Shapes as Product Differentiation: Neural Network Embedding in the Analysis of Markets for Fonts*

Sukjin Han[1], Eric Schulman[2],
Kristen Grauman[3], and Santhosh Ramakrishnan[3]

[1] *Department of Economics, University of Bristol*
[2] *Department of Economics, University of Texas at Austin*
[3] *Department of Computer Science, University of Texas at Austin*

This Draft: November 12, 2020

Abstract

Many differentiated products have key attributes that are high-dimensional (e.g., design, text). Quantifying these attributes is important for economic analyses. This paper considers one of the simplest design products, fonts, and quantifies their shapes by constructing embeddings using a deep convolutional neural network. The embedding maps a font's shape onto a low-dimensional vector. Importantly, we verify the resulting embedding is economically meaningful by showing that the mutual information is large between the embedding and descriptions assigned to each font by font designers and consumers. This paper then conducts two economic analyses of the font market. We first illustrate the usefulness of the embeddings by a simple trend analysis of font style. We then study the causal effect of a merger on the merging firm's creative product differentiation decisions by using the embeddings in a synthetic control method. We

*For helpful discussions, the authors are grateful to Jorge Balat, Phil Haile, Greg Howard, Roger Moon, Jesse Shapiro, Matt Shum, Haeyeon Yoon, participants in the 2020 North American Winter Meeting of the Econometric Society, the 25th Texas Econometrics Camp, and the KAEA Virtual Seminar Series. We acknowledge the generous provision of the data set by Monotype Inc. Sukjin especially thanks Nathan Ford, the Customer Analytics & Insight Director of Monotype, for fruitful conversations throughout the project, and graphic designers, Jaewon Seok and Jiwon Park, for early discussions on font design and font markets. Lastly, we appreciate research assistants, George Kyungho Lee and Po-Yuan Huang, for their excellent work.
Corresponding author: Sukjin Han sukjin.han@gmail.com

find that the merger causes the merging firm temporarily to increase the visual variety of font design.

JEL Numbers: L1, C8.
Keywords: Convolutional neural network, embedding, high-dimensional product attributes, visual data, product differentiation, merger.

1 Introduction

1.1 Markets for Products with Unstructured Attributes

Many differentiated products considered in economic analyses have important attributes that are unstructured. Examples are products with design elements like automobiles, houses, furniture, and clothing. Other obvious examples are creative works like books, musics, and movies. Unstructured attributes in these products are typically in visual or textual forms and thus high-dimensional. More generally, products well beyond these categories are often presented to consumers in visual and textual forms. Examples are product packages in supermarkets and online catalogs in e-commerce (e.g., Amazon, Airbnb, Yelp, Zillow). These visual attributes are one of the first pieces of information consumers receive along with more structured attributes such as price and product specifications. As a result, high-dimensional attributes are important decision factors for consumers as well as key decision variables for producers.

1.2 Economic Models and Policy Questions

Economists are well aware of the importance of unstructured attributes. Product attributes are an important aspect of economic models, such as discrete-choice models (McFadden (1973), Berry et al. (1995)) and hedonic models (Rosen (1974), Bajari and Benkard (2005)). These models treat product attributes as both low-dimensional observable variables and a (typically scalar) unobservable variable. In these models, the scalar unobservable normally captures the high-dimensional, unstructured attributes including design and other artistic features.

Although this tradition has its own merits, certain policy questions are better answered by treating unstructured attributes as observables and attempting to solve the resulting dimensionality problem. For example, one may ask how product differentiation in this creative dimension affect the product's market power or is affected by policies such as vertical integration. One can also subsequently ask how to design optimal policies for protecting the original features of a product. Finally, one can ask how the style of products evolves over time—i.e., how the fashion changes—in accordance with market conditions. In this paper, we propose to quantify the design-oriented attributes of a product by constructing a low-dimensional product space of them, and answer some of these questions.

1.3 Why Markets for Fonts?

In this paper, we consider a particular design product: fonts. There are a few reasons to study the market for fonts. First, font is one of the simplest visually-differentiated products. The shapes (i.e., two-dimensional monochrome visual information) of a fixed number of characters mostly describe the product. These shapes are called typefaces. Second, this visual information is simple to understand, but important in predicting the functionality and the value of the product. Third, fonts are ubiquitous products and fonts markets are large with frequent productions and transactions. The online marketplace we consider has over 28,000 fonts and 2,400,000 transactions in the past six years. Fourth, there are interesting policies involved in this market, such as vertical integration and license agreements. Finally, font is a stylized product that captures key aspects that a myriad of products in the market have in common: design attributes. We also believe that the analytical framework of this paper can be applied to other industries (e.g., those mentioned earlier) where unstructured product attributes are important elements in economic analyses.

1.4 Embedding and Low-Dimensional Product Space

The main challenge in economically analyzing the market for fonts is that the main product attributes, their shapes, are high-dimensional. To address this challenge, we represent font shapes as low-dimensional neural network embeddings and construct a corresponding product space. Specifically, we adapt a state-of-the-art method in convolutional neural networks (Schroff et al. (2015); Wang et al. (2014)), where the network directly learns to map font images to a compact Euclidean space, i.e., the embedding space. Importantly, we verify that the resulting embedding is economically meaningful. We demonstrate that *tags* assigned to each font by font designers and consumers are highly associated with our image embeddings through the mutually shared information. Tags are word phrases that describe the fonts (e.g., "curly," "flowing," "geometric," "organic"), and thus, we believe, represent the economic agents' perceptions. To calculate the mutual information, we construct word embeddings from these tags.

Why do we consider a convolutional neural network? The convolutional neural network is designed to capture spatial correlation between, e.g., nearby pixels or musical notes. Font shapes involve a non-linear interaction between many neighboring pixels. Knowing just one individual pixel provides very little information about the overall shape of a font. By considering how neighboring pixels interact, the deep neural network outperforms other machine learning methods such as LASSO, random forest, and boosting that use pixels or other hand-designed features (edges, corners, etc.); see Goodfellow et al. (2016). While neural networks are generally known to be less interpretable (Friedman et al. (2001))[1], we show how an interpretable embedding space can be learned through visual similarity. In particular, instead of the attempt to interpret each embedding value, we give meanings to the distance metric of the embedding space for subsequent economic analyses.

1.5 Economic Analyses Using the Embeddings

The low-dimensional embeddings we create can be a basis for performing various economic analyses. This paper conducts two analyses of the font market. The key insight of our analyses is that, given the product space constructed via embedding, firms (called *foundries* in this industry) engage in Hotelling-type of spatial competition. The main decision variable of this competition is font design as creative product differentiation.

First, we illustrate the usefulness of the embeddings by a simple trend analysis of font style. We analyze the trend in the distance between individual font and a benchmark font, for which we use "Averia," the average of embeddings of all the fonts in the market. A notable finding is that, compared to incumbent foundries, new entrants tend to create fonts that are more experimental, further away from the average shape.

Second, we study the causal effect of a merger on the merging firm's product differentiation decisions by using the embeddings in a synthetic control method. We show that this treatment effect approach is suitable in the presence of the embeddings.[2] For the degree of product differentiation, we use the distance to Averia and an alternative gravity measure, both calculated using the embeddings. In June 2014, one of the major font foundries is acquired by the company that owns the online marketplace. For a causal analysis of this merger case, it is important to make a good comparison between the merging foundry and the other untreated foundries, which is a challenge since there is only one treated unit. To address this challenge, we use the synthetic control method (Abadie and Gardeazabal (2003), Abadie et al. (2010)). We construct a comparable control unit by using the embeddings as the main predictor. We find that, relative to the synthetic control unit, the merging foundry has temporarily produced more experimental fonts after the merger, and that this

[1] One approach to overcome this is to consider "hand-crafted features." However, feature selection can generally be arbitrary and there can be arbitrarily many possible features, which hinders the interpretation.

[2] See Section 1.6 for more discussions on structural approaches as alternatives to this approach.

effect is statistically significant. That is, the merger results in increased visual variety in font design of the merging firm. This possibly implies that the merger has occurred from a preemptive motive. The treatment effect approach to study vertical integration appears in, e.g., Hastings (2004) and Ashenfelter and Hosken (2008), although they typically use difference-in-differences methods and focus on price responses.

1.6 Contributions and Related Literature

Machine Learning and Social Science Research

To our knowledge, this is the first paper to use neural network embedding for visual data in the economic analysis of markets and industries. Glaeser et al. (2018) uses visual data from Google Street View to predict the economic prosperity of neighborhoods. They assign scores to street images based on human surveys on visual perception of street quality and safety. Gross (2016) investigate how competition influences creative production in commercial logo design competition. While Gross (2016) uses hand-crafted features to create a perceptual hash code for comparing images, we learn image embeddings based on recent advances in deep learning that have been shown to work extremely well for high-dimensional image data (Krizhevsky et al. (2012); Simonyan and Zisserman (2014); He et al. (2016)).

This paper is also among the first social science studies that uses embeddings as part of empirical analyses. As another form of unstructured data, text data have recently gained much attention in economic analyses; see Gentzkow et al. (2019a) for a thorough review of the machine learning applications. Kozlowski et al. (2019) have used word embeddings to understand cultural norms. Gentzkow et al. (2019b) analyzes political polarization using congressional speeches as text data. Hoberg and Phillips (2016) use text data from firms' 10-K product descriptions across industries to classify competing products and construct a product location space as we do in our paper.[3] Unlike their paper, however, we use image embeddings and employ neural network as a classification method. Also, we focus on a particular industry as opposed to multiple industries, and utilize detailed structured and unstructured data about product offerings.

Merger and Product Differentiation

Market structures and product differentiation have been important themes in the field of economics. In a theoretical paper, Mazzeo et al. (2018) finds the effects of a merger on product differentiation can be ambiguous, implying that the question is more of empirical research

[3] In fact, our paper utilizes both image embeddings and word embeddings as part of our analysis.

as in empirical industrial organization. Berry and Waldfogel (2001) document the effect of mergers on product variety in local radio markets by exploiting the natural experiment provided by the 1996 Telecommunications Act. Sweeting (2013) studies the dynamic aspects of product differentiation in the radio industry. They find some evidence that increased concentration increases variety. Finally, Fan (2013) finds in newspaper markets that mergers between local competitors has effects on vertical differentiation, making firms reduce news quality. Unlike the structural approach of these papers, Hastings (2004) and Ashenfelter and Hosken (2008) study the effects of mergers on prices from an angle of the program evaluation literature. We follow the latter approach in the causal analysis of merger. The latter approach is not ideal in analyzing responses to changes in market environment that are never observed and calculating welfare.[4] Still, we believe this paper can be a good starting point to introduce embeddings in economic analyses. All the empirical papers above use structured data.[5] Our paper is distinguished from them in that we use unstructured data of images to derive new insights about the market, such as the effect of merger on creative product differentiation.

In the econometrics literature, structural approaches are considered in the presence of high-dimensional controls, but mainly from structured data. Gillen et al. (2015) develop BLP-LASSO with high-dim demographics as controls. Chernozhukov et al. (2016) consider orthogonal machine learning with high-dimensional demand signals as controls. Related to the latter, Chernozhukov et al. (2018) consider a robust inference of low-dimensional structural parameters in the presence of high-dimensional nuisance parameters. We are cautious about using such methods because in our setting, design attributes are endogenous decision variables that cannot simply be partialled-out. However, recent advances in machine learning may enable us to conduct more sophisticated policy analyses with both high-dimensional endogenous variables and control variables; see e.g., Foster and Syrgkanis (2019). This can be an interesting direction for future research.

Machine Vision

Recognizing letters (e.g., distinguishing handwritten "G" from "Q") is one of the most well-studied area of machine vision as with the MNIST database (LeCun et al. (2010)). Our paper, however, is one of the first that applies machine vision techniques to recognizing the *style* of font images (e.g., distinguishing typeface "G" from "G"), which is a more challenging vision problem. O'Donovan et al. (2014) develop a method of searching fonts using relative

[4] See Angrist and Pischke (2010) and Nevo and Whinston (2010) for pros and cons of these two approaches.

[5] For example, Fan (2013) uses data on the number of opinion section staffs, the number of reporters, the local news ratio, variety, the frequency of publication, and edition. Sweeting (2013) uses Neilson data on broadcasts.

attributes. It builds on the work on attributes and whittle search by Parikh and Grauman (2011) and Kovashka et al. (2012). Campbell and Kautz (2014) develop a procedure of learning font manifold by parametrizing font shapes and reducing the dimension of the resulting model.

The method of training the font embedding builds on Schroff et al. (2015) and Wang et al. (2014), who develop a face recognition algorithm that learns embedding for images directly in a neural network training. Schroff et al. (2015) show that their approach performs substantially better than the earlier approaches of training a classification network for face recognition as in Taigman et al. (2014) and Sun et al. (2015). The former approach is suitable for our purpose, since the procedure produces embeddings as the intermediate output of the algorithm of classification. Although we are not directly interested in the classification of font identity, embeddings serve as our object of primary interest.

Fonts can be viewed as fashion products. Our *quantitative* analysis of the trend in font style is related to, e.g., Al-Halah et al. (2017), Mall et al. (2019), and Yu and Grauman (2019), who apply advanced machine vision techniques they develop to recognize the style of clothes and shoes in the fashion industry and understand the trend. The analysis of visual attributes of design products has also been considered, e.g., in Burnap et al. (2016) and Dosovitskiy et al. (2016) using deep generative models with applications to furniture and automobile designs.

1.7 Organization of the Paper

In the next section, we provide the background about the font industry and the online marketplace for fonts considered in this paper. Section 3 describes the data obtained from this market. In Section 4, we construct the embedding and the product space using the neural network. In this section, we demonstrate that the embedding is meaningful by calculating the mutually shared information between the font embeddings and tags. Sections 5 and 6 contain two economic analyses that utilize the embeddings. The causal analysis of merger can be found in Section 6. Section 7 concludes.

2 Online Marketplace for Fonts

We consider the world's largest online market place MyFonts.com that sells around 30,000 different fonts. This market is a superset of all major global online stores. MyFonts.com and the other stores are all owned by Monotype Inc. To be precise, a font is a delivery mechanism for typefaces. Therefore, fonts are sold as a piece of software, for which consumers purchase

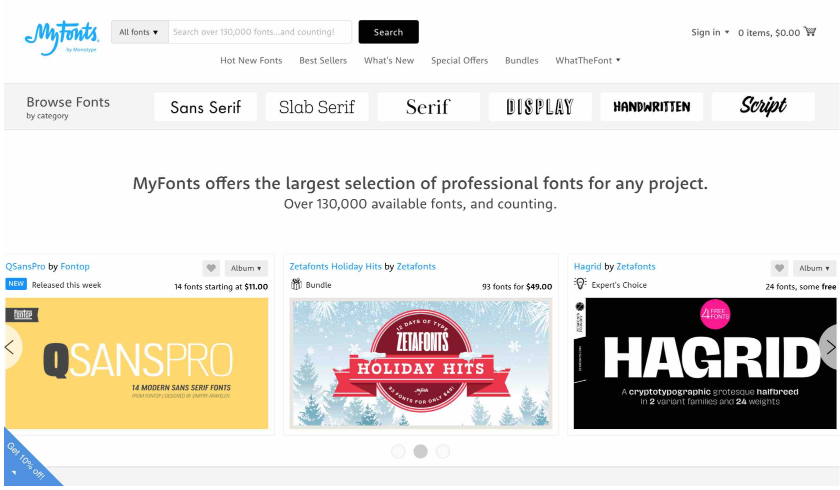


Figure 1: Main Page in MyFonts.com

a license. Licenses are protected by the End User License Agreement (EULA). There are generally two types of license sold to consumers. A web font license allows fonts displayed on a website, and a desktop license is for printed materials. The marketplace also serves as a platform for third party fonts. As a result, it sells fonts designed by foundries owned by Monotype, as well as fonts from third parties foundries.[6]

Figure 1 shows the main page in MyFonts.com. An example of a font family page in MyFonts.com is captured in Figure 2. In this market, typical consumers are other designers who use fonts as intermediate goods. They produce printed materials (e.g., posters, pamphlets, cards), for which a desktop license is purchased, or webpages and digital ads, for which a web license and digital ads license are purchased, respectively. Between the data period of 2012 and 2018, around 2,400,000 purchases were made.

3 Data

3.1 Overview

We have the sample of period from 2000 to 2017. There are in total 28,659 fonts and 2,446,604 orders in the data set. The main information contained in the data set is information on product attributes for each product with a unique ID and on transactions for each user (i.e., consumer) with a unique ID. In terms of high-dimensional attributes, we have images of typefaces and tags (i.e., descriptive words assigned by producers or consumers). In terms of structured characteristics, we have price, category types, license types, the number of

[6] A foundry is a group of designers that produce fonts.

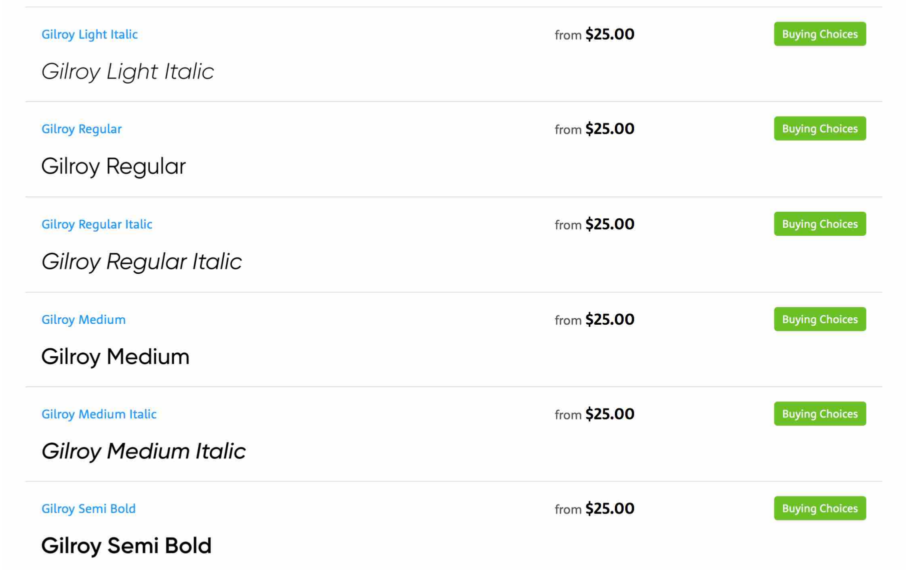


Figure 2: A Font Family Page in MyFonts.com

languages supported, the number of glyphs supported, the foundry and designer information, and the date introduced in the market. There are roughly six categories: sans serif, serif, display, handwritten, and script. There are five license types: desktop, web, apps, ePub, and digital ads.

Transactions data are individual orders made by users. For each order, we have information about the product ID, the time of purchase (by minutes), the total amount paid, and the user characteristics such as the country and city of origin, the order history, the life span and the total spending.

3.2 Visual Attributes

Fonts are displayed on the webpage using pangrams.[7] Pangrams effectively capture important design elements that cannot be seen in individual letters, such as spacing, deep-height, up-height, and ligature. Figure 3 shows the examples of pangrams that roughly correspond to the product categories (i.e., sans serif, serif, display, handwritten, and script). The format of pangram images is a bitmap with 200×1000 pixels. Each pixel is a vector of 3 integers between 0 and 255. We use 100×100 pixel (crops of) pangrams as direct inputs in the neural network in order to mimic a consumer's actual perception of the products.

[7] A pangram is a sentence that contains all the alphabet letters. In fonts markets, many different pangrams are used, such as "The quick brown fox jumps over a lazy dog.", "Six quite crazy kings vowed to abolish my pitiful jousts.", "Quincy Jones vowed to fix the bleak jazz program.", or "Mozart's jawing quickly vexed a fat bishop." Here we chose to use one of the shortest pangrams to minimize the size of the image.

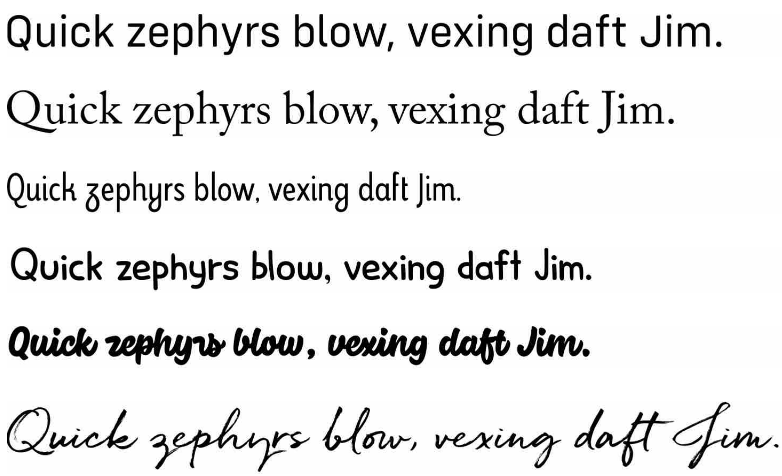


Figure 3: Examples of Pangrams (by categories)

4 Construction of Embeddings

4.1 Network Embedding

We employ a method where the network directly learns a mapping from pangram images to a compact Euclidean space. This mapping is called an embedding. We map each pangram to a 128-dimensional embedding, denoted as $f(x) \in \mathbb{R}^d$ for pangram image x and $d = 128$.[8] Then, L^2 distance corresponds to measure of similarity of font shape. For training the network embedding, we adapt a modern algorithm developed by Schroff et al. (2015) for face recognition. Their approach determines the identity of a person based on face images in two steps. In the first step, they train a deep convolutional network to learn an embedding space of faces. The rationale is that similar faces should occur closer in the embedding space than dissimilar faces. In the second step, they classify the identities of the images by choosing a threshold in the space below which the embeddings have the same identity. They show that this approach performs substantially better than the earlier approaches of training a classification network (Taigman et al. (2014); Sun et al. (2015)).[9] Schroff et al. (2015)'s approach is suitable for our purpose. First, as detailed below, pangram images can be naturally classified based on a structure that is analogous to that for face images where multiple images are associated with the same identity. Second, the procedure produces embeddings as the intermediate output of the algorithm. Although we are not directly interested in the classification of font identity, embeddings serve as our object of primary interest.

[8] We additionally normalize that this embedding lies on a d-dimensional hypersphere, $\|f\|_2 = 1$.

[9] The latter is an indirect approach where a classification network is trained and an intermediate bottleneck layer (typically a dimension of 1000s) is taken to generalize face recognition beyond identities in training.

4.2 Triplet Loss

The key idea of our strategy is to obtain positives from the same family (e.g., Helvetica) but different styles (e.g., Helvetica Regular, Helvetica Light, Helvetica Bold, Helvetica Italic) or crops (e.g. different words within the pangram), and negatives from different families (e.g., Time New Roman). The neural network then learns to group fonts that are within the same family in the embedding space.

In practice, we accomplish this by constructing triplets of images (Weinberger et al. (2006)). To be specific, triplet i consists of anchor x_i^a, positive x_i^p, and negative x_i^n, for which we want the following inequality to hold during training:

$$||f(x_i^a) - f(x_i^p)||_2^2 + \alpha \leq ||f(x_i^a) - f(x_i^n)||_2^2 \tag{1}$$

for all $(x_i^a, x_i^p, x_i^n) \in T$, where α is an enforced margin. Again, an anchor is a pangram crop of a given font family, positives are different crops of the same pangram or crops of pangrams of different styles in the same family, and negatives are crops of pangrams of different families. The way triplets are identified is analogous to that in the face recognition problem, where positives are images of the same person as the anchor and negatives are images of different persons. The margin α allows the images for one font family to stay on a manifold, while still discriminating the images of other families.

Then, a triplet-based loss function that is minimized in the network training is

$$L = \sum_i^N [||f(x_i^a) - f(x_i^p)||_2^2 - ||f(x_i^a) - f(x_i^n)||_2^2 + \alpha]_+.$$

We optimize this objective using stochastic gradient descent (SGD; Bottou (2010)). SGD is an iterative method for optimizing an objective function, in our case, to create a reasonable embedding space. Because our dataset is gigabytes in size, it would be too computationally challenging to compute gradient of the entire data set and optimize using a more traditional optimization algorithm (e.g., Nelder-Mead or the conjugate gradient method). SGD can be regarded as a stochastic approximation of gradient descent optimization. It replaces the actual gradient by an estimate of the gradient, based on a randomly selected subset of the data called batches.

We use roughly 20,000 images of fonts to train the neural network. The training iteratively improves the parameters using smaller batches of roughly 150 cropped images of fonts to estimate the gradient and then update the parameters accordingly (Wilson and Martinez

(2003)).[10] These 150 images are structured as 50 triplets. As the gradient is evaluated at more batches, the parameters in the network are adjusted.

Each embedding we produce has 128 dimensions. It is important to have enough dimensions so that the neural network can allow for variation in the embedding space based on the actual images. The 128 dimension was also used in the face classification task by Schroff et al. (2015).[11]

4.3 Constructed Product Space

From the trained neural network, we create a 128-dimensional space of font products. To visualize this space, we project it into a two-dimensional space using principal component analysis (PCA). Figure 4 depicts this two-dimensional product space. Each thumbnail corresponds to the embedding of each font style. For expositional purpose, it is created with the word "Quick" that is cropped from the full pangram. By visually inspecting the space, we can see even in this two-dimensional project that different styles of fonts are clustered together. More specifically, we can see the right hand side of the x-axis corresponds to fonts that are have narrower characters and the left hand side has fonts that are thicker. Additionally, the top of the y-axis includes fonts that are visually less standard and more dramatic.

To understand how well the fonts are clustered in the 128-dimensional space, we conduct the K-means clustering with the embedding. The result is visualized in the two-dimensional space in Figure 5, where each colored box corresponds to a particular font cluster. Even in this space with the dimension much lower than 128, we can see that fonts with similar styles are clustered together. To further illustrate the result of Figure 5, Figure 6 shows the examples of four different nearest neighbors and their corresponding pangram images.[12] We can see that visually similar fonts are near from each other in the resulting embedding space.

4.4 Internal Evaluation

First, we describe how the neural network performs in the original classification task of identifying the font family. Although we are interested in the embeddings and not the classification, it is important to evaluate the embeddings based on their ability to classify. If

[10] We cropped each image based on the number of characters in the image. For example, there are 20 characters in the first half of the pangram sentence, so to crop 5 characters, we would take 40 percent of the pixels in the first half of the image. We tried crops with 3, 4, 6, and 7 different characters. We also tried different cropping schemes like using the white space between characters.

[11] The embedding with a larger dimension would perform better, but it requires more training data to achieve the same level of accuracy while avoiding the risk of overfitting.

[12] To increase the visibility, we randomly sample a smaller set of images and obtain their nearest neighbors in the two-dimensional space.

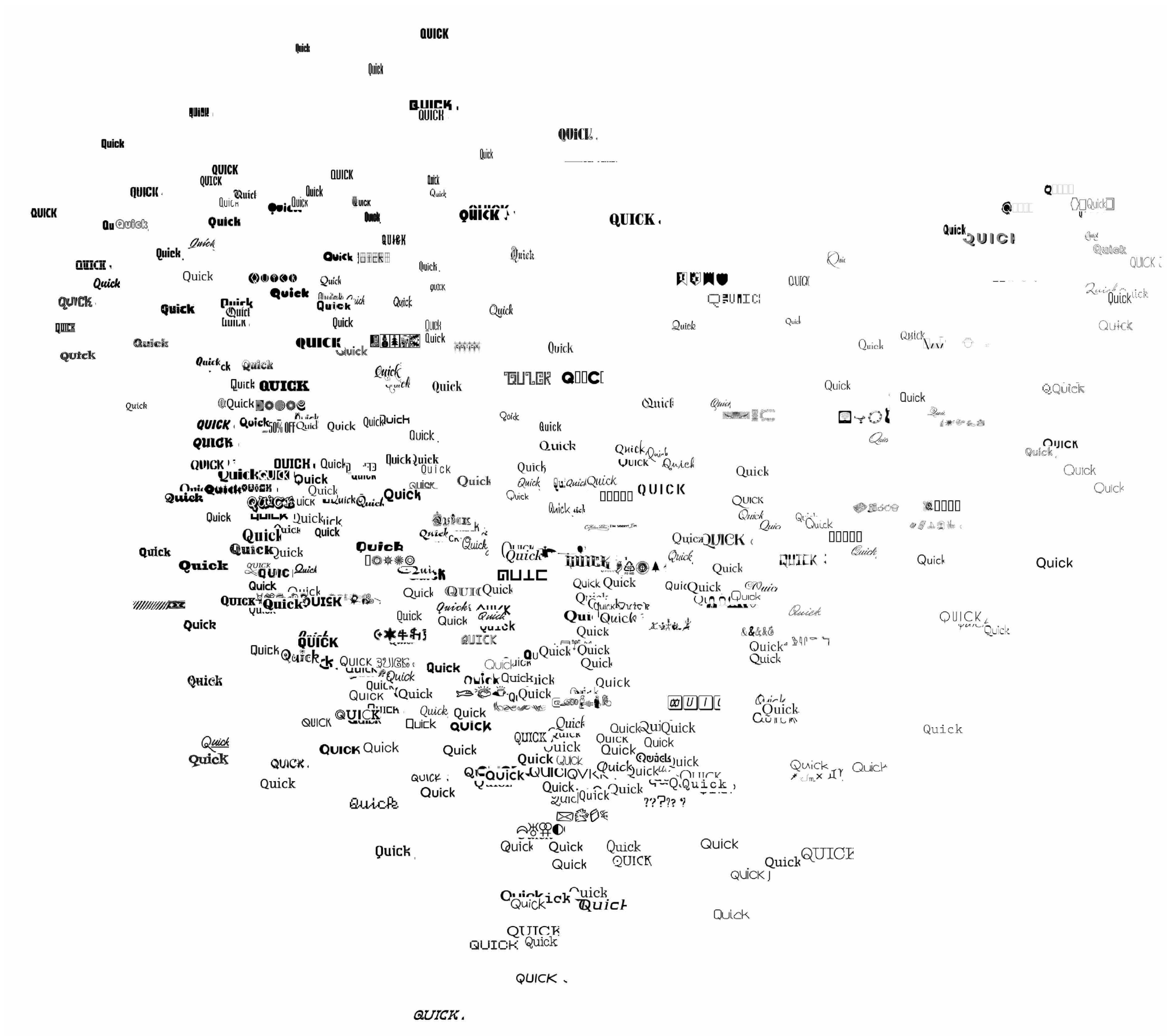


Figure 4: The Space for Fonts (projected onto two dimensions)

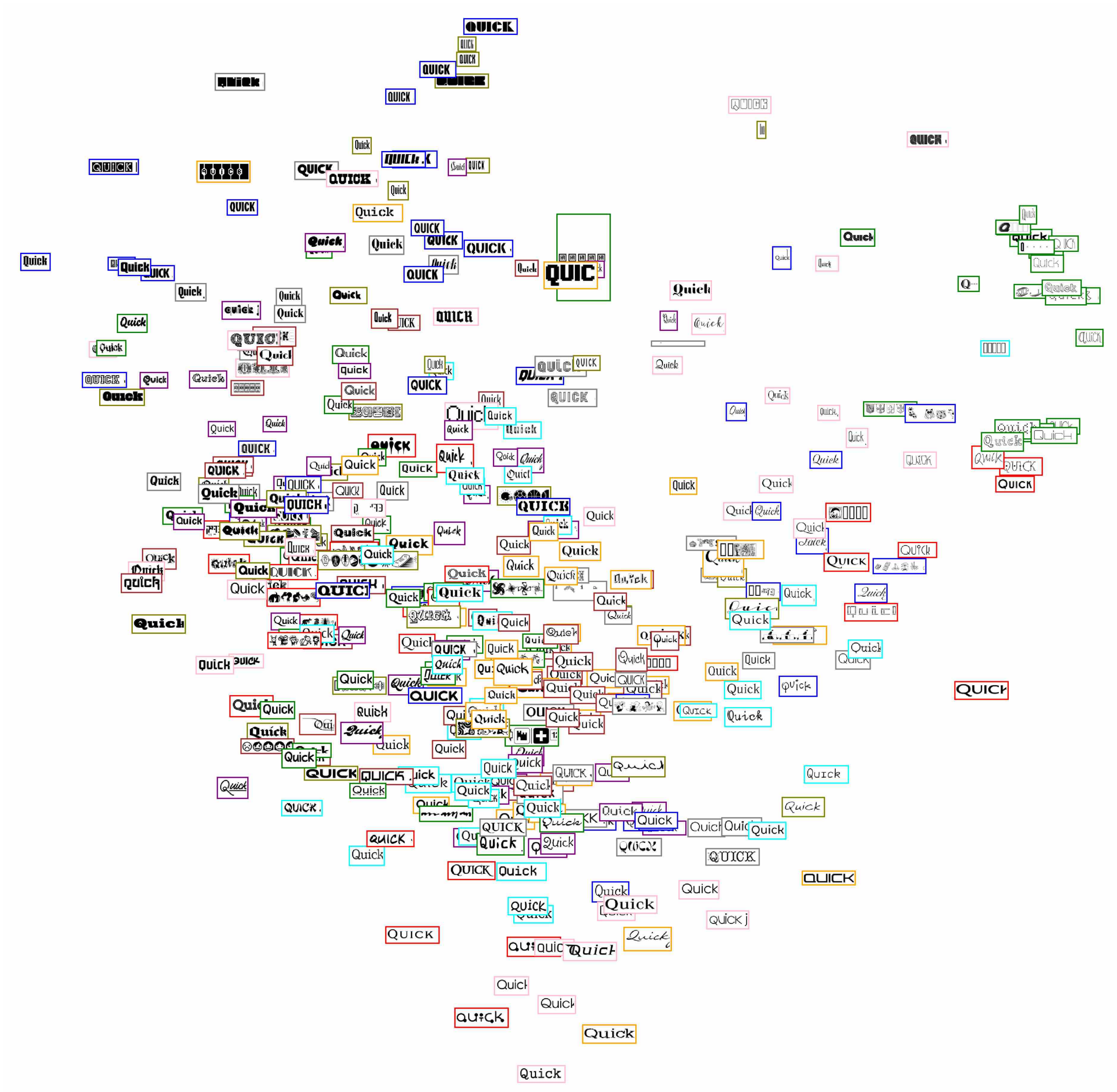


Figure 5: Clusters of Fonts Based on Embeddings (projected onto two dimensions)

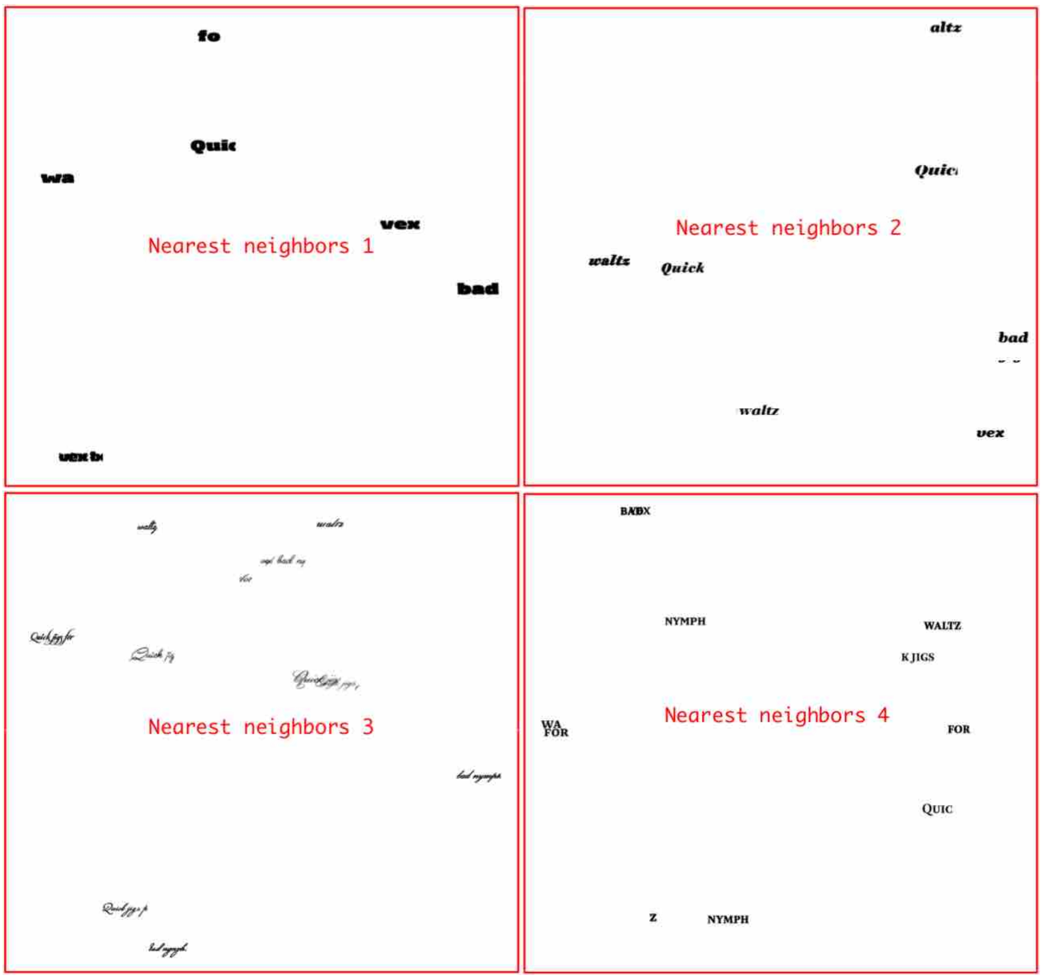


Figure 6: Examples of Nearest Neighbors and Corresponding Pangrams

Crop Size (Characters)	Test Set	Accuracy	Validation Rate	FAR
7	Hard	0.8925	0.09375	0
7	Easy	0.8975	0.47875	0
6	Hard	0.8825	0.04875	0
6	Easy	0.89667	0.53875	0
4	Hard	0.86333	0.02	0
4	Easy	0.8925	0.46875	0
3	Hard	0.76417	0.00875	0
3	Easy	0.88167	0.48625	0

Table 1: Internal Validation by Accuracy, Validation Rate, and FAR

the embeddings perform poorly, then they are not likely to be a reliable product space that is economically meaningful. As mentioned, classification is conducted by thresholding distances between embeddings. Overall, the neural network embeddings perform well in differentiating between fonts in different families.

To evaluate the neural network embeddings, we create test sets of triplets that the neural network has never seen during the training. With the test sets, the task is to identify whether the image in a triplet is a positive or negative. It is classified as a positive if its L^2 distance from the anchor in the trained embedding space is less than a pre-defined threshold (via cross validation) and a negative otherwise. We describe the results of tests on two different test sets. The first "easy" test set randomly samples pairs of a positive crop within the same family as an anchor and a negative crop from different families, for which the performance is evaluated. The second "hard" test set samples pairs of a positive crop within the same style of a family as an anchor and a negative crop from different styles within the same family.

The analysis involves true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN). In Table 1, we provide the accuracy and validation rates for the neural network in each of these test sets with different crop size. The accuracy is a measure of how well the neural network embeddings do at the classification task. Overall, it gets roughly 90 percent of triplets correct.

We also analyze the trade-off between Type-I and Type-II errors in classification. The validation rate is the true acceptance rate and is calculated as TP/(TP+FN). The false acceptance rate (FAR) is given by FP/(TN+FP). Additionally, we provide precision and recall curves. Precision is related to the Type-I Errors and is calculated as TP/(TP+FP). Recall is related to the Type-II Errors and is given by the formula TP/(TP+FN). The precision recall curve in Figure 7 shows the trade-off between these types of errors. There exists a steeper trade-off between precision and recall in the hard test set.

“Easy” Test Set | “Hard” Test Set

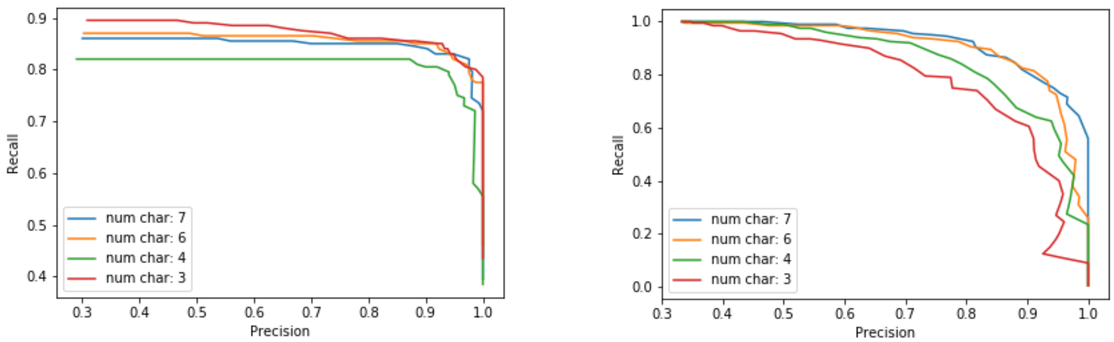


Figure 7: Precision Recall Curves

Overall, all the statistics we report here show that our neural network performs well. The low validation rate in the hard test sample suggests there is room for improvement for differentiating between different styles.

4.5 External Verification Using Word Embedding

Although the neural network embedding performs well in terms of prediction, it does not necessary imply the relevance to economic analyses, Next, we verify that visual attributes captured in the resulting embedding are relevant to economic agents' perceptions, by measuring a generalized notion of correlation with “perceived” attributes. We use information from “tags,” which are short descriptive words assigned to each font family by font designers and consumers. Examples of tags are “curly,” “flowing,” “geometric,” “organic,” “decorative,” and “contrast.”

These descriptive words of fonts are also high-dimensional, as there are nearly 30,000 different words in the tags. As a result, we consolidate tags that are synonyms to create meta-tags that encompass many related words (e.g. “script” and “cursive” would be in the same tag). To accomplish this task, we create clusters of tags using a standard word embedding “Word2vec” (2-layer neural network) by Mikolov et al. (2013). Then, we create clusters of tag embeddings using the K-means clustering from word embeddings. Finally, we measure how well they match clusters of visual embeddings we use mutual information.

To do the K-means clustering of font shape, we use the 128-dimensional shape embeddings we obtained, clustered directly into 60 clusters. To do the K-means clustering of tags, we use 100-dimensional word embeddings, reduced to 10-dimensional and then clustered into 6 clusters.[13] To evaluate how well the two clusters match, we used the normalized mutual

[13]For the number of clusters for image embeddings, we choose 60 clusters for two reasons. First, we use

F	W	$NMI(F, W)$
Image Embeddings	Tag Embeddings	0.473
Structured Tags	Tag Embeddings	0.261

Table 2: Normalized Mutual Information between Image Clusters and Word Clusters (first row), Compared to Baseline (second row)

information (NMI).

$$NMI(F, W) = \frac{I(F, W)}{[H(F) + H(W)]/2}$$

In this formula F is the distribution of clusters based on shape embeddings, W is the distribution of clusters based on word embeddings, $H(\cdot)$ is information entropy, and $I(F, W) = H(F) - H(F|W)$ is mutual information between F and W. The NMI can be interpreted as how informative W is in determining F. Its value will be between 0 and 1 (0 being W contains no information for F). Table 2 reports the values of NMI for two different pairs of distributions. The first row corresponds to the NMI for the image and word clusters described above. We obtain $NMI(F, W) = 0.473$, which is quite promising. To compare this value with the baseline case, the second row of the table shows the NMI when the pre-existing product categories are used in clustering as structured attributes instead of the image embeddings. In general, if images were treated as unobserved product attributes, then the main structured attribute available in many design products are product categories. In our context, the structured tags (i.e., sans serif, serif, display, handwritten, and script) are such product categories defined by the font industry and assigned by the producers and users. In this case, we obtain $NMI(F, W) = 0.261$, which is roughly only a half of the value in the first case. This indicates that the visual embeddings learned arguably capture economic agents' perceptions and thus can be relevant for economic analyses.

5 Trend Analysis

To illustrate the usefulness of the shape embeddings, we analyze the supply- and demand-side trends in font style. This part also provides relevant backgrounds for the causal analysis that

the heuristic elbow method. This involves calculating the mean squared error between the data and the cluster centers. One picks the point where you see an elbow or diminishing returns. This method is known to be sensitive to the coordinate scale. To choose a suitable scale, we use the fact that image embeddings had $2^6 = 64$ different mutually exclusive tag clusters. We try to find a similar number of mutually exclusive K-means clusters so that they would be comparable. We also select the number of tag clusters using the elbow method. We choose the scale by trying to match the 6 pre-existing structured tags.

follows in the next section. On the supply-side there are constant entries of new products in the marketplace. On the demand-side there are on average more than 1,000 fonts (stably) sold per day. Of course, demand and supply are endogenous, so the analysis is only a descriptive analysis. We, however, believe it reveals some interesting patterns in this market.

For this analysis, we calculate the L^2 distance between individual font and a benchmark font[14]: For font (image) x_i,

$$D_i \equiv \|f(x_i) - f_{averia}\|_2 ,$$

where $f(\cdot) \in \mathbb{R}^d$ is the embedding and f_{averia} is the average embedding of all font images. This distance measure D_i is intended to capture the degree of product differentiation for font x_i. For the benchmark font, we introduce "Averia," which is calculated by averaging the values of the embeddings of all the fonts existing in the market. The distance measure D_i, as well as an alternative measure of product differentiation, is used in the causal analysis in Section 6.

5.1 Supply-Side Trend

We analyze the supply trend in the L^2 distance from Averia. We are especially interested in how the style of fonts newly entering the market differs from the style of incumbent fonts. In Figure 8, each dot in the figure represents the daily mean (the upper panel) or median (the lower panel) distance from Averia for a range of periods. The red horizontal line is the mean/median distance of *all* fonts introduced before 2001, which we view as incumbents. The black line depicts the quarterly moving averages for the daily mean/median for fonts introduced from 2001, which we view as entrants. Overall, we find that entrants have font shapes that are more experimental or innovative than incumbents, possibly to avoid competition and establish market power distant from the incumbents in the product space.

5.2 Demand-Side Trend

We now analyze the trend in the L^2 distance between Averia and fonts purchased by consumers between 2014 and 2017.[15] To remove the supply-driven factor, we condition on price and focus on the price ranges of \$25-35 and \$35-45 where promotions are rare. This price range is the most frequent in the market as shown in the histogram of prices in Figure 9.

[14]For each font embedding in our economic analyses, we use a representative embedding in a given family, such as the embedding of a regular style.

[15]The time span is shorter than the supply-side analysis due to the missingness of the license type data for earlier periods.

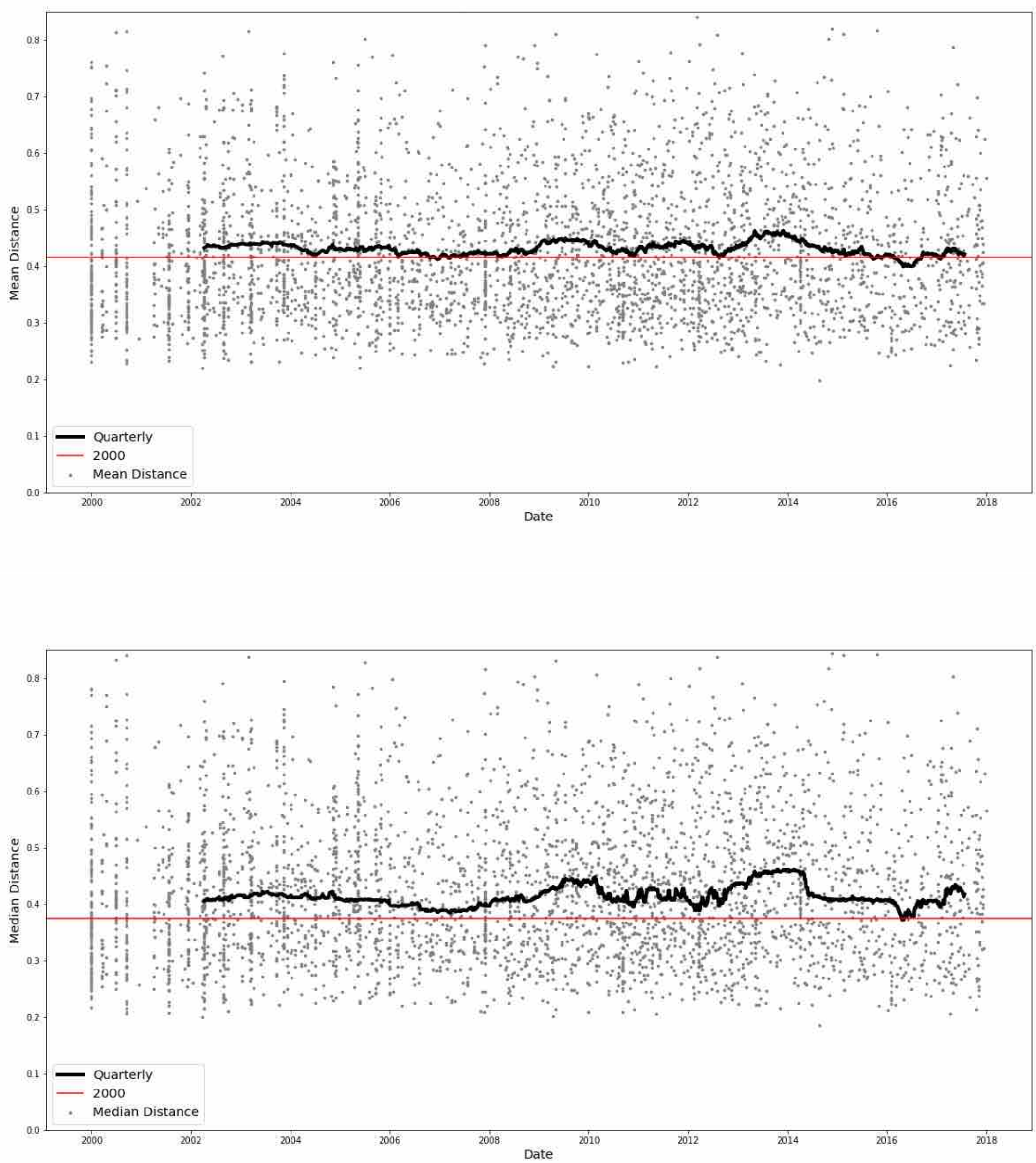


Figure 8: Trends in Mean/Median Distance of Incumbents ($\leq$ 2000, in red) and New Entrants (2001–2017)

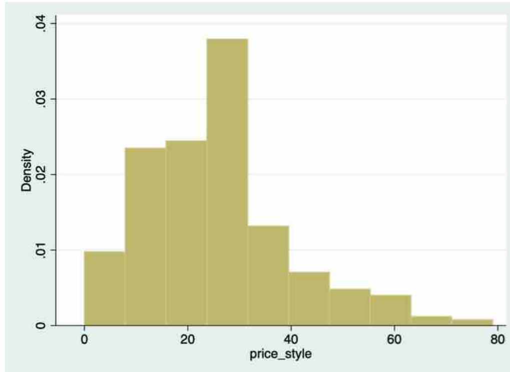


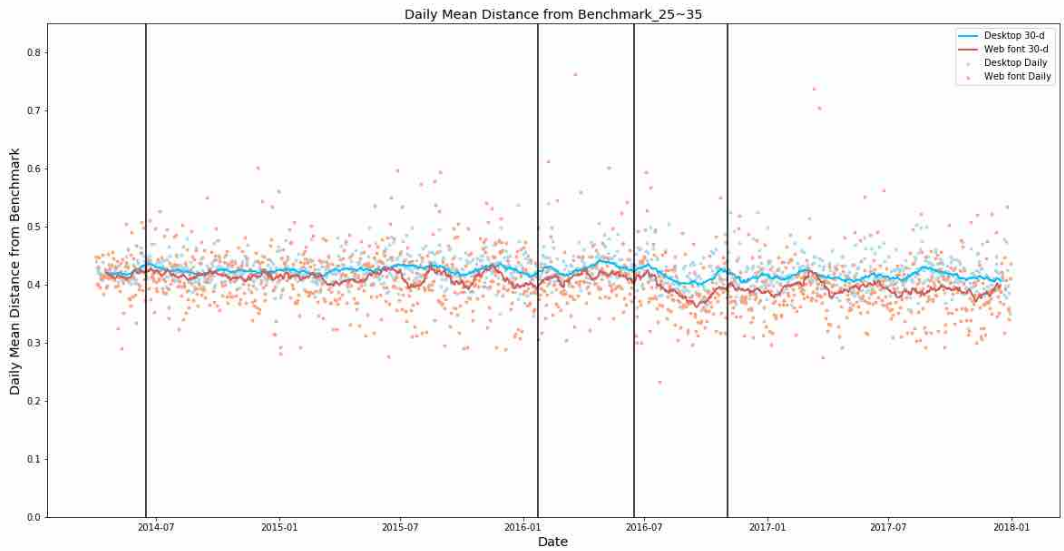
Figure 9: Price Distribution

Figure 10 plot the trend conditional on the price being \$25-35 and \$35-45. We also plot the trend for the two major license types: desktop license and web font license. Again, each dot represents the daily mean distance weighted by the number of purchases. We also plot monthly moving averages.

Interestingly, in both figures, we find that fonts that are sold as desktop license is more experimental (or less conservative) than fonts that are sold as web font license. Since both licenses are offered for most fonts, this difference cannot be attributed to the supply-side decision but are the result of consumer decisions.

5.3 Summary of Findings

Based on our analysis we find the following stylized facts: (i) Entrants are more innovative than incumbents; (ii) Demand is elastic to promotions (e.g., price-elastic). Although the timing of promotions is harder to predict and outside the scope of this analysis, we find expensive fonts are the ones that are usually promoted. Also, fonts that are promoted differ wildly in style compared to the average font. (iii) Consumers have different preferences over shapes depending on the license type they purchase. Consumers prefer more conservative shapes for web licenses and more experimental shapes for desktop license. Usually, web licenses are used in webpages, where legibility is important, while desktop license are used in printed materials (e.g., posters, cards), where designers (as consumers) have more control over the design environment.



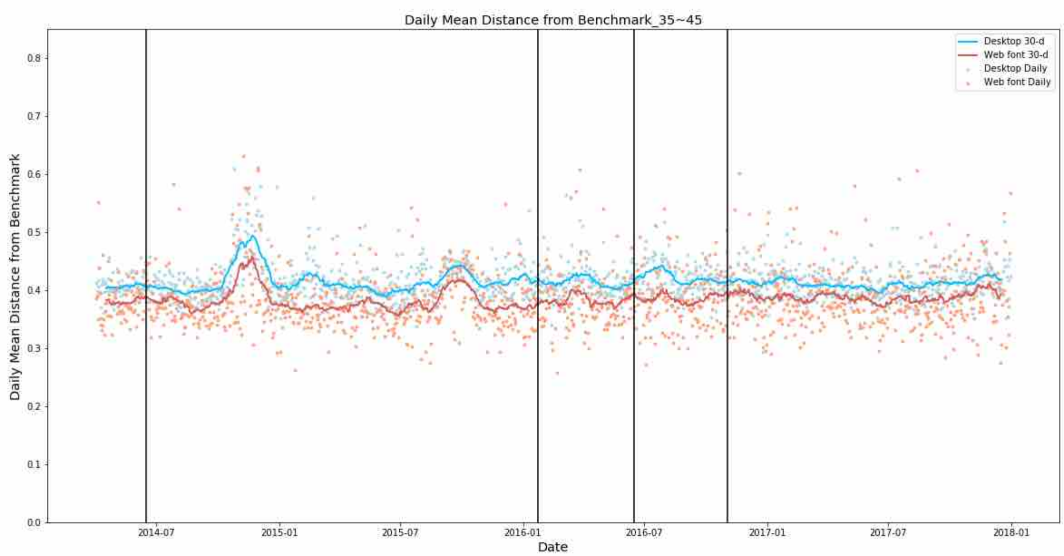


Figure 10: Trends in (Weighted) Mean Distance of Purchased Fonts (for \$25-35 and \$35-45, by license types)

6 Effects of Merger on Creative Product Differentiation

In June 15th, 2014, one of the major font foundries called FontFont is acquired by Monotype, which has been selling fonts from foundries it owns as well as third-party foundries. Using the synthetic control method as our empirical strategy, we study the causal effect of this merger on the change in the product differentiation decisions of the merging firm, i.e., FontFont. In this market, the major channel for product differentiation is the design of font shapes. In the presence of other competing foundries, this creative decision of a given foundry may be affected by many factors. The merging foundry and the platform owner may reduce cost and increase efficiency by the merger, in which case they can afford the creation of more experimental products. They may concern cannibalization, i.e., competition among their own products, in which case they again tend to increase product diversity. On the other hand, if Monotype has a preemptive motive with the merger, it will crowd its products to prevent entries. We empirically answer this question by using the embeddings we created as the main ingredient in our synthetic control method.

In this analysis, the outcome of interest is the mean distance from Averia of all new fonts created by a given foundry in a given period (i.e., $E[D_i|\text{foundry } k \text{ at } t]$ where D_i is defined in the previous section). This variable succinctly measures each foundry's creative decision of product differentiation. The product space allows us to create alternative measures for product differentiation and variety. In addition to D_i, we consider the following "gravity" measure: For font x_i,

$$\tilde{D}_i \equiv -\sum_{j \neq i} \frac{1}{\|f(x_i) - f(x_j)\|_2},$$

where the sum is for all other fonts j in the market. This measure effectively captures how font i is located relative to other fonts j's in the product space; it takes a large value when i is individually far from j's and a small value when it is close to any of them. We consider both $E[D_i|\text{foundry } k \text{ at } t]$ and $E[\tilde{D}_i|\text{foundry } k \text{ at } t]$ as the outcome variable, referred to as the distance and gravity measures, respectively.

Even before the merger, FontFont has sold its fonts in MyFonts.com as a third party. We want to measure how the mean distance has changed by the merger. The crucial step in order to recover the effect of merger that is causal, we need to have a control group that shares similarity with the treatment group, i.e., FontFont, if it were not for the merger. The challenge here is that there is only a single treated unit in the treatment group and it is difficult to find a single untreated unit that matches the treated unit. A naive average

	Mean	S.D.	Min	Max
Distance	0.41	0.11	0.18	0.74
Gravity	-9.68	0.16	-9.97	-9.24
Glyph Count	419.03	775.58	32	9,844
Release Frequency	6.25	7.02	0.00	30.00
Sales ($1K)	601	1,886	0	24,636
Order Count	7,517	23,239	1	360,435
Price per Order ($)	92.24	94.62	1.60	678.07
Age (Half Year)	10.32	5.34	0.00	17.50

Table 3: Summary Statistics (embeddings variables omitted)

of a control group would also be a poor comparison unit. The synthetic control method developed by Abadie and Gardeazabal (2003) and Abadie et al. (2010) is a useful technique to overcome this challenge. The method essentially makes comparison of the treated unit with a "synthesized treated unit" or a synthetic control unit from a weighted average of the control group. Here, the weights are estimated by minimizing the distance between the observed characteristics (including the outcome) of the treated unit and the weighted average of the characteristics of the control group.

The advantage of our setting is that the embeddings contain rich information about the design attributes of fonts that foundries created, from which the outcome variable is constructed. This information turns out to be valuable in obtaining a comparable synthetic control unit. For the control group that serves as the basis for constructing a synthetic control, we use foundries whose merger status have not changed during the period we consider.

The panel data has bi-annual time series between 2002 and 2018 and cross-sectional units are foundries. Table 3 shows the summary statistics for the variables (except the raw embeddings) we use in the analysis. Glyph count is considered to be a good measure of the quality of a font. The release frequency (i.e., maximum period between two releases), sales, number of orders, average price, and age of foundries introduced in that half year are other control variables we use. In the data, we have one treated unit (FontFont) and 76 control units.

Figure 11 captures the main result of our causal analysis using the gravity measure. The results with the distance measure are contained in the Appendix. The solid line in the left panel plots the trend of the mean gravity measure of fonts newly designed by FontFont in a given period. Around the period where FontFont was merged, which is indicated as a vertical line, the shape of FontFont's fonts substantially differs from that of other fonts in the market. This before and after comparison alone cannot yield the causal effect of the merger, since

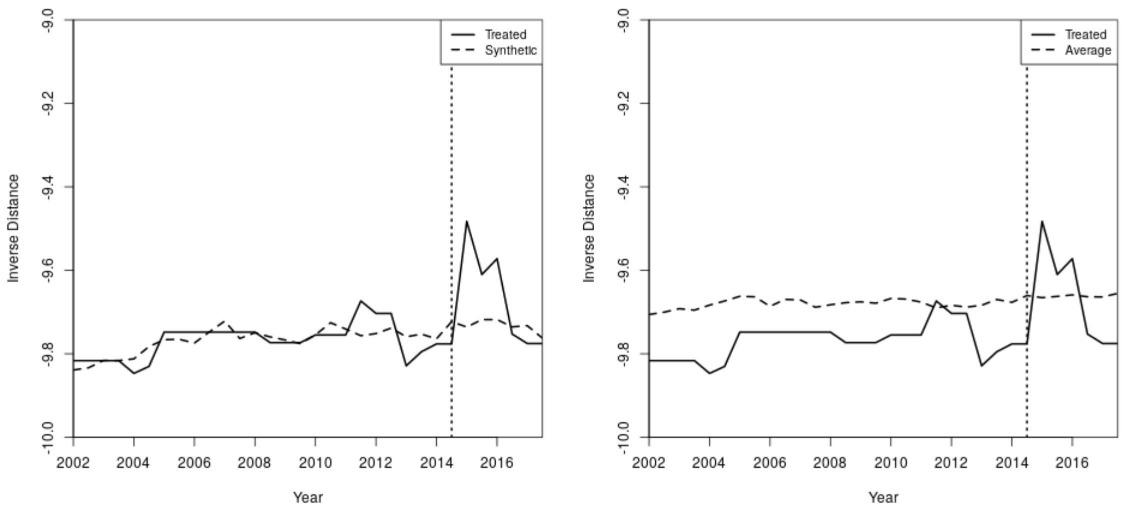


Figure 11: Trends of FontFont vs. Synthetic FontFont (left) and Trends of FontFont vs. Naive Control Group (right)—Vertical Line: Time of Merger

Years (After Merger)	2015	2016	2017
Treatment Effect	0.107	0.058	-0.019
p-Value (block)	0.037	0.074	1
p-Value (i.i.d.)	0.002	0.052	0.998

Table 4: Treatment Effects Averages (by year after the merger)

other market conditions may have changed around this period. Comparing this trend with the trend of the synthetic FontFont, depicted in a dashed line, removes possible confounding factors. First, the two trends before the treatment appear to be close to each other by construction. After the merger, however, FontFont has produced more experimental fonts (i.e., fonts that are far from others) after the merger, relative to the trend of the synthetic control.

To understand the virtue of our synthetic control method, we contrast the left panel with the right panel in Figure 11. The latter depicts the trends of the treated unit and the naive average of the control group. Inspecting the pre-treatment period in the right panel, the naive average fails to mimic the trend of the treated.

Finally, Table 4 reports the treatment effects averaged over each year after the merger. We also report their p-values using the permutation test by Chernozhukov et al. (2019). Consistent to Figure 11 (left), the short-run effects in the first and second years are significant whereas the effect in the third years is not. Also, the placebo test results, which are contained in Table 5 in the Appendix, show that the treatment effects are not significant before the merger.

In the Appendix, we produce analogous results using the distance measure instead of the gravity measure, and find that they are qualitative similar. This suggests that our findings are robust to the choice of the measure of product differentiation.

From this analysis, we conclude that the merger causes FontFont to explore a new territory of the product space at least temporarily. That is, now as part of the parent organization, Monotype, FontFont increases the visual variety in font design. Before the merger, Monotype owned multiple foundries that produce fonts and sold them in its platform, MyFonts.com. After the merger, Monotype may have incentives to diversify the product scope as the size and efficiency of the firm has increased. It may also be the case that Monotype tries to avoid cannibalization by spread apart its products, reducing competition of their own. Such tendency disappears after two years, perhaps because the company has either successfully foreclosed the market or found the strategy unprofitable. In the meantime, the synthetic FontFont tries to be more adventurous in the font design about two years after Monotype departs from conservative fonts. Since fonts with the same distance from Averia are not necessarily close to each other, this behavior of the synthetic FontFont can be viewed as either the increased competition with Monotype or a behavior to occupy other regions of the product space avoiding the competition.

7 Conclusion

Certain policy questions are better answered by treating high-dimensional, unstructured attributes as observables and attempting to solve the resulting dimensionality problem. As a result, we propose to quantify the design oriented attributes of a product by constructing a low-dimensional product space of these attributes. We use a modern convolutional neural network to accomplish this task. We find that these attributes are correlated with font designers and consumers perceptions by comparing the mutual information tags used to describe the fonts with the neural network embeddings. We conduct two economic analyses. We first analyze supply side and demand-side trends and find they are related to promotion policies and license types. We then turn to a causal analysis to understand the effects of merger on product differentiation. We find that the merger increases the product variety in this market.

This paper motivates interesting directions for future research, some of which may require a structural approach. For structural economic analyses, we envision a two-step approach of using the embeddings: first construct a neural network embedding to reduce the dimension of the product image, and then include the embeddings in structural economic models.

Although we could use the neural network to directly predict economic variables, such as demand and price, the neural network is not a causal model. As a result, its counterfactual predictions are not as credible as a more traditional structural model. An alternative structural approach would be to incorporate dimension reduction as part of the structural estimation (Chernozhukov et al. (2018); Foster and Syrgkanis (2019)).

Using these structural approaches, we can continue answering economic questions related to this market. Related to the economic analysis we conducted in this paper, we may view product differentiation as spatial competition. There is a close parallel between spatial differentiation and the neural network embedding product space we constructed. Analogous to Seim (2006) who studies the location choice of video rental stores, we can investigate the relationship between design choices and local market power of the designers. We can also study how third-party and in-house foundries differ in their product differentiation decisions. One relevant policy question is the effect of the commission fee of third parties. In fact, Monotype has changed the commission policy during the data period, which may serve as key policy variation.

Alternatively, we may view product differentiation as intellectual property. Agents in this industry are subject to license agreements, which aim to protect the originality of font shapes. Heuristically this policy says that "one cannot produce fonts which shapes are substantially similar to existing fonts." Given the product space we characterized, one can interpret this policy as imposing a ball centered around each font, preventing other productions within—producing another font inside is considered as violation. Then, one can ask what the welfare maximizing level of the policy (i.e., the optimal radius of the ball) would be and whether the current level in the market is optimal or suboptimal.

A Other Treatment Effects and Placebo Tests

A.1 Gravity Measure

	2012/1	2012/2	2013/1	2013/2	2014/1
Treatment Effects	0.081	0.070	-0.012	-0.085	-0.069
p-Value (block)	0.381	0.455	0.087	0.250	0.760
p-Value(i.i.d.)	0.369	0.467	0.084	0.248	0.756

Table 5: Placebo Test: Treatment Effects Before Merger (Using Gravity Measure)

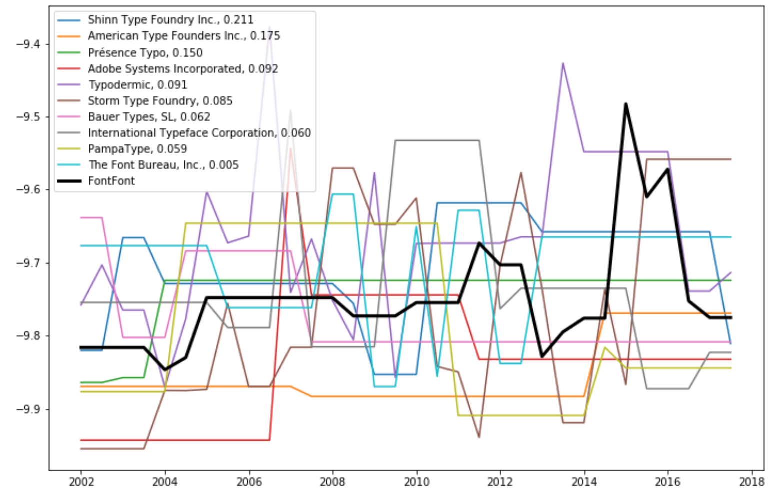


Figure 12: Trends of FontFont vs. Top 10 Control Units (Using Gravity Measure)

	2014/2	2015/1	2015/2	2016/1	2016/2	2017/1	2017/2
Treatment Effects	-0.055	0.087	0.060	0.072	-0.012	-0.028	-0.017
p-Value (block)	0.462	0.039	0.039	0.039	0.846	0.962	0.885
p-Value (i.i.d.)	0.452	0.040	0.039	0.041	0.845	0.957	0.883

Table 6: Treatment Effects After Merger (Using Gravity Measure)

A.2 Distance Measure

	2012/1	2012/2	2013/1	2013/2	2014/1
Treatment Effects	0.044	0.020	-0.025	-0.028	0.002
p-Value (block)	0.762	0.818	0.087	1	0.840
p-Value(i.i.d.)	0.763	0.825	0.086	1	0.835

Table 7: Placebo Test: Treatment Effects Before Merger (Using Distance Measure)

	2014/2	2015/1	2015/2	2016/1	2016/2	2017/1	2017/2
Treatment Effects	-0.007	0.154	0.079	0.101	0.000	-0.011	-0.008
p-Value (block)	0.539	0.039	0.039	0.039	0.885	0.346	0.500
p-Value (i.i.d.)	0.535	0.038	0.038	0.041	0.878	0.347	0.496

Table 8: Treatment Effects After Merger (Using Distance Measure)

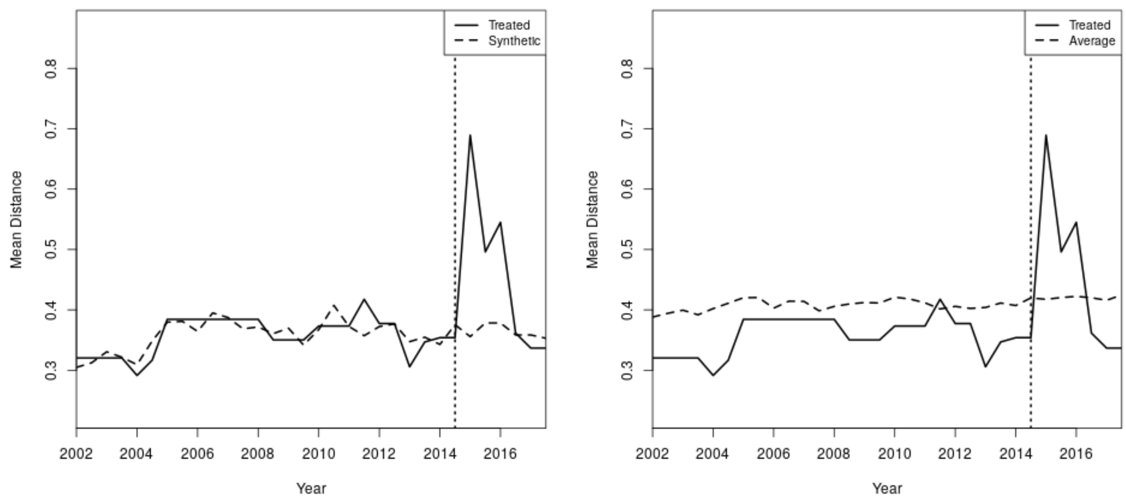


Figure 13: Trends of FontFont vs. Synthetic FontFont (left) and Trends of FontFont vs. Naive Control Group (right)—Vertical Line: Time of Merger (Using Distance Measure)

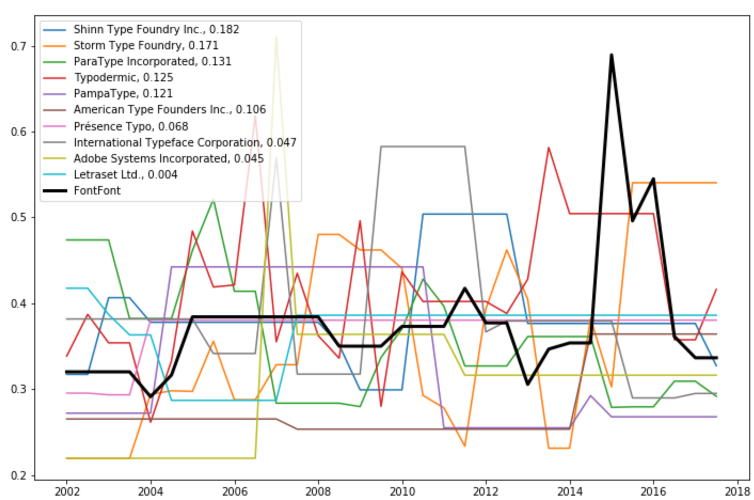


Figure 14: Trends of FontFont vs. Top 10 Control Units (Using Distance Measure)

References

Abadie, A., A. Diamond, and J. Hainmueller (2010): "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program," *Journal of the American statistical Association*, 105, 493–505. 1.5, 6

Abadie, A. and J. Gardeazabal (2003): "The economic costs of conflict: A case study of the Basque Country," *American economic review*, 93, 113–132. 1.5, 6

Al-Halah, Z., R. Stiefelhagen, and K. Grauman (2017): "Fashion forward: Forecasting visual style in fashion," in *Proceedings of the IEEE International Conference on Computer Vision*, 388–397. 1.6

Angrist, J. D. and J.-S. Pischke (2010): "The credibility revolution in empirical economics: How better research design is taking the con out of econometrics," *Journal of economic perspectives*, 24, 3–30. 4

Ashenfelter, O. and D. Hosken (2008): "The effect of mergers on consumer prices: Evidence from five selected case studies," Tech. rep., National Bureau of Economic Research. 1.5, 1.6

Bajari, P. and C. L. Benkard (2005): "Demand estimation with heterogeneous consumers and unobserved product characteristics: A hedonic approach," *Journal of political economy*, 113, 1239–1276. 1.2

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile prices in market equilibrium," *Econometrica: Journal of the Econometric Society*, 841–890. 1.2

Berry, S. T. and J. Waldfogel (2001): "Do Mergers Increase Product Variety? Evidence from Radio Broadcasting," *The Quarterly Journal of Economics*, 116, 1009–1025. 1.6

Bottou, L. (2010): "Large-scale machine learning with stochastic gradient descent," in *Proceedings of COMPSTAT'2010*, Springer, 177–186. 4.2

Burnap, A., Y. Liu, Y. Pan, H. Lee, R. Gonzalez, and P. Y. Papalambros (2016): "Estimating and exploring the product form design space using deep generative models," in *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, American Society of Mechanical Engineers, vol. 50107, V02AT03A013. 1.6

Campbell, N. D. and J. Kautz (2014): "Learning a manifold of fonts," *ACM Transactions on Graphics (TOG)*, 33, 91. 1.6

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018): "Double/debiased machine learning for treatment and structural parameters," . 1.6, 7

Chernozhukov, V., J. C. Escanciano, H. Ichimura, and W. K. Newey (2016): "Locally Robust Semiparametric Estimation," *arXiv preprint arXiv:1608.00033*. 1.6

Chernozhukov, V., K. Wuthrich, and Y. Zhu (2019): "An exact and robust conformal inference method for counterfactual and synthetic controls," *arXiv preprint arXiv:1712.09089*. 6

Dosovitskiy, A., J. T. Springenberg, M. Tatarchenko, and T. Brox (2016): "Learning to generate chairs, tables and cars with convolutional networks," *IEEE transactions on pattern analysis and machine intelligence*, 39, 692–705. 1.6

Fan, Y. (2013): "Ownership Consolidation and Product Characteristics: A Study of the US Daily Newspaper Market," *American Economic Review*, 103, 1598–1628. 1.6, 5

Foster, D. J. and V. Syrgkanis (2019): "Orthogonal statistical learning," *arXiv preprint arXiv:1901.09036*. 1.6, 7

Friedman, J., T. Hastie, and R. Tibshirani (2001): *The elements of statistical learning*, vol. 1, Springer series in statistics New York. 1.4

Gentzkow, M., B. Kelly, and M. Taddy (2019a): "Text as Data," *Journal of Economic Literature*, 57, 535–74. 1.6

Gentzkow, M., J. M. Shapiro, and M. Taddy (2019b): "Measuring group differences in high-dimensional choices: method and application to congressional speech," *Econometrica*, 87, 1307–1340. 1.6

Gillen, B., S. Montero, H. R. Moon, and M. Shum (2015): "Blp-lasso for aggregate discrete choice models of elections with rich demographic covariates," *USC-INET Research Paper*. 1.6

Glaeser, E. L., S. D. Kominers, M. Luca, and N. Naik (2018): "Big data and big cities: The promises and limitations of improved measures of urban life," *Economic Inquiry*, 56, 114–137. 1.6

Goodfellow, I., Y. Bengio, and A. Courville (2016): *Deep Learning*, MIT Press, http://www.deeplearningbook.org. 1.4

Gross, D. P. (2016): "Creativity under fire: The effects of competition on creative production," *Review of Economics and Statistics*, 1–17. 1.6

Hastings, J. S. (2004): "Vertical relationships and competition in retail gasoline markets: Empirical evidence from contract changes in Southern California," *American Economic Review*, 94, 317–328. 1.5, 1.6

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*, 770–778. 1.6

Hoberg, G. and G. Phillips (2016): "Text Based Network Industries and Endogenous Product Differentiation," *Journal of Political Economy*, 124, 1423–1465. 1.6

Kovashka, A., D. Parikh, and K. Grauman (2012): "Whittlesearch: Image search with relative attribute feedback," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, 2973–2980. 1.6

Kozlowski, A. C., M. Taddy, and J. A. Evans (2019): "The Geometry of Culture Analyzing the Meanings of Class through Word Embeddings," *American Sociological Review*, 84, 905–949. 1.6

Krizhevsky, A., I. Sutskever, and G. E. Hinton (2012): "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 1097–1105. 1.6

LeCun, Y., C. Cortes, and C. Burges (2010): "MNIST handwritten digit database," . 1.6

Mall, U., K. Matzen, B. Hariharan, N. Snavely, and K. Bala (2019): "Geostyle: Discovering fashion trends and events," in *Proceedings of the IEEE International Conference on Computer Vision*, 411–420. 1.6

Mazzeo, M. J., K. Seim, and M. Varela (2018): "The Welfare Consequences of Mergers with Endogenous Product Choice," *The Journal of Industrial Economics*, 66, 980–1016. 1.6

McFadden, D. (1973): *Conditional logit analysis of qualitative choice behavior*, Institute of Urban and Regional Development, University of California Berkeley. 1.2

Mikolov, T., K. Chen, G. Corrado, and J. Dean (2013): "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*. 4.5

Nevo, A. and M. D. Whinston (2010): "Taking the dogma out of econometrics: Structural modeling and credible inference," *Journal of Economic Perspectives*, 24, 69–82. 4

O'Donovan, P., J. Lībeks, A. Agarwala, and A. Hertzmann (2014): "Exploratory font selection using crowdsourced attributes," *ACM Transactions on Graphics (TOG)*, 33, 1–9. 1.6

Parikh, D. and K. Grauman (2011): "Relative attributes," in *2011 International Conference on Computer Vision*, IEEE, 503–510. 1.6

Rosen, S. (1974): "Hedonic prices and implicit markets: product differentiation in pure competition," *Journal of political economy*, 82, 34–55. 1.2

Schroff, F., D. Kalenichenko, and J. Philbin (2015): "FaceNet: A unified embedding for face recognition and clustering," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 815–823. 1.4, 1.6, 4.1, 4.2

Seim, K. (2006): "An Empirical Model of Firm Entry with Endogenous Product-Type Choices," *The RAND Journal of Economics*, 37, 619–640. 7

Simonyan, K. and A. Zisserman (2014): "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*. 1.6

Sun, Y., X. Wang, and X. Tang (2015): "Deeply learned face representations are sparse, selective, and robust," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2892–2900. 1.6, 4.1

Sweeting, A. (2013): "Dynamic Product Positioning in Differentiated Product Markets: The Effect of Fees for Musical Performance Rights on the Commercial Radio Industry," *Econometrica*, 81, 1763–1803. 1.6, 5

Taigman, Y., M. Yang, M. Ranzato, and L. Wolf (2014): "Deepface: Closing the gap to human-level performance in face verification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1701–1708. 1.6, 4.1

Wang, J., T. Song, Yangand Leung, C. Rosenberg, J. Wang, Jingbinand Philbin, B. Chen, and Y. Wu (2014): "Learning Fine-grained Image Similarity with Deep Ranking," *arXiv preprint arXiv:1404.4661*. 1.4, 1.6

Weinberger, K. Q., J. Blitzer, and L. K. Saul (2006): "Distance metric learning for large margin nearest neighbor classification," in *In NIPS*, MIT Press. 4.2

Wilson, D. R. and T. R. Martinez (2003): "The general inefficiency of batch training for gradient descent learning," *Neural networks*, 16, 1429–1451. 4.2

Yu, A. and K. Grauman (2019): "Thinking outside the pool: Active training image creation for relative attributes," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 708–718. 1.6