

Copyright and Competition: Estimating Supply and Demand with Unstructured Data

Sukjin Han and Kyungho Lee

University of Bristol and Yale University

MLESC 2025

Copyright for Creative Goods

[Copyright policy](#) plays pivotal role in the functioning of markets for creative works

- ▶ e.g., books, music, movies, illustrations, photos, fashion, mobile apps, cartoons, ...

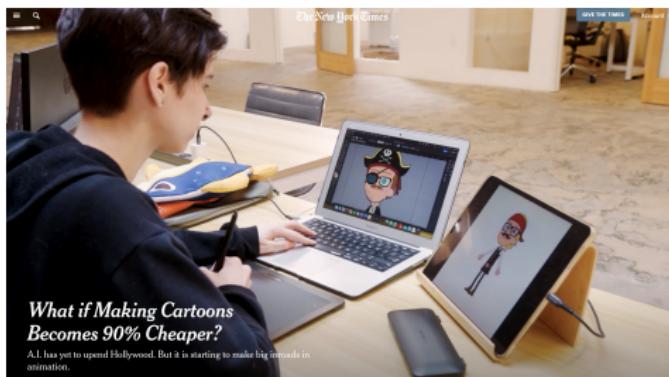
The role of copyright becomes even more relevant with cost-reducing technologies, esp. [generative AI](#)

Copyright for Creative Goods

Copyright policy plays pivotal role in the functioning of markets for creative works

- ▶ e.g., books, music, movies, illustrations, photos, fashion, mobile apps, cartoons, ...

The role of copyright becomes even more relevant with cost-reducing technologies, esp. generative AI

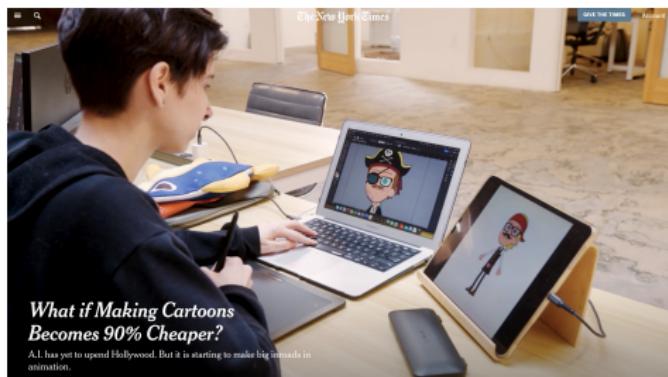


Copyright for Creative Goods

Copyright policy plays pivotal role in the functioning of markets for creative works

- ▶ e.g., books, music, movies, illustrations, photos, fashion, mobile apps, cartoons, ...

The role of copyright becomes even more relevant with cost-reducing technologies, esp. generative AI



*What if Making Cartoons
Becomes 90% Cheaper?*

A.I. has yet to upend Hollywood. But it is starting to make big inroads in animation.

- ▶ policymakers are seeking effective solutions: U.S. Copyright Office Reports 24, 25; Japan's Copyright Guidelines 24

This Paper: Questions

Q: How does copyright protection affects competition?

Q: How should we design the copyright policy in the context of low-cost technologies?

This Paper: Questions

Q: How does copyright protection affects competition?

Q: How should we design the copyright policy in the context of low-cost technologies?

A common feature of products with creative elements:

- ▶ key attributes are unstructured (e.g., design)
- ▶ focus of copyright policy
- ▶ judging infringement via product similarity

This Paper: Questions

Q: How does copyright protection affects competition?

Q: How should we design the copyright policy in the context of low-cost technologies?

A common feature of products with creative elements:

- ▶ key attributes are unstructured (e.g., design)
- ▶ focus of copyright policy
- ▶ judging infringement via product similarity

Therefore, crucial to quantify these attributes and build an economic model incorporating them

Challenges

Challenges:

1. products' complex unstructured attributes (hard to standardize, compare and analyze)
2. mathematical characterization of copyright policy
3. for some products, consumers may not value unstructured attributes as much as structured ones (e.g., product specs)
4. for some artistic products, markets are thin

Challenges

Challenges:

1. products' complex unstructured attributes (hard to standardize, compare and analyze)
2. mathematical characterization of copyright policy
3. for some products, consumers may not value unstructured attributes as much as structured ones (e.g., product specs)
4. for some artistic products, markets are thin

To make progress, focus on a specific creative product: **fonts**

This Paper: Markets for Fonts

Why fonts?

1. design is central to functionality and value
2. visual info is one of the simplest, easy to interpret
3. copyright (Carroll 94; Lipton 09; Manfredi 10; Evans 13) and AI-assisted design (Zeng et al 19; Wang et al 20) have been important policy issues
4. rich transaction data with consumer, product, and firm info
5. stylized products, capturing essence that many products have in common: design attributes and copyrights

This Paper: Markets for Fonts

Why fonts?

1. design is central to functionality and value
2. visual info is one of the simplest, easy to interpret
3. copyright (Carroll 94; Lipton 09; Manfredi 10; Evans 13) and AI-assisted design (Zeng et al 19; Wang et al 20) have been important policy issues
4. rich transaction data with consumer, product, and firm info
5. stylized products, capturing essence that many products have in common: design attributes and copyrights

Use data from the world's largest online marketplace for fonts

- ▶ nearly 33,000 fonts, 3,000,000 transactions

This Paper: Embeddings for Unstructured Attributes

Represent font's visual attributes as **embeddings**

- ▶ by specifically training convolutional neural network with triplet loss + institutional knowledge (Schroff et al 15; Chen et al 20; Han et al 23)
- ▶ show the resulting embeddings have properties appropriate for our analysis

Given the embedding space, characterize the competition of firms in the visual dimension as **spatial competition**

- ▶ and compute **visual similarity**, which serves as a crucial metric in our policy analyses

This Paper: Exploratory Analyses

Reduced-form analysis reveals...

- ▶ firms engage in local competition in the visual characteristics space
- ▶ business stealing has significant and lasting impacts on revenue

Implications on the role of copyright policy:

- ▶ providing local protection

This Paper: Model for Demand and Supply

Develop an equilibrium model of supply and demand that integrates the embeddings

- ▶ to study competition and welfare effects of copyright policy
- ▶ traditional econ models include unstructured attributes as unobservables (not enough!)

This Paper: Model for Demand and Supply

Develop an equilibrium model of supply and demand that integrates the embeddings

- ▶ to study competition and welfare effects of copyright policy
- ▶ traditional econ models include unstructured attributes as unobservables (not enough!)

1. Supply side:

- ▶ firms' entry and location choices (within the visual characteristics space) and pricing
- ▶ copyright policy: restrictions on the area of possible choices

2. Demand side:

- ▶ consumers' heterogeneous preferences over visual attributes

This Paper: Model for Demand and Supply

Incorporate the embeddings into supply and demand models after further reducing dimension by...

- ▶ principal component analysis (unsupervised); or
- ▶ partial least squares (supervised with demand info)

This Paper: Model for Demand and Supply

Incorporate the embeddings into supply and demand models after further reducing dimension by...

- ▶ principal component analysis (unsupervised); or
- ▶ partial least squares (supervised with demand info)

Counterfactual welfare analyses reveal the interplay between copyright policy and cost-reducing technologies

This Paper: Model for Demand and Supply

Incorporate the embeddings into supply and demand models after further reducing dimension by...

- ▶ principal component analysis (unsupervised); or
- ▶ partial least squares (supervised with demand info)

Counterfactual welfare analyses reveal the interplay between copyright policy and cost-reducing technologies

Our approach can be used for other similar industries with unstructured attributes

Related Literature and Contributions

Unstructured data (text, images) in economics

- ▶ Gentzkow & Shapiro 10; Gross 16; Hoberg & Philips 16; Glaeser et al 18; Gentzkow et al 19a, 19b; Dell 24; Ludwig & Mullainathan 24; Gorin et al 25; Modarressi et al 25;...
- ▶ Han et al 23; Bajari et al 25; Compiani et al 25
- ▶ **model for supply and demand with unstructured data via embeddings**

Related Literature and Contributions

Copyright policy:

- ▶ Waldfogel 12, 16; Biasi & Moser 21; Giorcelli & Moser 20; Oberholzer-Gee & Strumpf 07; Li et al 18; MacGarvie & Moser 15; Gans 24; Rassenfosse et al 24
- ▶ **similarity judgment:** Lemley 10; Balganesh et al 14; Scheffler et al 23
- ▶ **quantitative evaluation of the similarity-based copyright policy**

Optimal product variety:

- ▶ Spence 76; Mankiw & Whinston 86; Berry & Waldfogel 99, 01; Sweeting 13; Berry et al 16
- ▶ **allowable similarity as a policy tool to achieve optimal variety**

Product positioning:

- ▶ Seim 06, Sweeting 10, 13, Fan 13, Eizenberg 14, Wollmann 18
- ▶ **positioning in high-dimensional setting**

Background, Data and Embeddings

Background: Marketplace for Fonts

The screenshot shows the MyFonts website interface. At the top, there's a search bar with the placeholder "Search over 300,000 fonts & families". Below the search bar are various navigation links: "Browse by", "Categories", "Bestsellers", "Hot New Fonts", "Premium Typefoundries", and "WhatTheFont?". On the right side, there are buttons for "Logout", "USD (US)", "EN", and a "Log In" button. A "Subscription Plans" button is also present.

Below the header, a message states: "These are the top 50 selling fonts from all of MyFonts in the last month. Updated hourly!"

The main content area displays a grid of 15 font preview cards, each featuring a font sample and a "Buy Now" button. The fonts shown are:

- COLLECT WORKS F STRINGS DESIDERIO QUARTANA**: Sans Serif font by Myfonts. Sample text: "The quick brown fox jumps over the lazy dog". Price: From \$177.00 USD.
- Neue Helvetica**: Sans Serif font by Linotype. Sample text: "12345ABCDEFabcdef". Price: From \$42.99 USD.
- Helvetica® Now**: Sans Serif font by Linotype. Sample text: "The quick brown fox jumps over the lazy dog". Price: From \$42.99 USD.
- Ultra Light Helvetica**: Sans Serif font by Linotype. Sample text: "The quick brown fox jumps over the lazy dog". Price: From \$39.00 USD.
- Avenir® Next**: Sans Serif font by Linotype. Sample text: "The quick brown fox jumps over the lazy dog". Price: From \$67.99 USD.
- Avenir®**: Sans Serif font by Linotype. Sample text: "The quick brown fox jumps over the lazy dog". Price: From \$42.99 USD.
- Gilroy**: Sans Serif font by Myfonts. Sample text: "The quick brown fox jumps over the lazy dog".
- Neue Haas Grotesk**: Sans Serif font by Myfonts. Sample text: "The quick brown fox jumps over the lazy dog".
- Frutiger**: Sans Serif font by Myfonts. Sample text: "The quick brown fox jumps over the lazy dog".
- FF DIN**: Sans Serif font by Albert-Jan Paul. Sample text: "The quick brown fox jumps over the lazy dog".
- Univers®**: Sans Serif font by Myfonts. Sample text: "The quick brown fox jumps over the lazy dog".
- Korolev**: Sans Serif font by Myfonts. Sample text: "The quick brown fox jumps over the lazy dog".

At the bottom of the page, there are several navigation icons: back, forward, search, and other site navigation symbols.

Background: Marketplace for Fonts

Fonts are “software products”

- ▶ consumers download software with licenses
- ▶ consumers are mostly designers
 - design projects for books, ad posters, websites, apps,...

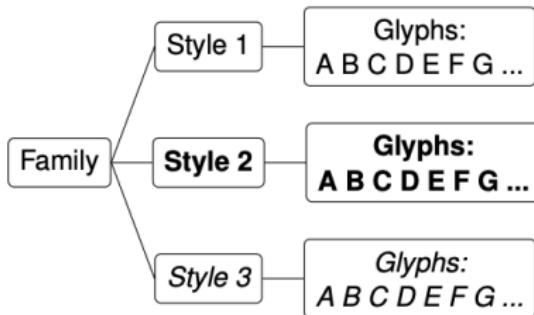
Plagiarism policy: ban designs nearly identical to existing fonts

- ▶ also court cases in the US and UK

Shares features with broader markets of creative goods:

- ▶ high-dim attributes
- ▶ high fixed costs of development (designing & coding)
- ▶ concerned with copyright infringement

Font Family, Styles and Glyphs



(a) Font Family Structure

Gilroy Semi Bold	from \$20.37	<button>Buying Choices</button>
Gilroy Semi Bold		
Gilroy Ultra Light	from \$20.37	<button>Buying Choices</button>
Gilroy Ultra Light		
Gilroy Regular	from \$20.37	<button>Buying Choices</button>
Gilroy Regular		
Gilroy Regular Italic	from \$20.37	<button>Buying Choices</button>
Gilroy Regular Italic		
Gilroy Medium Italic	from \$20.37	<button>Buying Choices</button>
Gilroy Medium Italic		
Gilroy Medium	from \$20.37	<button>Buying Choices</button>
Gilroy Medium		

(b) Example: Gilroy font family

Figure: Font Family

Image Data and Embedding Construction

We use pangram images of styles of families:

Quick zephyrs blow, vexing daft Jim.

Quick zephyrs blow, vexing daft Jim.

Quick zephyrs blow, vexing daft Jim.

Quick zephyrs blow, vexing daft Jim.

We employ an encoding method:

- ▶ input: pangram images \Rightarrow output: embeddings ($x_j^{emb} \in \mathbb{S}^{128}$)
- ▶ i.e., embedding is a low-dim normalized vector
- ▶ L^2 distance corresponds to measure of similarity

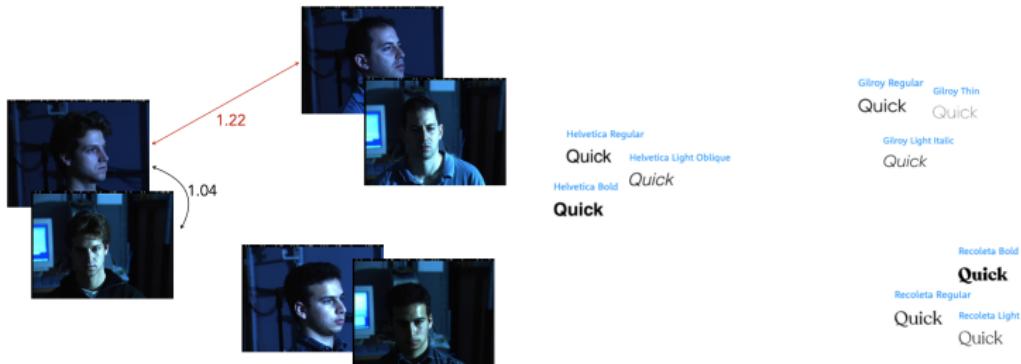
Important to capture style differences (e.g., \mathcal{A} vs A)

- ▶ more difficult than recognizing different letters (e.g., A vs B)

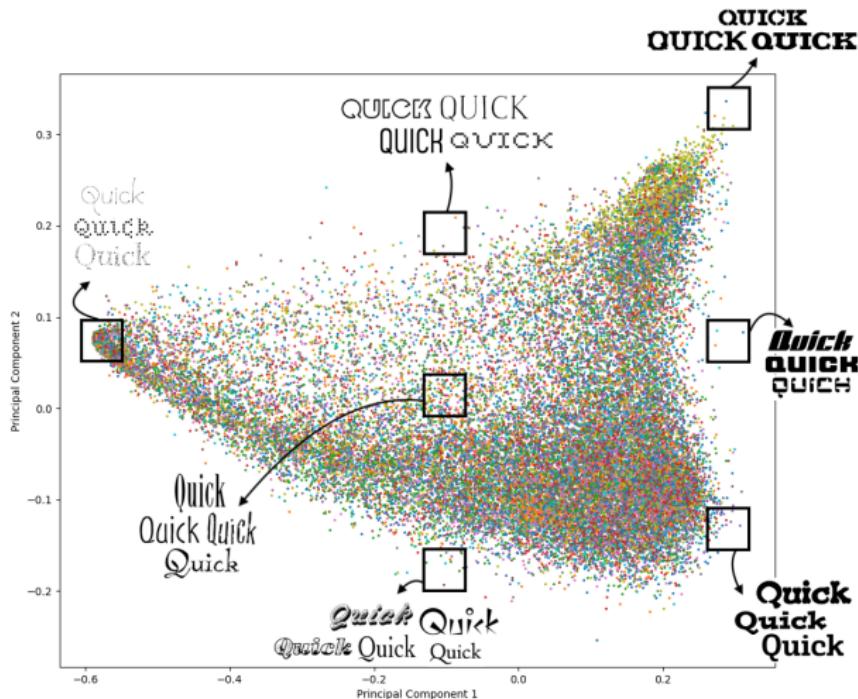
Image Data and Embedding Construction

Train CNN algorithm with triplet loss function:

- ▶ special case of contrastive learning (Chen et al 20)
 - improved uniformity (i.e., evenly-spaced embeddings) (Deli 24)
- ▶ key is to construct triplets
 - use product info as “labels” (Han et al 23)
 - multiple images of a face \approx multiple styles of a font family



2D Visualization of Embeddings



Embedding Distance as Visual Similarity

Font Name	Distance	Pangram Shape
Minion	0.000	The quick brown fox jumps over
Alia JY	0.057	The quick brown fox jumps over
Garamond	0.081	The quick brown fox jumps over
Bauhaus B. S.	0.090	The quick brown fox jumps over
Andrea H. II	0.149	<i>The quick brown fox jumps over</i>
Ruling Script	0.375	<i>The quick brown fox jumps over</i>
Scruff	0.477	The quick brown fox jumps over

- ▶ additional validation using *tags* (i.e., phrases assigned by consumers and producers; e.g., curly, geometric) as “gold standard” (Han et al 21)

Market Data

Aggregation: scanner data to panel data

- ▶ transaction record
 - ⇒ aggregation product × license type × country × time
 - product: font family
 - license type: desktop and web font
 - country: 12 countries using Roman alphabet
 - time: monthly from 2014 to 2017

Sample size ($N = 478,212$)

- ▶ large number of products (with frequent entries)

Price has large cross-sectional variation but small time-series variation

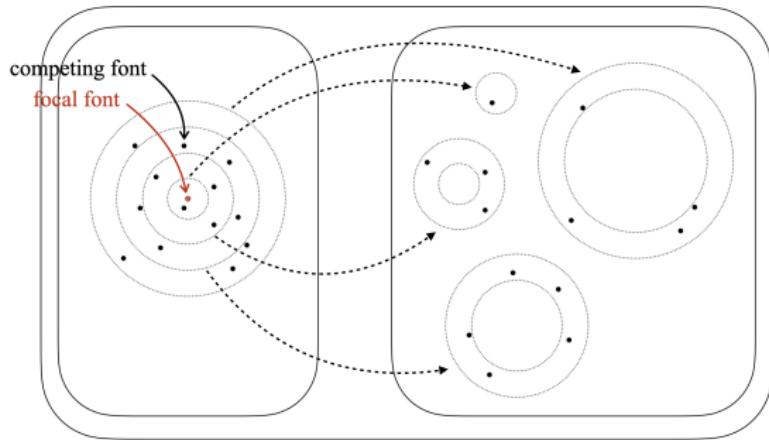
Exploratory Analyses

Analysis of Spatial Competition

Q: How does competition look in the visual characteristics space?

- ▶ Does visual similarity (calculated by embeddings) matter for competition?
- ▶ Is competition *local* in the space?

Counting Number of Spatial Competitors



$$R_{jt}^{r,r'} \equiv \sum_{j' \in J_t} \mathbf{1}\{r < \|x_{j'}^{\text{emb}} - x_j^{\text{emb}}\|_2 < r'\} \text{ for } r, r' \in \mathbb{R}$$

Analysis of Spatial Competition

$$y_{jlct} = \sum_{r \in \{0.1, 0.2, 0.3, 0.4\}} \gamma_r R_{jt}^{r-0.1,r} + \alpha_j + \alpha_l + \alpha_c + u_{jlct}$$

- ▶ y_{jlct} : revenue, quantity, or price
 - product j , license l , country c , time t
- ▶ γ_r : the **impact of additional competitor** in the certain distance range
- ▶ $\alpha_j, \alpha_l, \alpha_c, \alpha_t$: product, license, country, time FEs

Business Stealing by Nearest Competitors

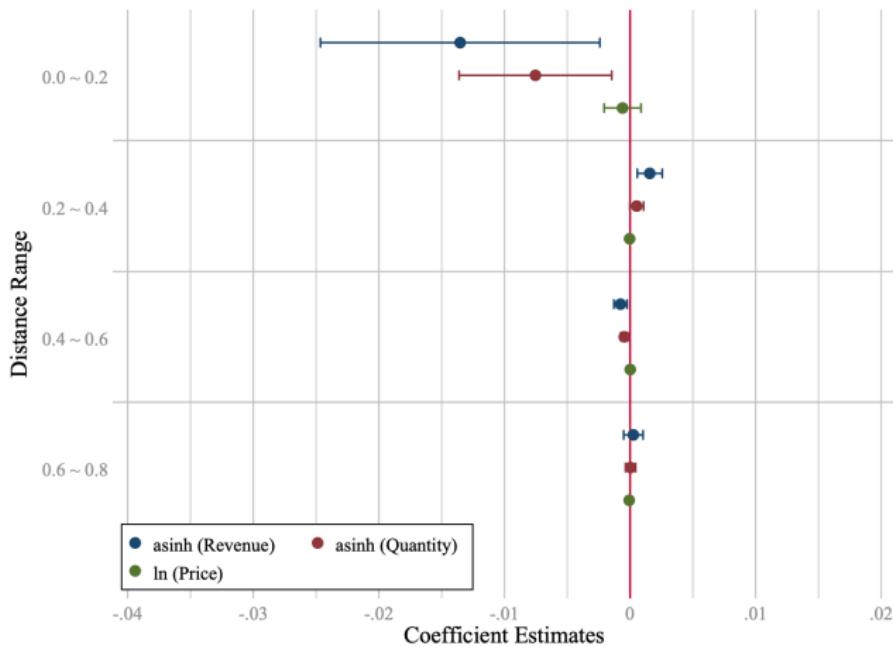


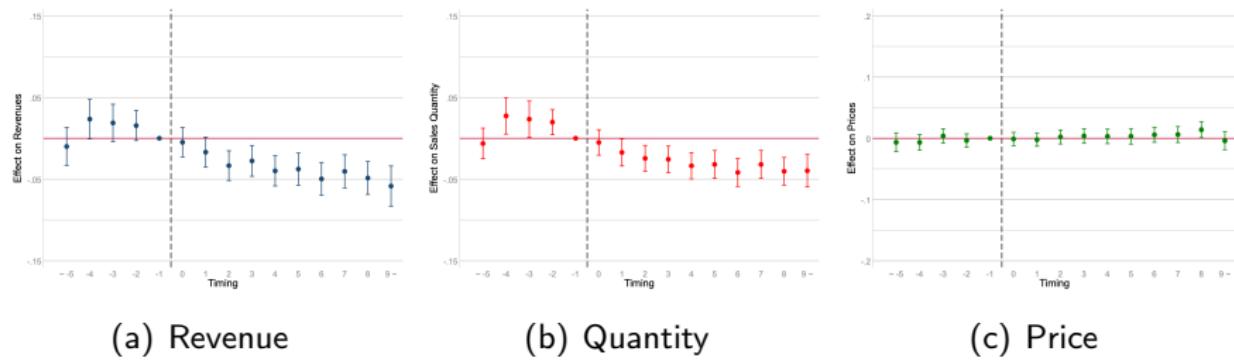
Figure: Spatial Regression Results (γ_r)

Event Study of Spatial Competition

$$y_{jlct} = \sum_{s=-5}^9 \beta_s E_{jt}^s + \alpha_f + \alpha_l + \alpha_c + \alpha_t + e_{jlct}$$

- ▶ y_{jlct} : revenue, quantity, or price
- ▶ E_{jt}^s : treatment = a new entrant appears within 5 visually closest competitors
- ▶ $\alpha_f, \alpha_l, \alpha_c, \alpha_t$: firm, license, country, time FEs

Business Stealing by Visually Similar Entrant



(a) Revenue

(b) Quantity

(c) Price

Figure: Event Study Regression Results (β_s)

- ▶ revenues reduced by 5%: business stealing
- ▶ spatial regression analyses suggest local competition

Business Stealing by Visually Similar Entrant

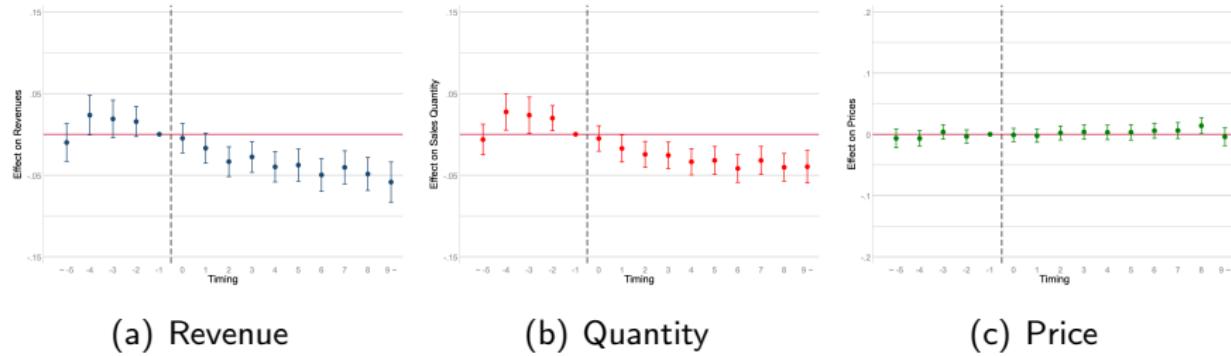


Figure: Event Study Regression Results (β_s)

- ▶ revenues reduced by 5%: business stealing
- ▶ spatial regression analyses suggest local competition

Local protection by copyright would have direct welfare effects

⇒ structural models

Structural Analyses

Structural Models

Counterfactual welfare evaluation through supply and demand models that capture...

- ▶ positioning under similarity constraints imposed by copyright policy
- ▶ heterogeneous preferences to high-dim characteristics

Incorporate the embeddings into these models by adding further dim-reduction layer:

- ▶ principal component analysis (unsupervised); or
- ▶ partial least squares (supervised with market share)

Demand Model

We specify the indirect utility model, building on Berry et al 95

$$U_{ijt} \equiv \bar{\beta}^p p_{jt} + \bar{\beta}^{str} x_j^{str} + x_j^{emb} \beta_i^{img} + \xi_{jt} + \epsilon_{ig(j)t} + (1 - \rho) \bar{\epsilon}_{ijt}$$

- ▶ i.e., nested logit with random coefficients
- ▶ p_{jt} : prices; x_j^{str} : glyphs (# of supported characters)
- ▶ x_j^{emb} : the vector of embeddings (i.e., PCs by PCA or PLS)
- ▶ β_i^{img} : random coefficients (heterogeneous preferences)
- ▶ $g(j)$: nest based on menu tags for browsing (search design)
- ▶ ρ : nesting parameter

IVs: exchange rates + differentiation IVs (Gandhi & Houde 19)

Demand Estimation

Variables/Parameters	$\bar{\beta}$	σ
Prices	-0.156	-
	(0.001)	-
Glyph Counts	0.001	-
	(0.000)	-
PC 1	5.292	9.500
	(0.082)	(0.096)
PC 2	-6.328	2.499
	(0.109)	(0.458)
PC 3	-11.823	7.652
	(0.177)	(0.209)
PC 4	-11.661	5.582
	(0.226)	(0.720)
PC 5	2.374	11.567
	(0.140)	(0.504)
PC 6	10.145	0.113
	(0.242)	(0.005)

Demand Estimation

Quantile 1
1680 Images

Quantile 2
1630 Images

Quantile 3
1757 Images

Quantile 4
1744 Images

Quantile 5
1595 Images



Figure: Mean of Pixel Values (across fitted utilities)

Demand Estimation

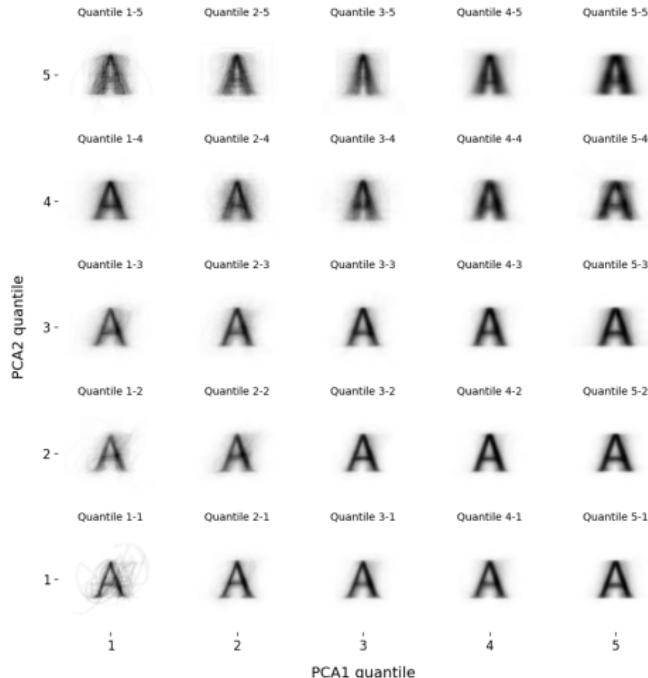


Figure: Mean of Pixel Values (across each PC)

Demand Estimation

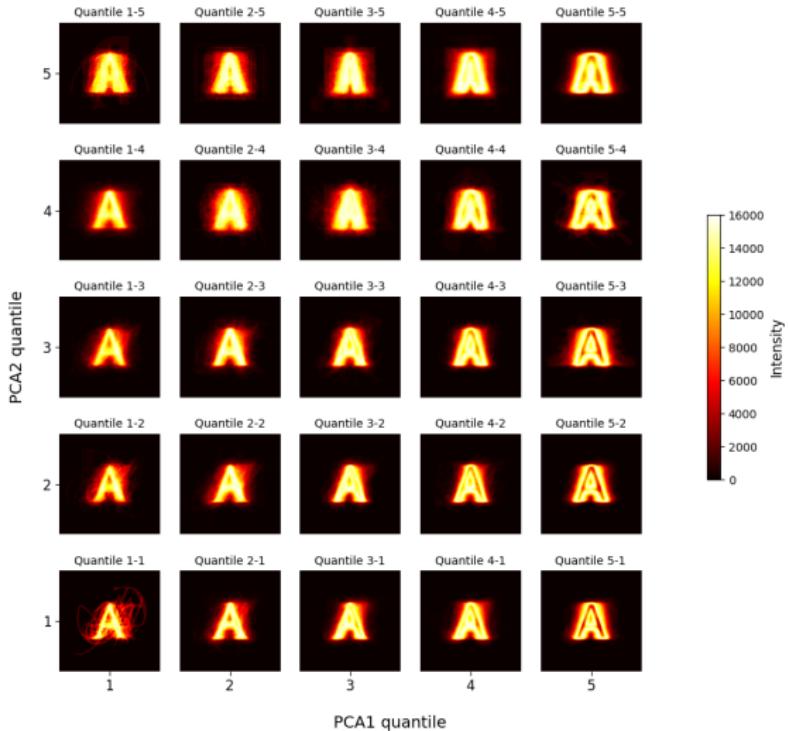


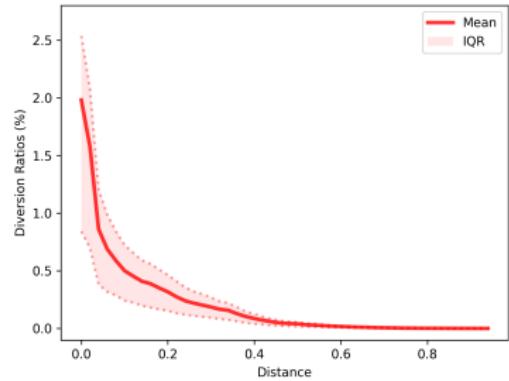
Figure: Variance of Pixel Values (across each PC)

Demand Estimation

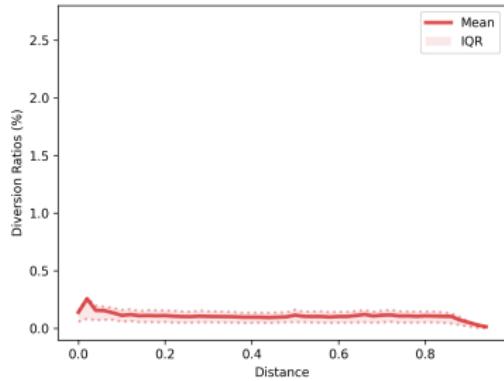


Figure: Lasso Regression Results: PC1 and PC2 on Tags

Demand Estimation



(a) Embeddings Used



(b) Embeddings Ignored

Figure: Degree of Substitution vs. Original Embedding Distance

Supply Model

Consider a multi-stage model:

1. Entry decision
 2. Product positioning decision
 3. Pricing decision
- ▶ an empirical counterpart to theory of spatial location choice
(Hotelling 29, Salop 79),
 - ▶ based on the characteristics space defined by embeddings

At each t , firm (font designers) f decides whether to launch product k or not

▶ Timing Assumptions

Supply Model

Total profit:

$$\Pi_{ft,k} \equiv \sum_{j \in J_{ft}} \pi_{jt} + 1\{E_{kt} = 1\} (\pi_{kt} - F(\mathbf{x}_t, \nu_k))$$

- ▶ J_{ft} set of product launched by firm f up to t
- ▶ $\pi_{jt} \equiv s_{jt} M_t (p_{jt} - mc_{jt})$ variable profit (of product j at t)
- ▶ E_{kt} entry decision for *new* product k at t
- ▶ $F(\mathbf{x}_t, \nu_k)$ fixed cost where $\mathbf{x}_t \equiv \{x_k^{emb}\}_k$

$$F(\mathbf{x}_t, \nu_k) \equiv \nu_{k0} + \sum_{\ell} \left[(\eta_{0\ell} + \nu_{k\ell}) \mathbf{x}_{k\ell}^{emb} + \sum_{j \neq k} \left(\eta_{1\ell} \mathbf{d}_{kj}^{\ell} + \eta_{2\ell} (\mathbf{d}_{kj}^{\ell})^2 + \eta_{3\ell} (\mathbf{d}_{kj}^{\ell})^3 \right) \right]$$

- $d_{kj}^{\ell} \equiv \|\mathbf{x}_{k\ell}^{emb} - \mathbf{x}_{j\ell}^{emb}\|_2$: distance of product k to incumbent j

Solving Supply Model

Final stage:

$$p_{ft}^* = \arg \max_{p_{jt} \in \{p_{jt}: j \in J_{ft} \cup \{k\}\}} \sum_{j \in J_{ft} \cup \{k\}} s_{jt} M_t (p_{jt} - mc_{jt})$$

- ▶ optimal price as function of prices and observed and unobserved characteristics of all (possibly neighboring) products through demand
- ▶ effectively captures competition in the marketplace

Solving Supply Model

Second stage:

$$x_k^{emb,*} = \arg \max_{x_k^{emb} \in \mathbb{S}^d} E_{\xi_{kt}} \left[\sum_{j \in J_{ft} \cup \{k\}} \pi_{jt} \right] - F(\mathbf{x}_t, \nu_k)$$
$$\text{s.t. } \|x_k^{emb} - x_{j'}^{emb}\|_2 \geq d \text{ for all } j' \in J_{-ft}$$

- ▶ similarity constraint due to copyright policy
- ▶ necessary condition for optimality:

$$\sum_{j \in J_{ft} \cup \{k\}} E_{\xi_{kt}} \left[\frac{\partial \pi_{jt}}{\partial x_k^{emb}} + \sum_{j' \in J_{-ft}} + \frac{\partial \pi_{jt}}{\partial p_{j'}} \frac{\partial p_{j'}}{\partial x_k^{emb}} \right]$$
$$+ \sum_{j' \in J_{-ft}} \left[\lambda_{kj'} \left(\frac{\partial \|x_k^{emb} - x_{j'}^{emb}\|_2}{\partial x_k^{emb}} - d \right) \right] = \frac{\partial F(\mathbf{x}_t, \nu_{kt})}{\partial x_k^{emb}}$$

- estimating equation (with/without IVs)

Solving Supply Model

First stage:

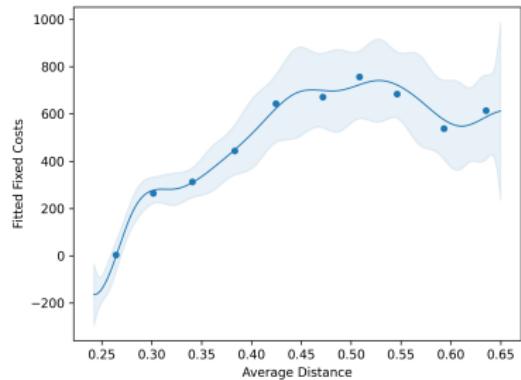
$$E_{\xi_{kt}} [\Pi_{ft,k}(E_{kt} = 1)] - \Pi_{ft,k}(E_{kt} = 0) \geq 0$$

- ▶ recall $\Pi_{ft,k} = \sum_{j \in J_{ft}} \pi_{jt} + 1\{E_{kt} = 1\} (\pi_{kt} - F(\mathbf{x}_t, \nu_k))$
- ▶ **revealed profit approach** (Bresnahan & Reiss 91; Berry 92; Berry & Waldfogel 99; Seim 06): firm f pays fixed costs if the expected net profit is greater than zero
- ▶ partial identification of fixed cost *level*

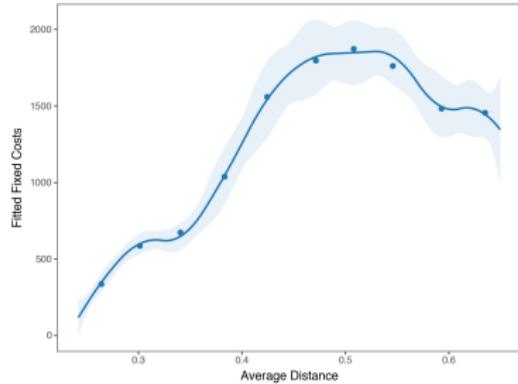
Supply-Side Estimation

Parameters	(1) $\partial F / \partial x_1^{emb}$	(2) $\partial F / \partial x_2^{emb}$	(3) $\partial F / \partial x_3^{emb}$	(4) $\partial F / \partial x_4^{emb}$	(5) $\partial F / \partial x_5^{emb}$	(6) $\partial F / \partial x_6^{emb}$
$\eta_{0\ell}$	3400.8 (218.35)	-2916.9 (100.01)	-6742.9 (292.15)	-5549.6 (185.88)	1699.0 (104.96)	5158.2 (182.42)
$\eta_{1\ell}$	0.15 (0.06)	0.12 (0.05)	0.41 (0.28)	-0.04 (0.16)	0.10 (0.15)	-0.19 (0.10)
$\eta_{2\ell}$	0.41 (0.17)	-0.34 (0.29)	-3.61 (2.41)	0.23 (1.99)	0.43 (2.23)	2.59 (1.62)
$\eta_{3\ell}$	-0.22 (0.14)	0.57 (0.48)	10.85 (5.96)	-0.28 (6.82)	-0.98 (8.92)	-6.65 (7.00)
R^2	0.33	0.06	0.07	0.00	0.25	0.01
F-stat	271.96	33.71	40.66	0.73	177.85	4.44
Observations				1,630		

Supply-Side Estimation



(a) OLS



(b) IV

Figure: Fixed Cost vs. Embedding Distance

- ▶ mimicking advantages when close to incumbents

Counterfactual Welfare Analyses

Counterfactual Welfare Analyses

Given the estimated structural model, investigate the role of copyright policy in competition and welfare

Two analyses:

1. Enforcement of stricter copyright protection
 - ▶ by increasing copyright protection level \underline{d}

Copyright Policy and Consumer Surplus

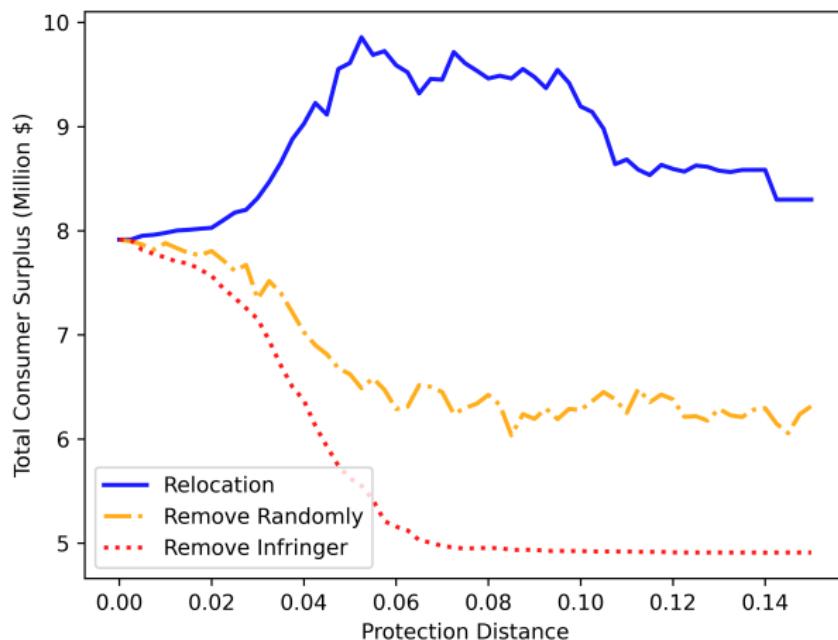


Figure: Consumer Surplus by Varying Copyright Protection d

Counterfactual Welfare Analyses

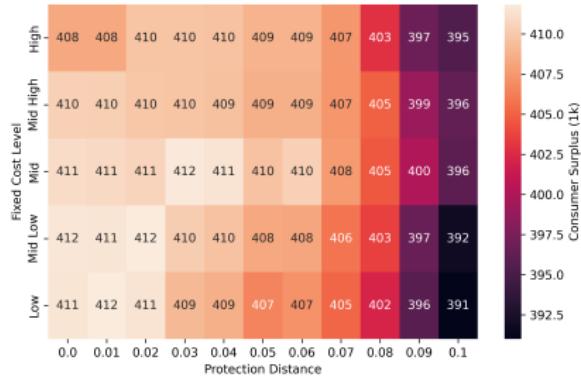
2. Introduction of cost-reducing technologies (e.g., GenAI) in product design

- ▶ Scenario A: GenAI as an assistant
 - concave cost function; lower level
- ▶ Scenario B: GenAI as a substitute
 - flat cost function; lower level

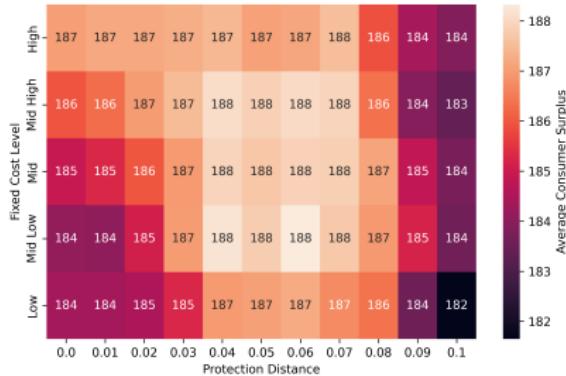
Consider both aggregate and average (per product) welfares

- ▶ infringement cases typically focus on individual products

Scenario A: Consumer Surplus



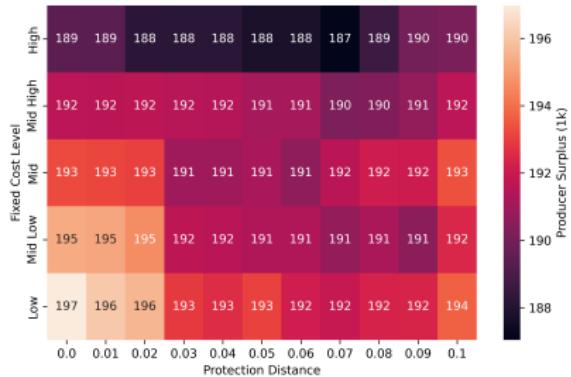
(a) Aggregate



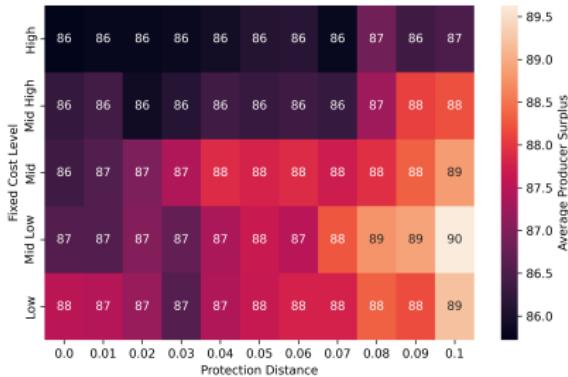
(b) Average

Figure: Consumer Surplus Across Protection Levels and Fixed Costs

Scenario A: Producer Surplus



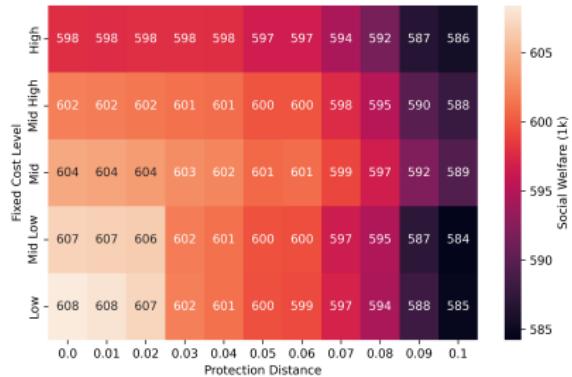
(a) Aggregate



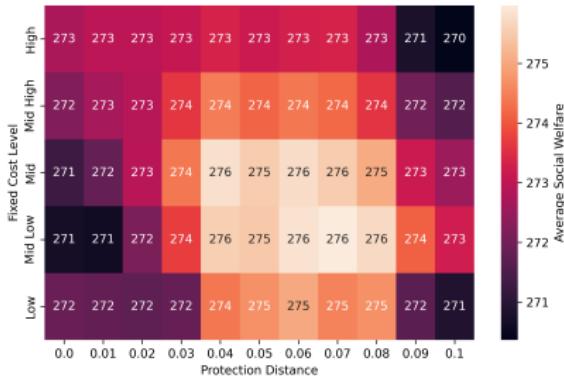
(b) Average

Figure: Producer Surplus Across Protection Levels and Fixed Costs

Scenario A: Social Welfare



(a) Aggregate



(b) Average

Figure: Social Welfare Across Protection Levels and Fixed Costs

Demand Distribution and Entry Patterns (High Cost)

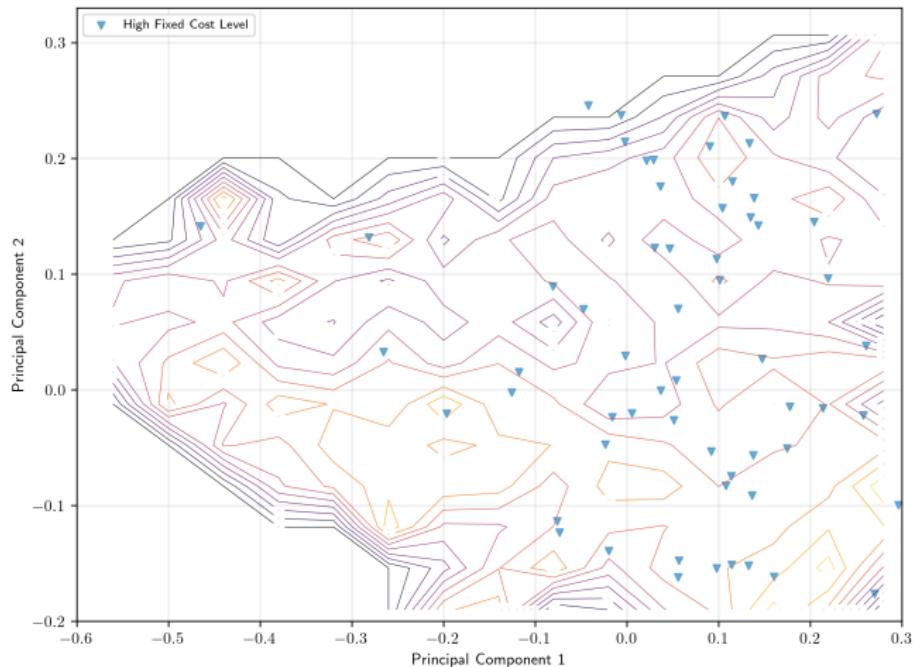


Figure: Entries under Copyright $d = 0.00$

Demand Distribution and Entry Patterns (High Cost)

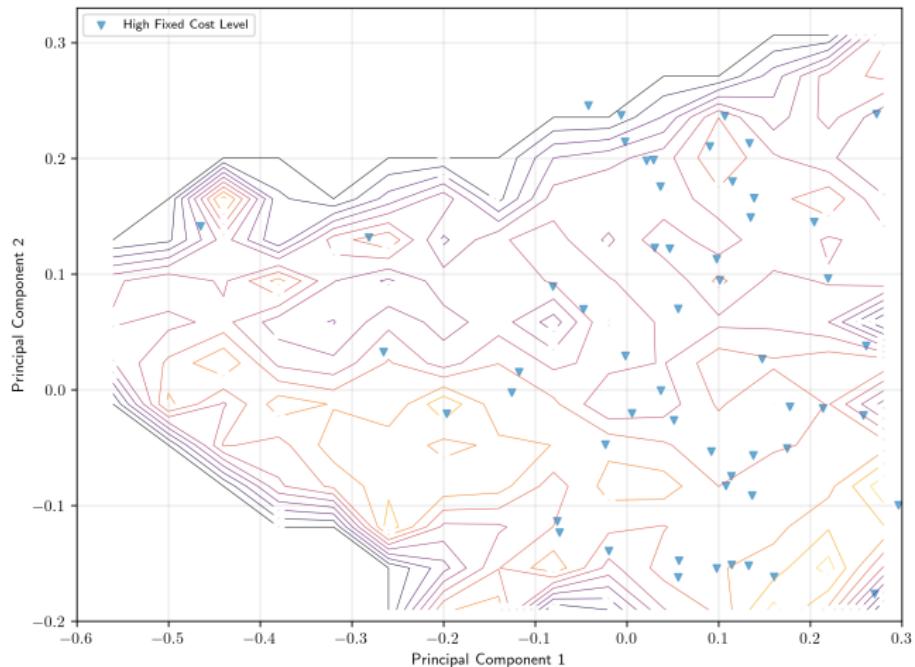


Figure: Entries under Copyright $d = 0.02$

Demand Distribution and Entry Patterns (High Cost)

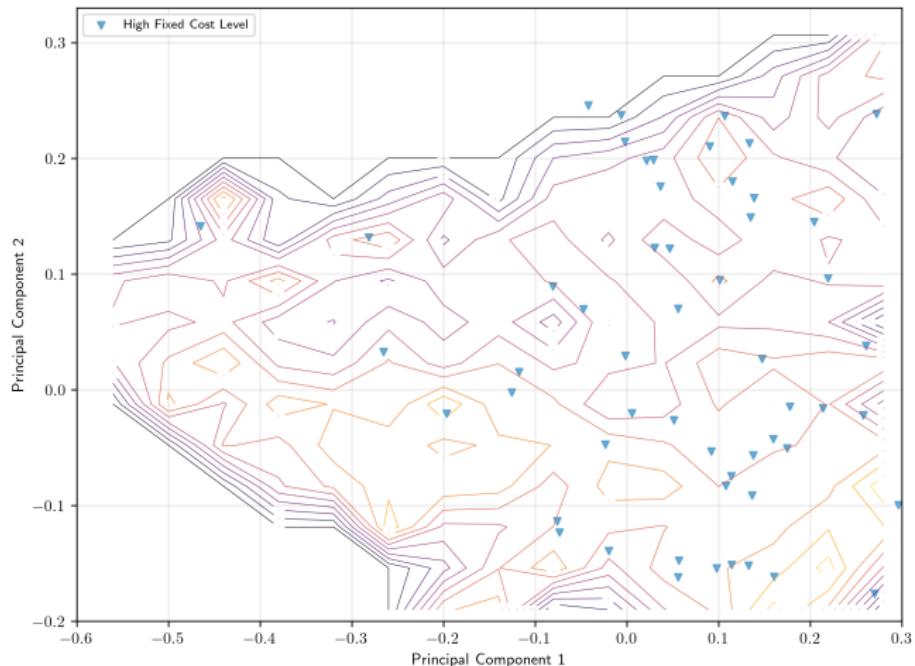


Figure: Entries under Copyright $d = 0.04$

Demand Distribution and Entry Patterns (High Cost)

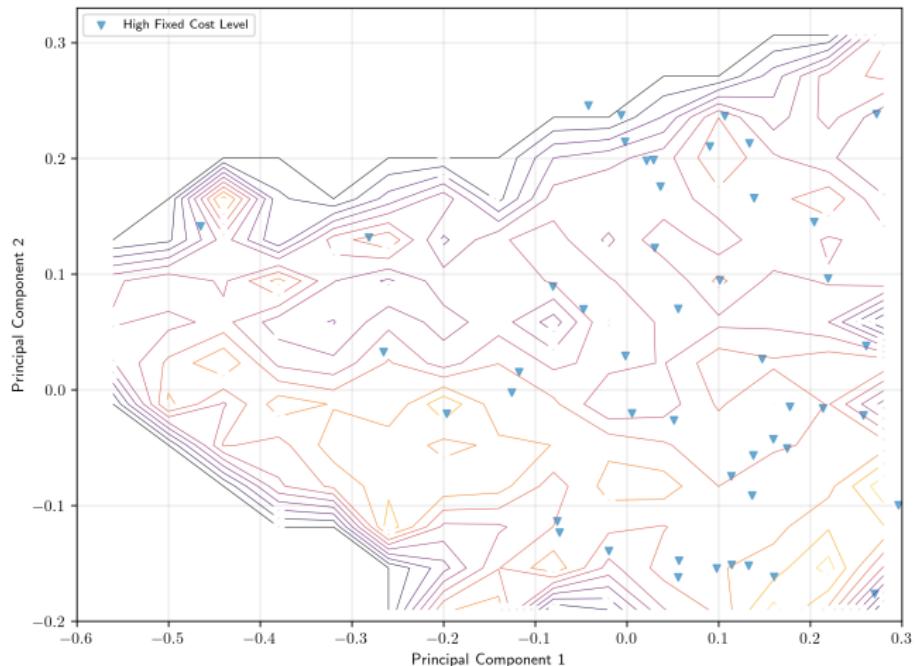


Figure: Entries under Copyright $d = 0.06$

Demand Distribution and Entry Patterns (High Cost)

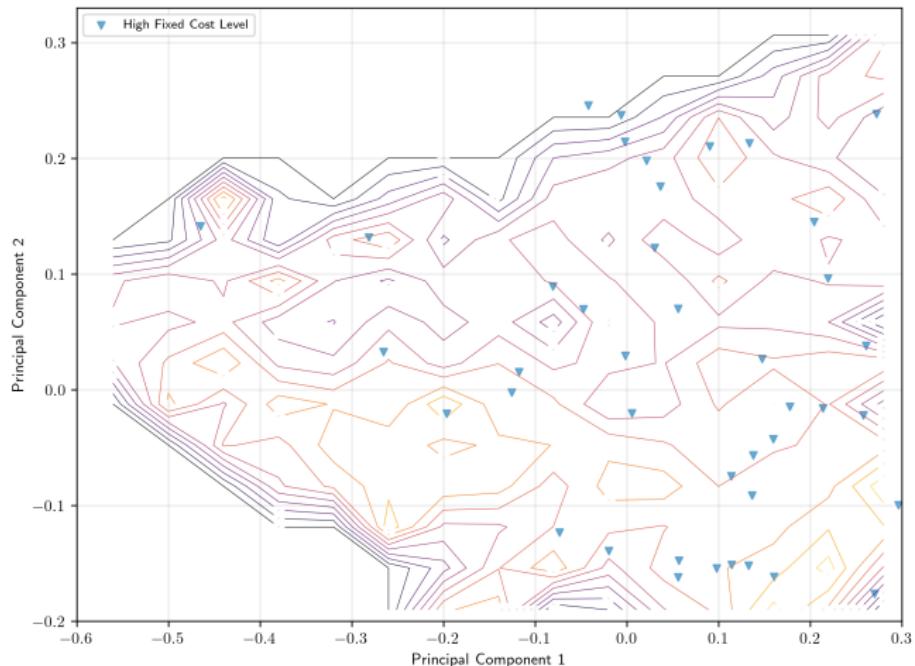


Figure: Entries under Copyright $d = 0.08$

Demand Distribution and Entry Patterns (High Cost)

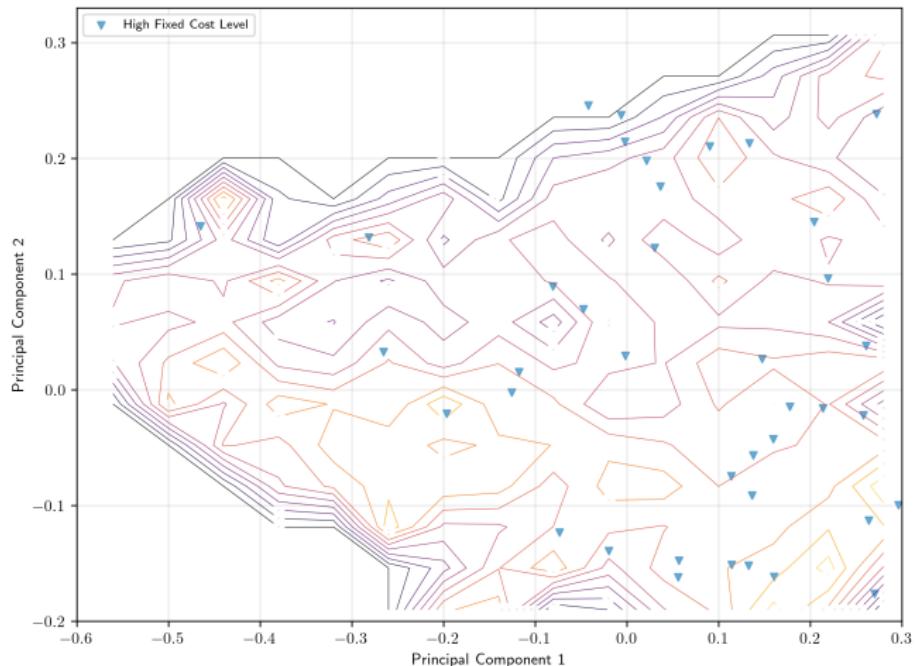


Figure: Entries under Copyright $d = 0.10$

Demand Distribution and Entry Patterns (Low Cost)

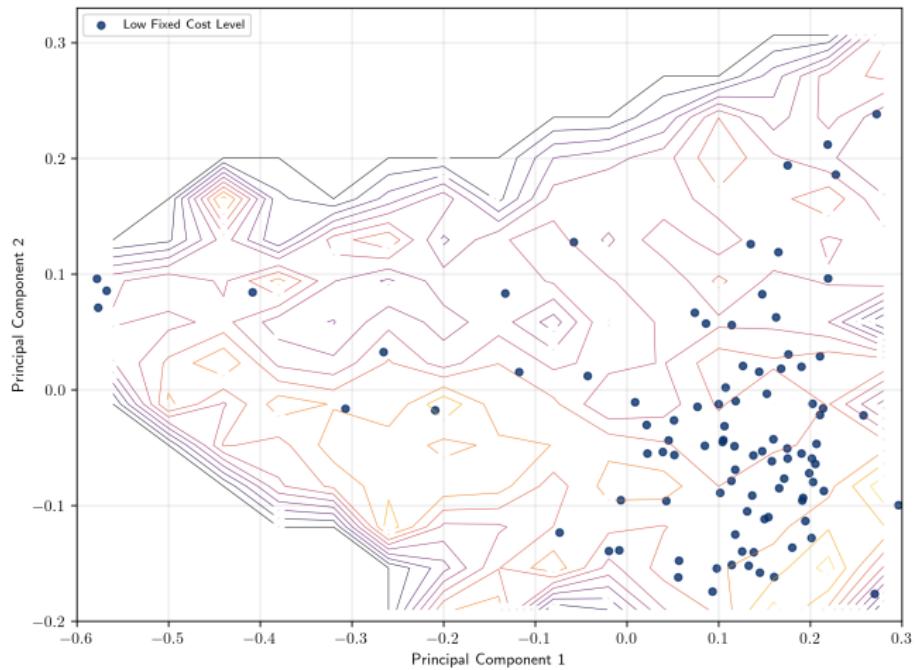


Figure: Entries under Copyright $d = 0.00$

Demand Distribution and Entry Patterns (Low Cost)

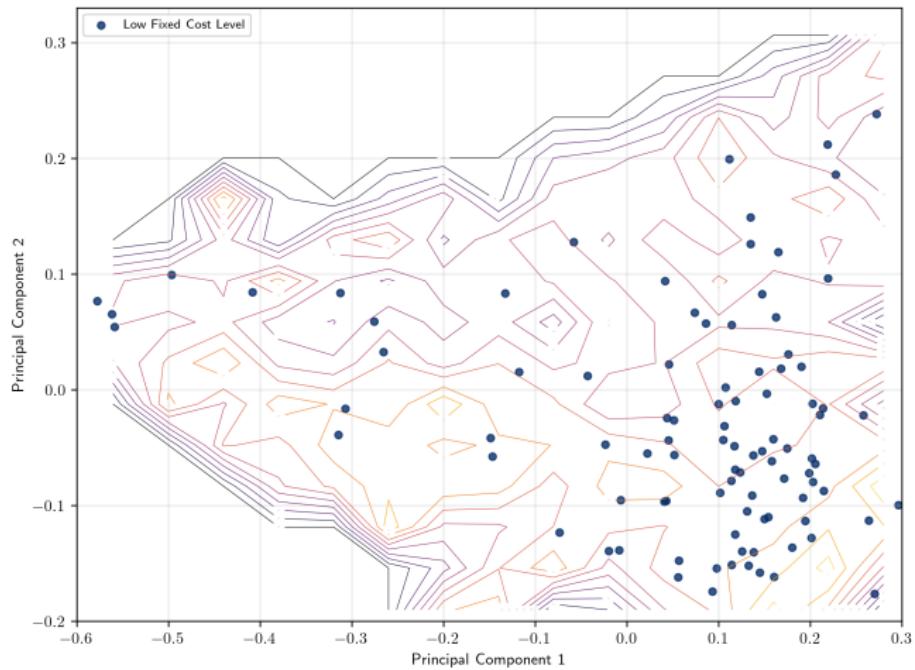


Figure: Entries under Copyright $d = 0.02$

Demand Distribution and Entry Patterns (Low Cost)

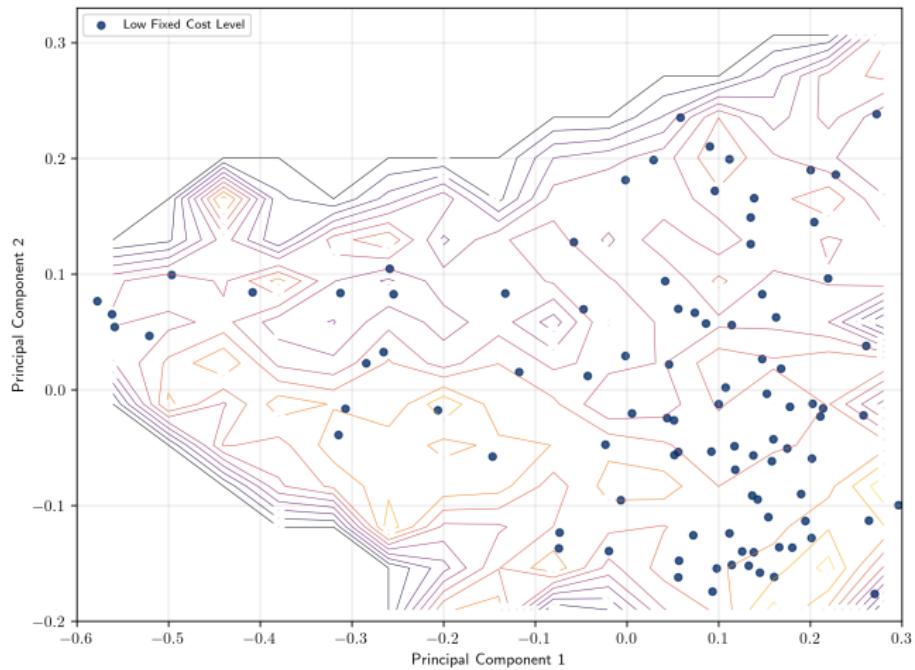


Figure: Entries under Copyright $d = 0.04$

Demand Distribution and Entry Patterns (Low Cost)

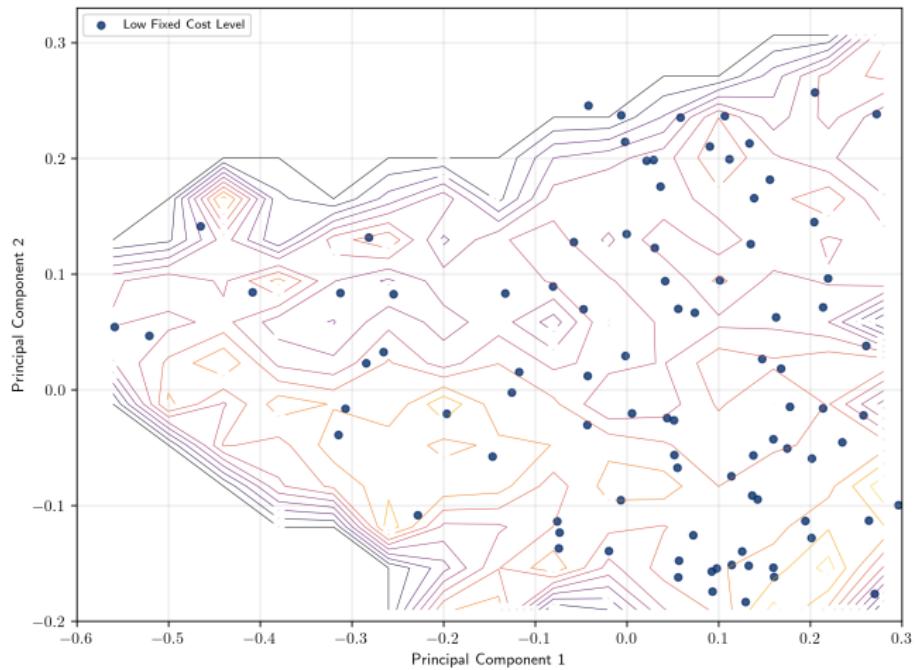


Figure: Entries under Copyright $d = 0.06$

Demand Distribution and Entry Patterns (Low Cost)

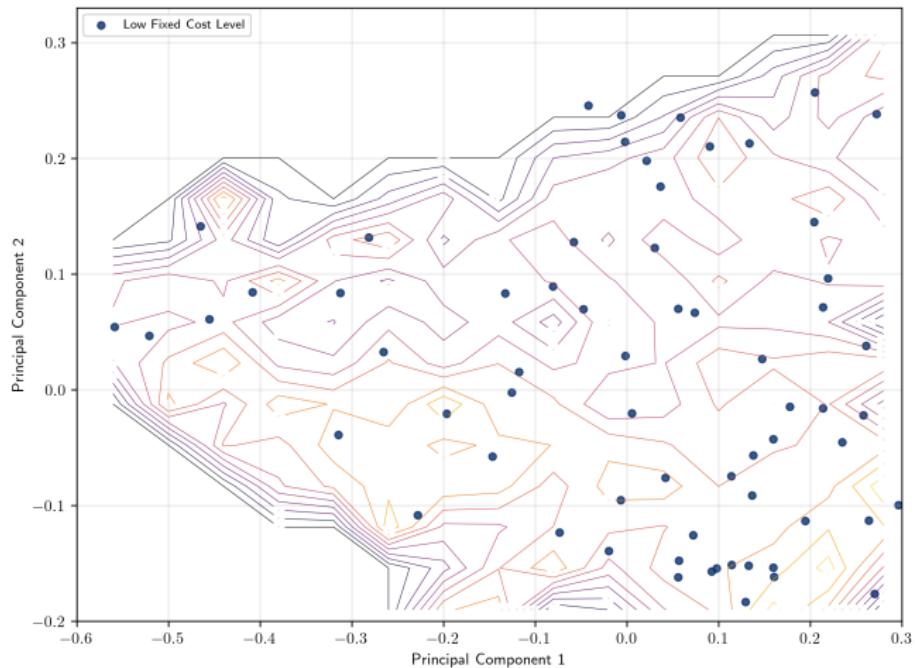


Figure: Entries under Copyright $d = 0.08$

Demand Distribution and Entry Patterns (Low Cost)

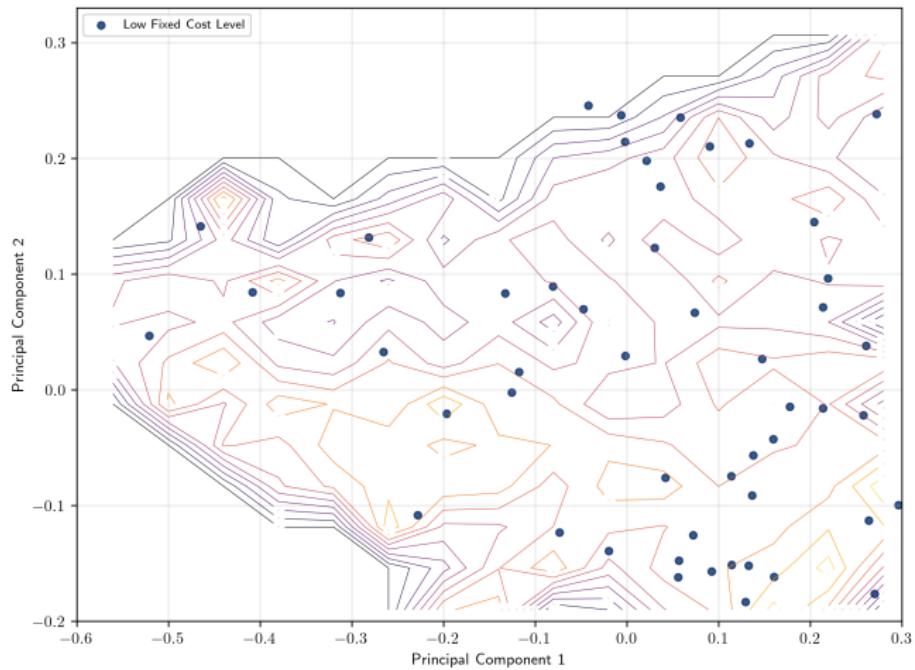
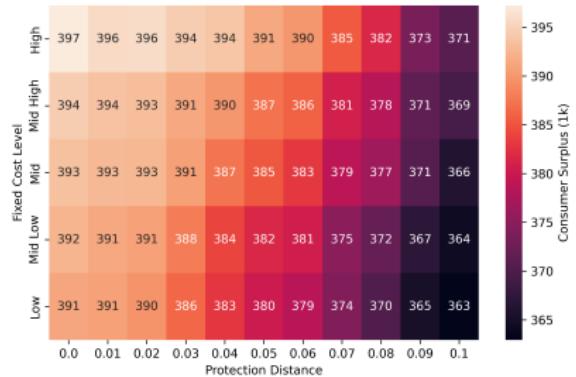
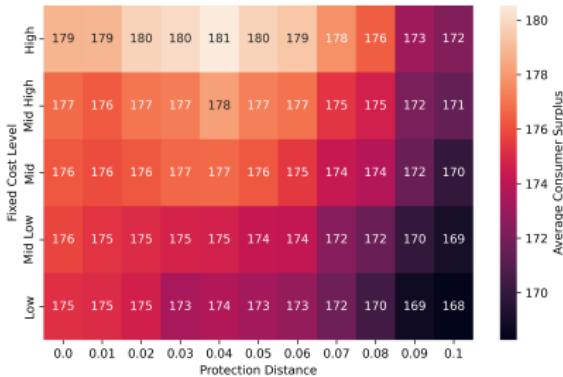


Figure: Entries under Copyright $d = 0.10$

Scenario B: Consumer Surplus



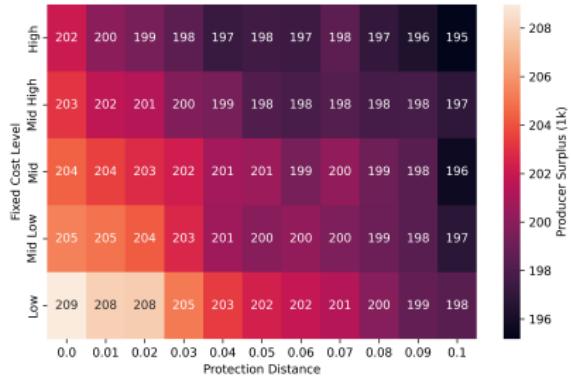
(a) Aggregate



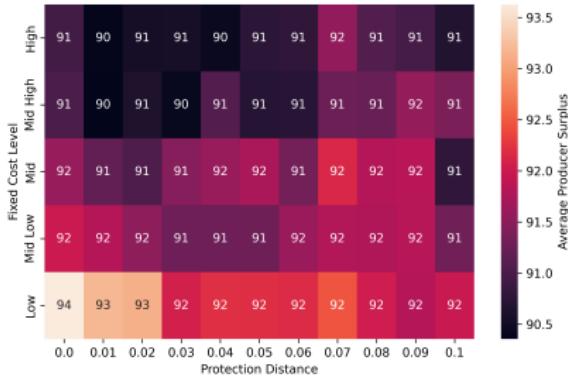
(b) Average

Figure: Consumer Surplus Across Protection Levels and Fixed Costs

Scenario B: Producer Surplus



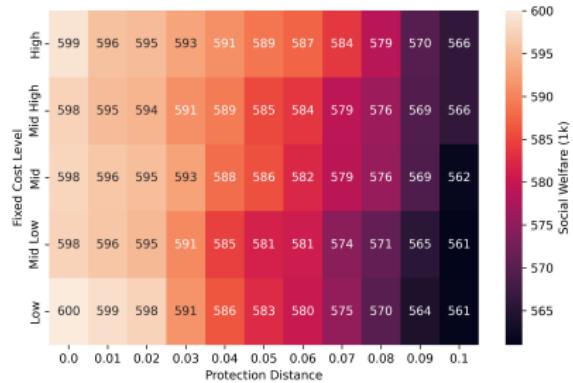
(a) Aggregate



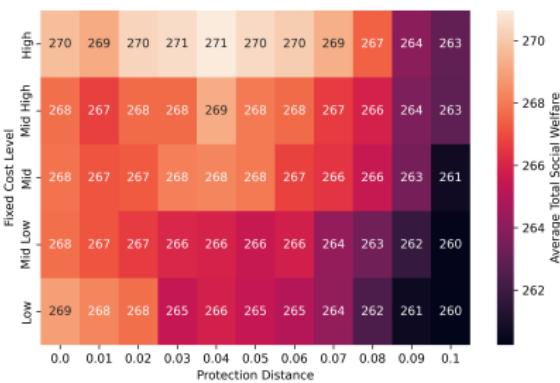
(b) Average

Figure: Producer Surplus Across Levels and Fixed Costs

Scenario B: Social Welfare



(a) Aggregate



(b) Average

Figure: Social Welfare Protection Across Levels and Fixed Costs

Counterfactual Analyses: Summary

Consumer surplus:

- ▶ inverse U-shape
 - esp. average surplus
- ▶ stricter protection incentivizes firms (except with low costs) to choose consumer-favorable locations
- ▶ while excessively strict protection harms consumers
 - entrants disappear from higher-demand areas
- ▶ consistent with Analysis 1's relocation exercise

Counterfactual Analyses: Summary

Producer surplus:

- ▶ U-shape
 - esp. aggregate surplus and lower costs
- ▶ business stealing effects are offset
- ▶ as protection becomes stricter
 - mimicry advantage drops (while # of entrants stays)
 - market expansion effect decreases (esp. with lower costs)
- ▶ as stringency further increases
 - plateaued high fixed costs, while dominating market expansion
 - but paid by less entrants

▶ Number of Entrants

Interaction between copyright policy and cost-reducing technology is essential for determining the optimal stringency of policy

Conclusions

Concluding Remarks

In a stylized market for design products, we demonstrate how unstructured data can be used in structural economic analysis

- ▶ type of policy experiments that are *not* possible with traditional data and approaches
- ▶ offer a scientific reference for copyright infringement judgments

Growing availability of unstructured data and ML tools

- ▶ new economic and policy questions
- ⇒ this paper's empirical models can be broadly applicable to other similar industries

Concluding Remarks

One important question: whether the embeddings capture...

- ▶ context-specific economic features (e.g., substitution patterns, local competition),
- ▶ while maintaining general interpretability (e.g., distance, visual similarity)

Thank You! ☺

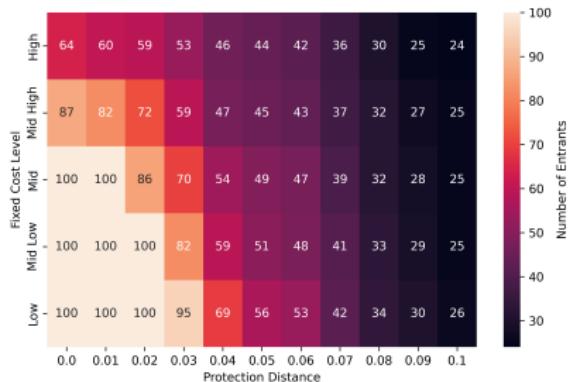
Supply Model

Timing:

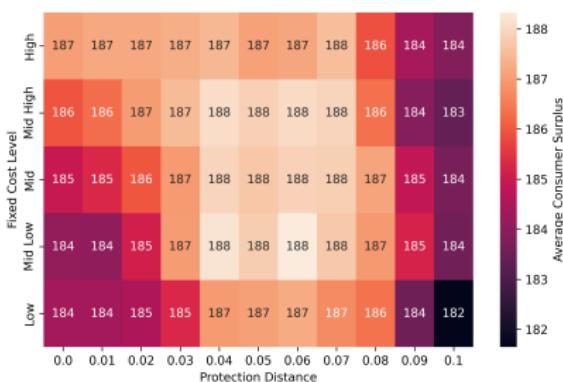
1. Firm f makes entry decision after cost shock is realized (and before the demand shock is realized)
2. Upon entry, f chooses the optimal location of product k
3. Unobserved demand shock (ξ_{kt}) is realized and f chooses price
(Eizenberg 14)

 Return

Scenario A: Number of Entrants and Fixed Costs



(a) Number of Entrants



(b) Fixed Costs

Figure: Number of Entrants and Fixed Costs

◀ Return