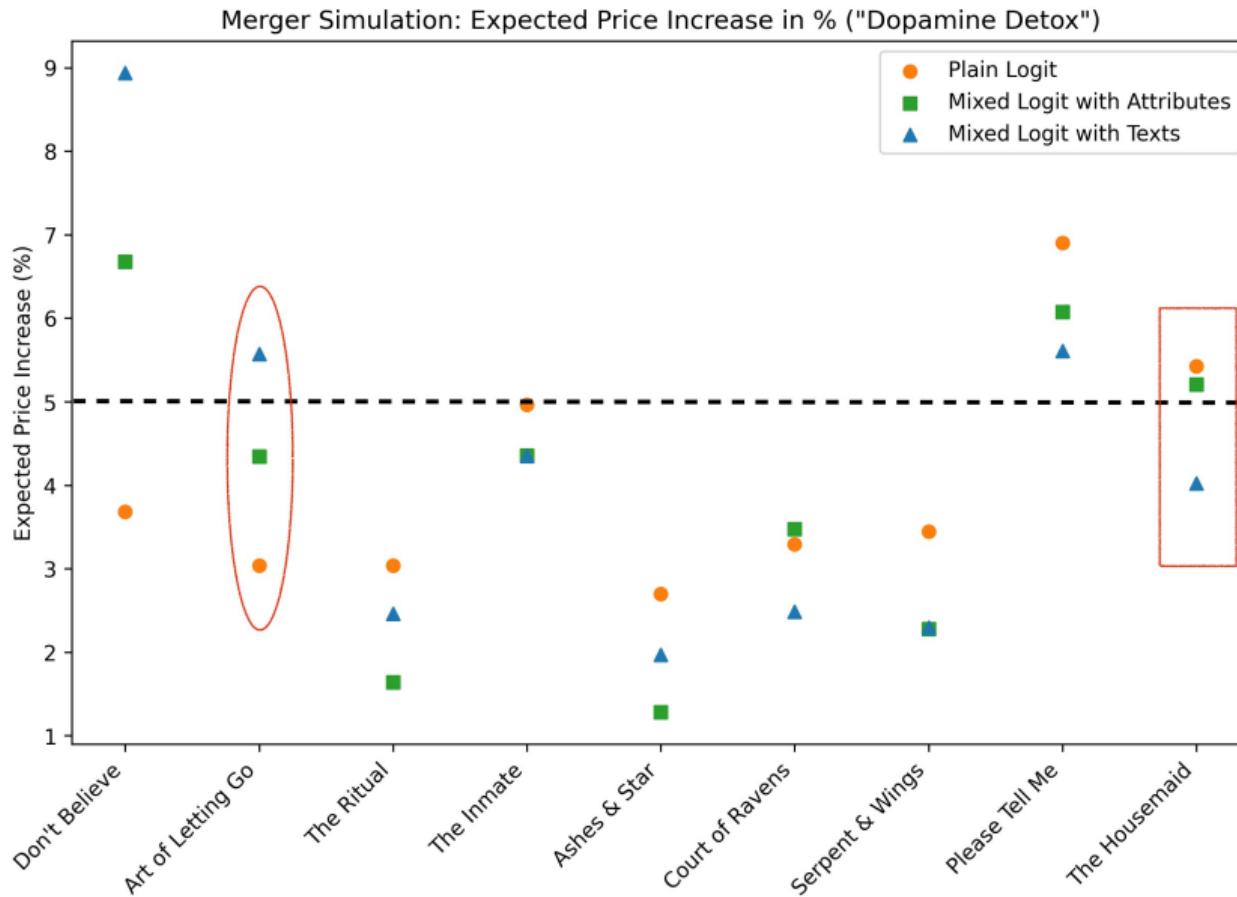
Please Tell Me

Top Second Choices

Implications for predicting the effects of a merger



Implications for predicting the effects of a merger



Does functional form matter?

- ▶ We also considered another way to incorporate embeddings
 - **Pairwise Combinatorial Logit** (Chu, 1981; Koppelman and Wen, 2000)
 - Every product pair (j, k) is a nest with correlation λ_{jk} (nests overlap)
 - Include Euclidean distances $D_{jk} = ||e_j - e_k||$ as shifters of λ_{jk}

$$\lambda_{jk} = 1 - \eta \cdot \exp(-\beta D_{jk}^\rho), \quad D_{jk} \geq 0,$$

with constraints: $\beta, \rho \geq 0$ and $\eta \in [0, 1]$

- ▶ **Pro:** Numerically light (closed-form likelihood, no numerical integration)
- ▶ **Con:** Worse Performance (also when using attribute distances)
- ▶ **Message:** Functional form matters in discrete choice models

⇒ **Ideally would want to relax parametric restrictions** (Compiani, 2022)

Demand Estimation in Online Retail

Data

- ▶ Consumer-level data of purchases from many categories on Amazon.com
 - Source: Comscore Web-Behavior Panel
 - Period: 2019-2020
 - Category definitions from Greminger et al. (2024)
- ▶ Historical price data from Keepa.com
- ▶ Sample selection:
 - Take top 15 products within each category
 - Consumers who purchased at least one product (no outside option)

We Look Across 40 Categories

Examples:



Clothing Active



Bedroom Beds



Pet Cat Snacks



Video Games PS4

Text and Image Data



Apple iPad Air 2, 64 GB, Space Gray

4.4 ★★★★★ 4,799 ratings | 213 answered questions

Climate Pledge Friendly

Price: \$146.98

Brand Apple

Model Name iPad Air 2

Memory Storage Capacity 64 GB

Screen Size 9.7 Inches

Display Resolution 2048 x 1536 Pixels
Maximum

product title

About this item

- Apple iOS 8; 9.7-Inch Retina Display; 2048x1536 Resolution
 - A8X Chip with 64-bit Architecture; M8 Motion Coprocessor
 - Wi-Fi (802.11a/b/g/n/ac); 16 GB Capacity; 2GB RAM
 - 8 MP iSight Camera; FaceTime HD Camera - Up to 10 Hours of Battery Life
- [See more product details](#)

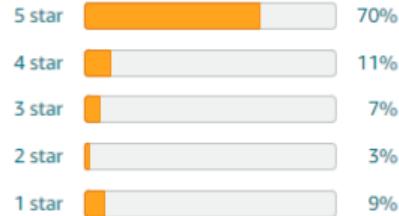
product description

Text and Image Data

Customer reviews

★★★★★ 4.3 out of 5

3,166 global ratings



⌄ How customer reviews and ratings work

By feature

Easy to use	★★★★★ 4.4
Screen quality	★★★★★ 4.3
Touch Screen	★★★★★ 4.3

Most recent

From the United States



JustinAvo

★★★★★ Apple iPod air

Reviewed in the United States us on June 23, 2023

Verified Purchase

Arrived on time, works great, overall good service! Really happy with the product!!

Helpful

| Report



Sabrina Castillo

★★★★★ NEVER AGAINNN

Reviewed in the United States us on June 22, 2023

Verified Purchase

IPAD DIES REALLY FAST AND FREEZES ON A BLANK SCREEN NEVER AGAINN !!'

Helpful

| Report

product reviews

Diversion Matrix: Our Approach

Tablets

	Fire 7	Fire HD 8	Fire HD 10	Fire 7 Kids	iPad 10.2	Fire HD 8 Kids	Dragon Touch	iPad 9.7
Fire 7	0	0.651	0.113	0	0	0	0	0
Fire HD 8	0.935	0	0.514	0	0	0	0	0.028
Fire HD 10	0.065	0.344	0	0	0.450	0	0	0.444
Fire 7 Kids	0	0	0.005	0	0.207	0.931	0.903	0.083
iPad 10.2	0	0	0.297	0.175	0	0	0	0.444
Fire HD 8 Kids	0	0	0.003	0.623	0.086	0	0.097	0
Dragon Touch	0	0	0	0.193	0	0.069	0	0
iPad 9.7	0	0.005	0.069	0.009	0.257	0	0	0

Our approach identifies “intuitive” closest substitutes

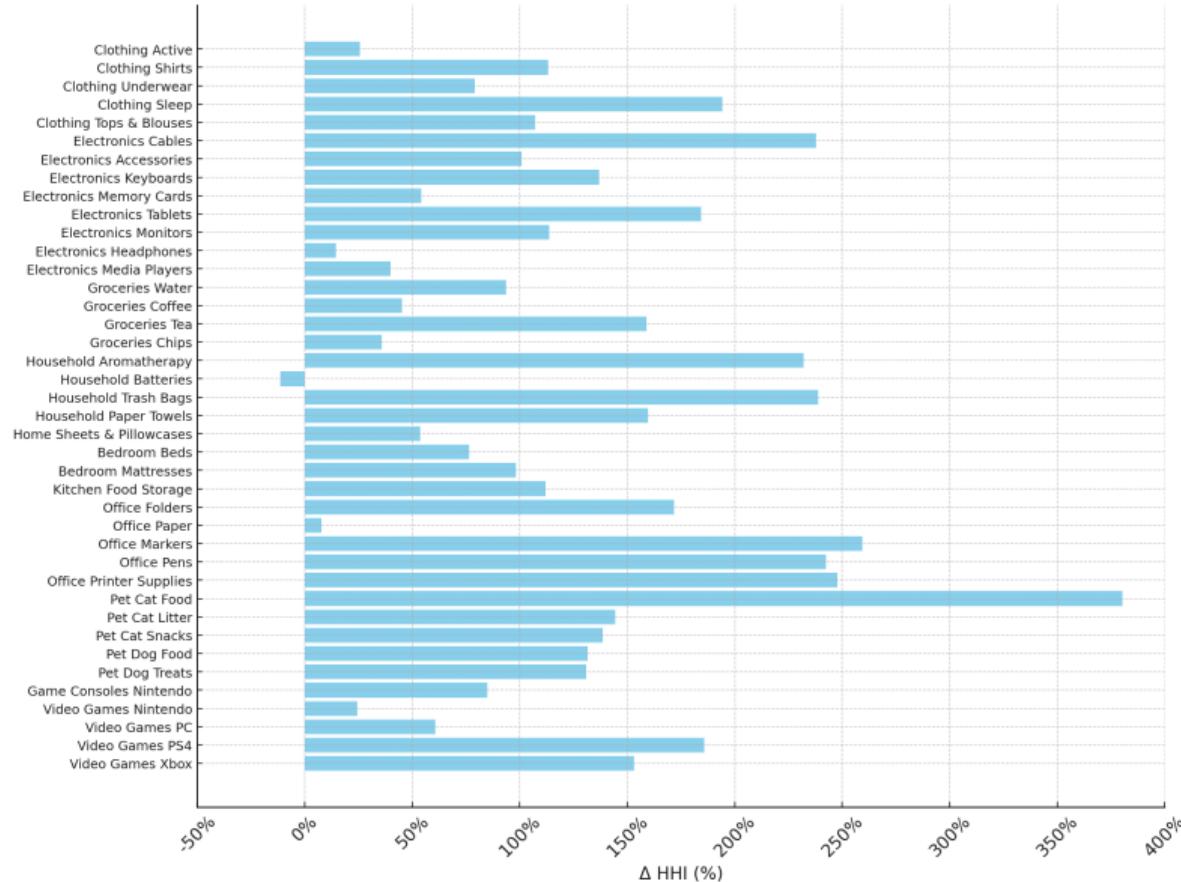
Diversion Matrix: Logit

Tablets

	Fire 7	Fire HD 8	Fire HD 10	Fire 7 Kids	iPad 10.2	Fire HD 8 Kids	Dragon Touch	iPad 9.7
Fire 7	0	0.380	0.385	0.363	0.356	0.222	0.450	0.315
Fire HD 8	0.274	0	0.239	0.133	0.163	0.306	0.150	0.093
Fire HD 10	0.329	0.322	0	0.265	0.319	0.278	0.250	0.278
Fire 7 Kids	0.134	0.078	0.136	0	0.062	0.056	0.050	0.130
iPad 10.2	0.146	0.112	0.131	0.142	0	0.083	0.050	0.130
Fire HD 8 Kids	0.052	0.044	0.058	0.044	0.031	0	0	0.037
Dragon Touch	0.017	0.029	0.028	0.027	0.013	0.056	0	0.019
iPad 9.7	0.047	0.034	0.023	0.027	0.056	0	0.050	0

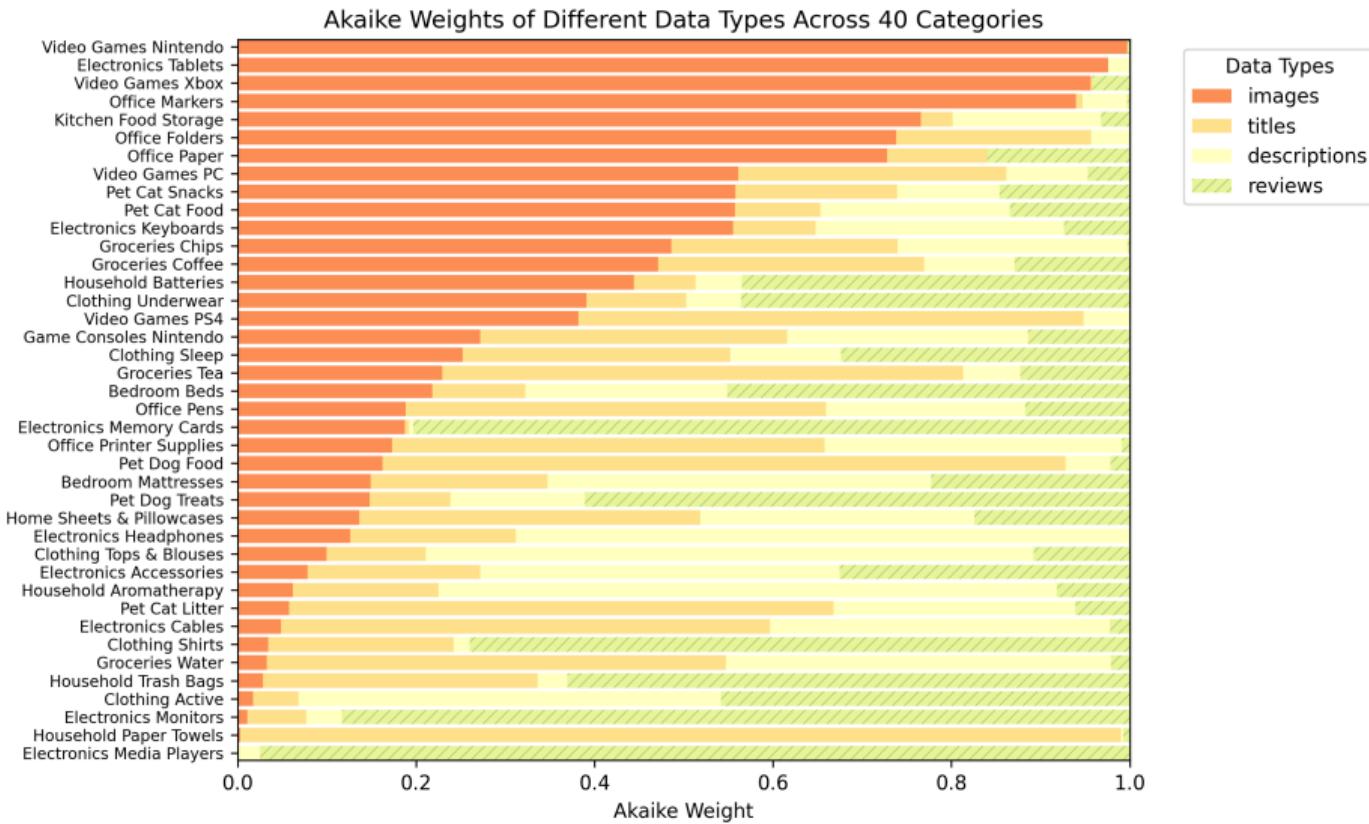
Logit mechanically predicts substitution to popular products (IIA)

Diversion Ratios are Much More Concentrated than in Logit Model

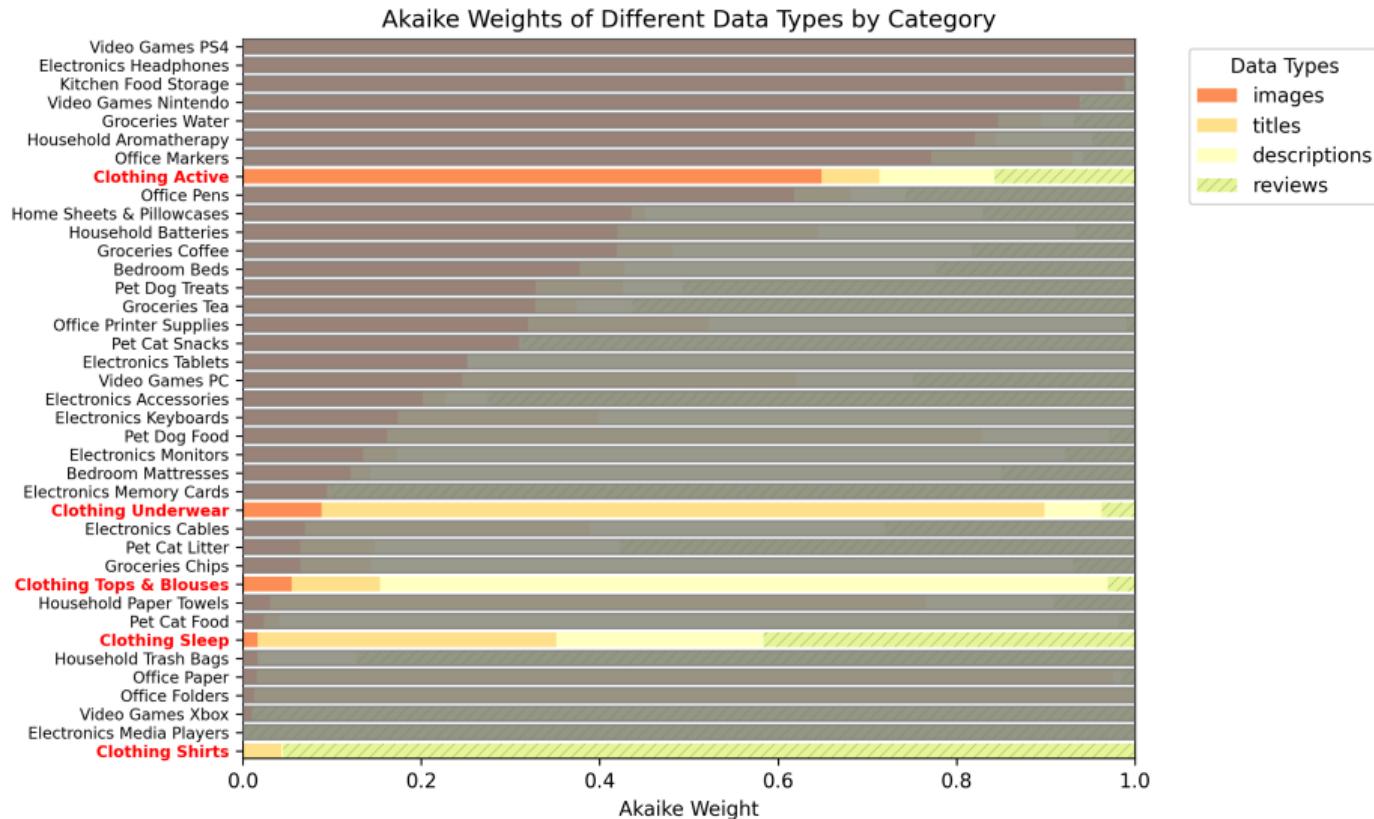


- ▶ Likelihood-ratio test rejects Logit across all categories
- ▶ Diversion ratios are more concentrated ($> 2x$) ⇒ **Unstructured data captures closest substitutes**

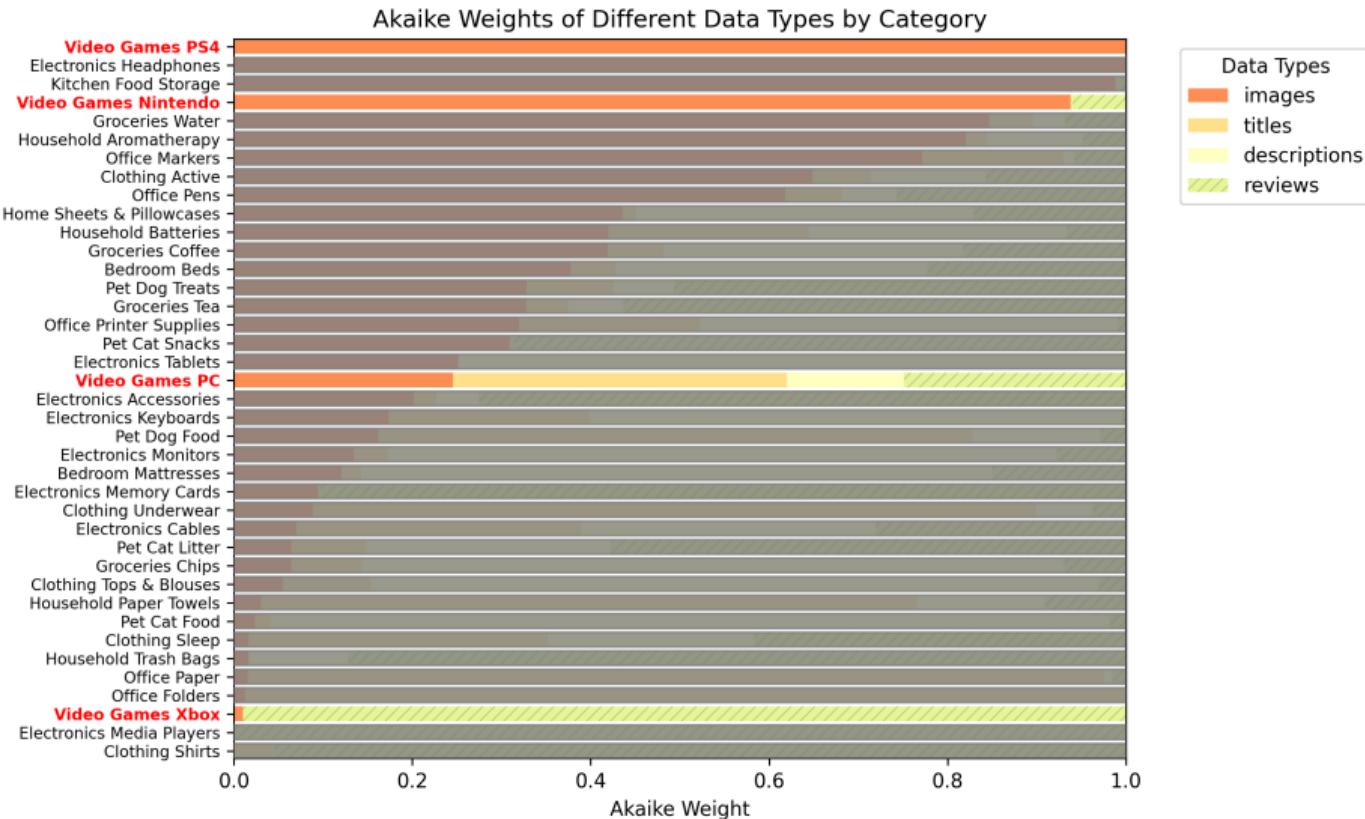
Hard to Predict Which Data Performs Best ⇒ Model Selection is Key



Hard to Predict Which Data Performs Best ⇒ Model Selection is Key



Hard to Predict Which Data Performs Best ⇒ Model Selection is Key



Conclusion & Next Steps

Recap:

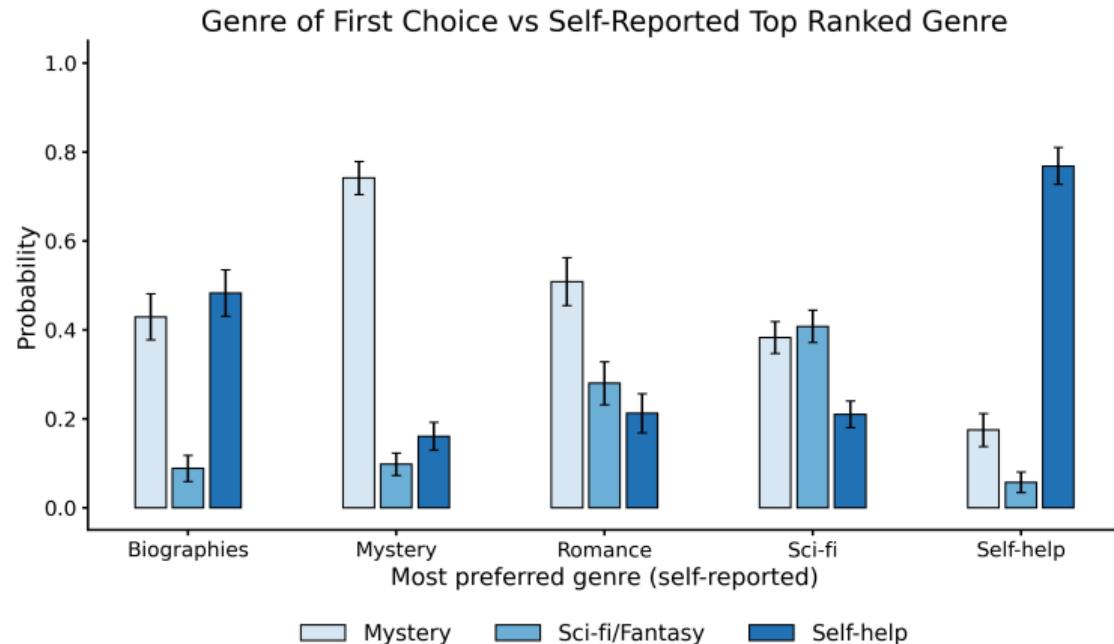
- ▶ Proposed two models that incorporate text and image distances
- ▶ Validated the models via an experiment
- ▶ Showed that our preferred model performs well with real data from online retail

Next steps / next papers:

- ▶ How much better can we do by **fine-tuning the ML models** (vs off-the-shelf models)?
- ▶ Apply this to do **empirical analyses at scale** (e.g., merger analysis, markup estimation)
- ▶ **Econometrics** of demand estimation with unstructured data

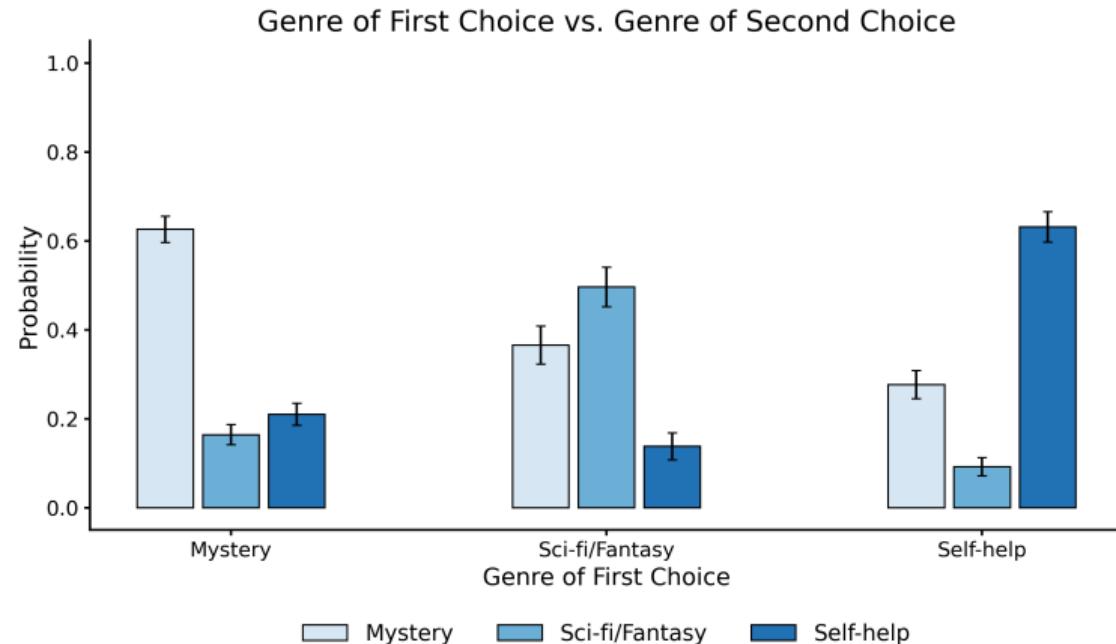
Thank You!

First choices align with self reported preferences



◀ Return

First choices and second choices are consistent



[◀ Return](#)