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 18, 2021

Abstract

Many differentiated products have key attributes that are *unstructured* and thus high-dimensional (e.g., design, text). Instead of treating unstructured attributes as unobservables in economic models, quantifying them can be important to answer interesting economic questions. To propose an analytical framework for this type of products, this paper considers one of the simplest design products—fonts—and investigates merger and product differentiation using an original dataset from the world's largest online marketplace for fonts. We quantify font shapes by constructing embeddings from a deep convolutional neural network. Each embedding maps a font's shape onto a low-dimensional vector. In the resulting product space, designers are assumed to engage in Hotelling-type spatial competition. From the image embeddings, we construct two alternative measures that capture the degree of design differentiation. We

*For helpful discussions, the authors are grateful to Jorge Balat, Aureo de Paula, Phil Haile, Greg Howard, Alessandro Iaria, Roger Moon, Jesse Shapiro, Matt Shum, Haeyeon Yoon, participants in the 2021 North American Winter Meeting, the 2020 European Winter Meeting, 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

then study the causal effects of a merger on the merging firm's creative decisions using the constructed measures in a synthetic control method. We find that the merger causes the merging firm to increase the visual variety of font design. Notably, such effects are not captured when using traditional measures for product offerings (e.g., specifications and the number of products) constructed from structured data.

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

1 Introduction

Many differentiated products considered in economic analyses have important attributes that are *unstructured*. Examples include design elements in products such as automobiles, houses, furniture, and clothing or digital products such as mobile applications. Other obvious examples include creative features in books, music, movies, and fine arts. Unstructured attributes in these products are typically in visual or textual forms and thus are high-dimensional. More generally, products well beyond these categories are often presented to consumers in visual and textual forms: for example, product packages in supermarkets and online catalogs in e-commerce (e.g., Amazon, Airbnb, Yelp, Zillow). These 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, unstructured attributes are important decision factors for consumers and thus are key decision variables for producers.

Economists are aware of the role of unstructured attributes. Product attributes are an important component 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 variable normally captures the high-dimensional, unstructured attributes, including design and other original features of the products.

Although this tradition has its own merits, certain economic questions are better answered by treating unstructured attributes as observables. For example, one may ask how the style of products evolves over time—that is, how the fashion changes—in accordance with market conditions. One may further ask how product differentiation in this creative dimension affects the product's market power or is affected by market shocks such as vertical integration. These questions entail the issues of measurability and dimensionality of unstructured attributes. Due to these challenges, there has been little understanding about the production of creative attributes—namely, product differentiation decisions—in the realm of quantitative economics.[1]

In this paper, we propose a framework to quantify the design-oriented attributes of a product and construct a low-dimensional space of products and measures for the degree of product differentiation using the Euclidean distance endowed in the space. We then illustrate how some of the economic questions above can be answered using this framework, such as market-driven product differentiation decisions of product designers in multi-product firms.

For the purpose of this paper, we consider a particular design product: fonts. We use a dataset obtained from the world's largest online marketplace for Roman alphabet 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.[2] Second, this visual information is simple to understand but important in predicting the functionality and value of the product. Third, fonts are ubiquitous products; thus, the market for fonts is large with frequent productions and transactions. The online marketplace we consider has over 28,000 fonts (produced by font design firms called foundries) and 2,400,000 transactions over the past six years. Fourth, interesting policies are involved in this market, such as vertical integration. Finally and most importantly, font is a stylized product that saliently captures a key aspect many products in the market have in common: design attributes.

The main challenge in quantitatively 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. In particular, we adapt a state-of-the-art method in convolutional neural networks (Wang et al. (2014), Schroff et al. (2015)), where the network directly learns to map font images to a compact Euclidean space—that is, the embedding space—in which perceived visual similarity is preserved. The algorithm is sophisticated enough to recognize the *style* of font shapes, which is crucial for our purpose.

Another challenge is to ensure that the resulting embeddings represent economic agents' perceptions of the product's visual information. To this end, we employ two strategies. First, we use the images of entire *pangrams* (instead of individual alphabet letters) as inputs in the neural network.[3] Because pangrams effectively capture important design elements that cannot be seen in individual letters (e.g., spacing, deep-height, up-height, and ligature), they are the most relevant decision variables for font designers and consumers. Second, we demon-

[1] Galenson and Weinberg (2000, 2001) study artists and their career choices and paths in fine arts. Their main approach to quantitative analyses is to use price as a proxy to measure the value of artworks.

[2] These shapes are called typefaces.

[3] A pangram is a sentence that contains all the alphabet letters.

strate that the obtained image embeddings contains a substantial amount of information that is mutually shared with *tags*, which are word phrases that describe fonts (e.g., "curly," "flowing," "geometric," "organic"). Tags are assigned to each font by font designers and consumers and thus, we believe, represent the economic agents' perceptions of the product. We construct word embeddings from these tags using a simple neural network and calculate the mutual information between the word and image embeddings.

Why do we consider a convolutional neural network? Font shapes involve a non-linear interaction between many neighboring pixels. Considering individual pixels separately or recognizing interactions in a restrictive model provides little information about the overall shape of a font. By considering how neighboring pixels interact in a very flexible model, 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). In particular, the deep convolutional neural network is designed to effectively capture the spatial correlation between nearby pixels. Although neural networks are generally known to be less interpretable (Friedman et al. (2001)) than other learning methods due to model flexibility,[4] we show how an interpretable embedding space can be learned through visual similarity. Most importantly, instead of attempting to interpret each embedding value, our approach is to give meanings to the distance metric of the embedding space and subsequently construct the product differentiation measures based on it.

Given the embedding space, a font designer's decision of a typeface design is equivalent to choosing a location in the space. This location choice is a strategic decision that depends on the choices of other designers in the space. In this sense, the abstract product space we construct can be viewed as a location-analog model (Hotelling (1929)) with Lancasterian characteristics (Lancaster (1966, 1971)). Based on this space, we construct two alternative differentiation measures using the image embeddings, namely a distance to Averia (i.e., the average font) and a gravity measure, which succinctly represent the location choice of a designer relative to others' choices.

To illustrate the usefulness of our approach, we conduct a causal analysis of how a merger affected the merged firm's design decisions before and after the merger. In June 2014, a major font foundry was acquired by the company that owns the online marketplace. Before the merger, the merging firm was selling fonts as a third party and competing with the foundries owned by the marketplace. The main motivation of the analysis is that designers are not only artists but also economic agents who are affected by market conditions. We use the constructed design differentiation measures as the main dependent variables. To rule out

[4] 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.

other factors confounded with the merger, we use the synthetic control method (Abadie and Gardeazabal (2003), Abadie et al. (2010)). This method is suitable in our context because there is only one merged foundry, for which finding a good comparison group from untreated foundries is challenging. The advantage of our setting in using the synthetic control method is that the embeddings we obtained serve as the main predictors in constructing a comparable synthetic control.

Our main finding is that, relative to the synthetic control unit, the merged foundry produced fonts with greater visual variety after the merger and that this effect was statistically significant. One of the explanations is that the degree of product differentiation may have increased after the merger to avoid cannibalization. Notably, we find that such effects are not captured when traditional measures for product offerings are used from structured data (i.e., the number of products and specifications; Berry and Waldfogel (2001)). This illustrates the importance of more sophisticated product offerings measures as employed in the current paper.

1.1 Contributions and Related Literature

Machine Learning and Social Science Research

To our knowledge, this is one of the first few papers using neural network embeddings for visual data in the economic analysis of markets and industries. Earlier work that analyzes visual data uses methods that are partly or fully human-aided. Glaeser et al. (2018) use visual data from Google Street View to predict the economic prosperity of neighborhoods, but they assign scores to street images based on human surveys on the visual perception of street quality and safety. Gross (2016) investigates how competition influences creative production in a commercial logo design competition. Whereas 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)). Deep learning methods are used in more recent studies. Zhang et al. (2017) show how the quality and specific attributes of property images on Airbnb can affect the demand. The quality and attributes are human-labeled, and a convolutional neural network is used to train the images based on the labels. Our approach does not require subjective human labeling. As independent and contemporaneous work, Bajari et al. (2021) estimate a hedonic function for apparel consumption using a deep neural network based on visual and textual data. We also utilize textual data but in such a way that gives economic meanings to the image embeddings we obtain. The most significant difference between our approach and the two aforementioned studies is that we consider images as the main response variables of market shocks and construct differentiation measures based on neural network embeddings. Using these measures also helps gain interpretability and robustness in the quantities resulting from the network training. In addition, we use a different neural network algorithm based on *image triplets* that is suitable to our setting.

This paper is also among the first social science studies that make use of 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) use word embeddings to understand cultural norms. Gentzkow et al. (2019b) analyze 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. Unlike their paper, however, we use image embeddings and employ neural networks as a classification method.[5] Moreover, 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) find that the effects of a merger on product differentiation can be ambiguous, implying that the question is more of empirical research, as in empirical industrial organization. Sweeting (2013) studies the dynamic aspects of product differentiation in the radio industry and finds some evidence suggesting that increased concentration increases variety. Fan (2013) finds in newspaper markets that mergers between local competitors has effects on vertical differentiation, leading to a decrease in the news quality. Besides these two papers, the literature using a structural approach to study the effects of merger on non-price attributes (let alone unstructured attributes considered in this paper) is scarce due to modeling difficulties. This is in contrast to merger effects on price, which have received relatively more attention in the literature, e.g., building on the structural framework of Nevo (2001).

The structural approach has been complemented in the literature by studies of merger and concentration from a viewpoint of treatment effects and program evaluation. Berry and Waldfogel (2001) document the effects of mergers on product variety in local radio markets by exploiting a natural experiment. Hastings (2004) and Ashenfelter and Hosken (2008)

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

study the effects of mergers on prices using the difference-in-differences and instrumental variable methods, respectively. Although the treatment effect approach to merger analyses has limitations (Nevo and Whinston (2010)), it is still suitable to highlight the rich information contained in the visual dimension of product attributes previously neglected in the literature on mergers.[6] We introduce unstructured data of images and related machine learning methods to derive new insights into creative product differentiation and its relationship to mergers. On the other hand, all the empirical studies listed in this and the previous paragraphs use structured data to construct response variables including price and non-price attributes (e.g., product variety measures).[7] As mentioned, we cannot find in our analysis the merger effects on the traditional product offerings measures. We believe this paper is a good starting point for introducing embeddings into economic analyses.

Machine Vision

Recognizing letters (e.g., distinguishing handwritten "G" from "Q") is one of the most well-studied areas of machine vision as is done 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. Tenenbaum and Freeman (2000) propose bilinear models that separate style and content with fonts as one of the examples. O'Donovan et al. (2014) develop a method for searching fonts using relative attributes based 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 for learning a font manifold by parametrizing font shapes and reducing the dimension of the resulting model.

The method of training the font embeddings builds on the works of Schroff et al. (2015) and Wang et al. (2014), who develop a face recognition algorithm that directly learns an embedding for images via 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, such as those in Taigman et al. (2014) and Sun et al. (2015). The former approach is suitable for our purpose, as the procedure produces embeddings as the intermediate output of the classification algorithm. Although we are not directly interested

[6]We further discuss the structural and treatment effect approaches in our specific context in Section 5. See Angrist and Pischke (2010) and Nevo and Whinston (2010) for discussions on how the two approaches can complement each other and what their pros and cons are.

[7]For example, Fan (2013) uses data on the number of opinion section staff members, the number of reporters, local news ratio, variety, frequency of publication, and edition. Sweeting (2013) uses Neilson data on broadcasts. Berry and Waldfogel (2001) use the number of stations and programming formats as product offerings.

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 work by, 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 the 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. However, none of the studies conduct causal analyses to answer economic questions.

1.2 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 embeddings are meaningful by calculating the mutually shared information between the font embeddings and tags. Section 5 presents the causal analysis of a merger using the embeddings. Section 6 concludes. Sections A.2–C in the Appendix respectively present the internal evaluation of the trained network, the descriptive analyses of style trends, and supplemental results from the merger analysis.

2 Online Marketplace for Fonts

We consider the world's largest online marketplace MyFonts.com that sells around 30,000 different fonts. This market is a superset of all major global online stores for fonts. MyFonts.com and the other stores are all owned by Monotype Inc. A font is a delivery mechanism for typefaces. Therefore, fonts are sold as a piece of software, for which consumers purchase a license. Licenses are protected by the End User License Agreement (EULA). Mostly two types of licenses are sold to consumers. A web font license allows fonts to be displayed on a website, and a desktop license is for printed material. 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.[8]

Figure 1 shows the home page of MyFonts.com. An example of a font family page on MyFonts.com is captured in Figure 2. In the font industry, a *family* represents the identity of a font, which name is the name of the font. A font family is consist of several different *styles*,

[8]A foundry is a group of designers who create fonts.

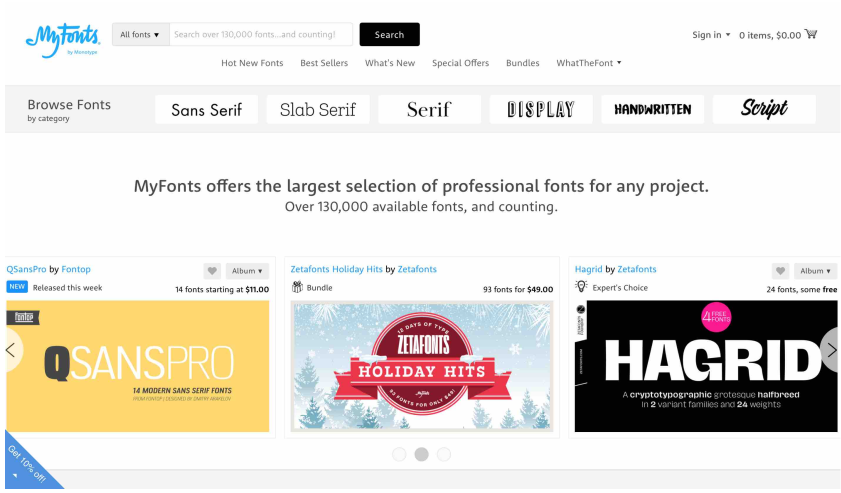


Figure 1: Home Page of MyFonts.com

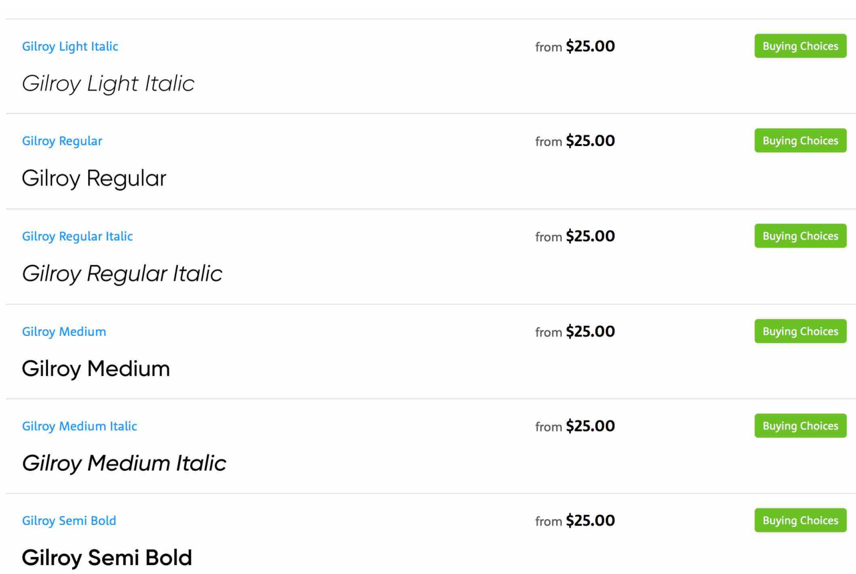


Figure 2: A Font Family Page on MyFonts.com

such as regular, light, bold, and italic. Figure 2 shows different styles of Gilroy family. In this market, typical consumers are independent designers who use fonts as intermediate goods. They produce printed material (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. In the data collection period of 2012–2018, around 2,400,000 purchases were made.

3 Data

3.1 Overview

Our sample comprises data from 2002 to 2017. The dataset includes, in total, 28,659 fonts and 2,446,604 orders. The main information contained in the dataset is product attributes and transactions for each consumer. The unstructured high-dimensional attributes include images of typefaces and tags (i.e., descriptive words assigned by producers or consumers). The structured attributes include price, category types, license types, the number of languages supported, the number of glyphs supported, the foundry and designer information, and the date of introduction in the market. There are roughly six category types: sans serif, serif, slab serif, display, handwritten, and script. Transaction data include information on individual orders made by consumers and consumer characteristics such as the country and city of origin. We will revisit this dataset in Section 5 for the merger analysis. For now, we focus on the unstructured data.

3.2 Visual Attributes

Fonts are displayed on the webpage using pangrams that contain all the alphabet letters in one sentence.[9] Pangrams effectively capture important design elements that cannot be seen in individual letters, such as spacing, deep-height, up-height, and ligature. We use pangrams as direct inputs in the neural network in order to mimic a consumer's actual perception of the products. Figure 3 shows the examples of a pangram that roughly correspond to the product categories (i.e., sans serif, serif, slab serif, display, handwritten, and script). The format of pangram images is a bitmap with 200×1000 pixels, where each pixel is a greyscale with a value between 0 and 255. We use random crops of 100×100 pixels as inputs in the

[9] In font markets, many different pangrams are used: "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." "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.

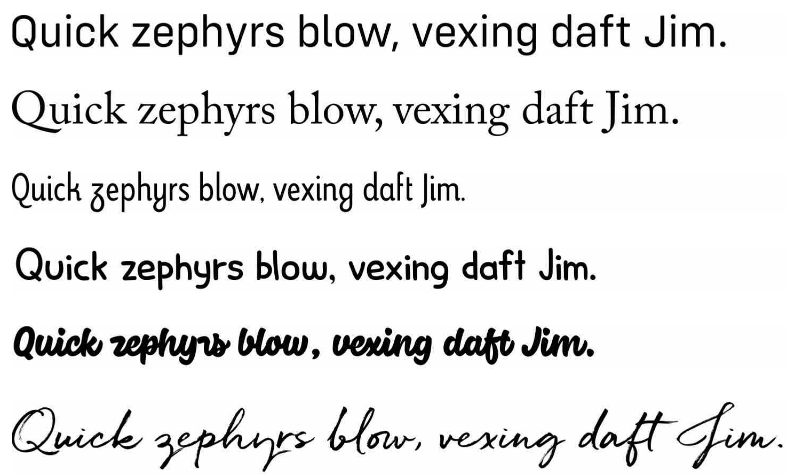


Figure 3: Examples of a Pangram (by category types)

network.

4 Construction of Embeddings

4.1 Neural Network Embedding

We employ a method in which 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 with $d = 128$.[10] The Euclidean distance in the resulting embedding space corresponds to the 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 algorithm determines the identity of a person based on face images in two steps. In the first step, a deep convolutional network is trained 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, the identities of the images are classified 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)). Our algorithm builds on Schroff et al. (2015), which approach is suitable for our purpose. First, pangram images can be classified based on a structure similar to

[10] It is important to have sufficient dimensions so that the neural network can allow for variation in the embedding space based on the actual images. The embedding with a larger number of dimensions would perform better, but it requires more training data to achieve the same level of accuracy while avoiding the risk of overfitting. We additionally normalize each embedding so that it lies on a 128-dimensional hypersphere, i.e., $\|f\|_2 = 1$ with the Euclidean norm $\|\cdot\|_2$.

that for face images. In the dataset of face images, multiple images are associated with the same identity. Analogously, multiple styles are associated with the same font family as detailed below. Second, the approach 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.

4.2 Triplet Loss and Network Training

The neural network is trained to produce embeddings and classify fonts that are within the same family in the resulting embedding space. In practice, we accomplish this by constructing triplets of images. Triplet i comprises anchor x_i^a, positive x_i^p, and negative x_i^n. An anchor is a pangram image of a given font family (e.g., Helvetica), positives are pangram images of the same family but different styles (e.g., Helvetica Regular, Helvetica Light, Helvetica Bold, Helvetica Italic), and negatives are pangram images of different families (e.g., Time New Roman).[11] The way triplets are sampled 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. For triplet (x_i^a, x_i^p, x_i^n) in the entire set of font images, we enforce the following inequality during the network training:

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

where $f(x) \in \mathbb{R}^{128}$ is the embedding of image x, $\|\cdot\|_2$ is the Euclidean norm, and α is an enforced margin. That is, we ensure that the distance between an anchor and positive is smaller than that between the anchor and a negative.[12] The margin α allows the images for one font family to stay closer in the embedding space, while still discriminating the images of other families.

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

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

We optimize this objective using stochastic gradient descent (SGD; Bottou (2010)). SGD is an iterative method for optimizing an objective function—in our case, for creating a reasonable embedding space. Because our dataset size is in gigabytes, it would be computationally

[11] To be precise, pangram images here refer to crops of the images. Therefore, we also use different crops within the same pangram as positives.

[12] The choice of triplets is very important for the fast convergence of the algorithm. As such, we make sure we choose sufficiently many triplets with hard positives and negatives, namely triplets that violate (1).

challenging to compute the gradient of the entire dataset and optimize it 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.

We use approximately 20,000 images of fonts to train the neural network. The training iteratively improves the parameters of the network using small batches of images to estimate the gradient and then update the parameters accordingly. As the gradient is evaluated at more batches, the parameters in the network are adjusted. The training of the network is completed when the loss function reaches below a certain threshold. Section A.1 in the Appendix contains the details.

As internal evaluation, we show in Section A.2 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 the embeddings perform poorly, then they are not likely to be a reliable product space. As mentioned, the classification of font families is conducted by thresholding the distances between embeddings. Overall, the neural network embeddings perform well in differentiating between fonts in different families.

4.3 Constructed Product Space

The trained neural network produces embeddings with 128 dimensions, which define a 128-dimensional space of font products. Figure 4 visualizes this space by projecting it onto a two-dimensional space using t-distributed stochastic neighbor embedding (t-SNE).[13] For expositional purposes, the thumbnail is created with the word “Quick” that is cropped from the full pangram. Each thumbnail corresponds to the embedding of the *regular style* of each font family as a representative style. Therefore, for example, boldface in this figure is a design aspect of a particular font family, *not* a bold style within a family. A visual inspection reveals that different styles of fonts are well-clustered together even in this two-dimensional projection. More specifically, we can see that the western region is populated with fonts that have narrower characters and the eastern region with fonts that are thicker. In addition, the shape of fonts in the northeast is more geometric while the shape in the southwest is more curly. It is worth noting that these patterns are observed even in this space with restricted dimensions. The space we base our economic analysis on is the original 128-dimensional space.

[13] t-SNE is a useful tool to visualize high-dimensional data in a two or three dimensional space. In our setting, we visualize 128-dimensional objects in a two-dimensional space.

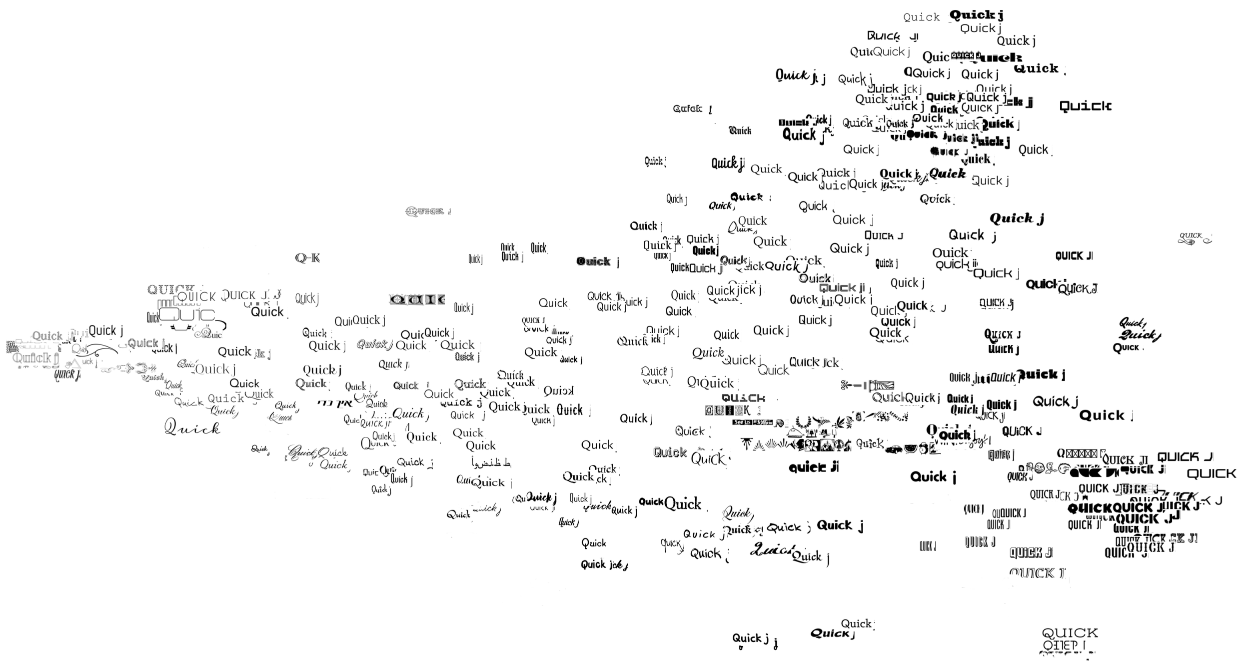


Figure 4: Two-Dimensional Visualization of the Space for Fonts

Note: This figure is created using a randomly selected subsample of 400 fonts projected onto two dimensions using t-SNE.

To further understand how well the fonts are clustered in the 128-dimensional product space, we present in Figure 5 the examples of regular fonts that are close to each other in the space. In Figure 5, the images of six nearest neighbors in terms of the Euclidean distance are listed in each row of the table. Even though we only list regular fonts, it is clear that fonts with different thicknesses are clustered together in the space. Also, the middle and the last rows show that geometric and curly fonts tend to cluster together, respectively.

Unlike individual embeddings, the distance metric endowed in this space has the clear interpretation of visual similarity and is more robust to model specifications. These aspects allow us to use the distance metric as the main building block in the economic analysis in Section 5.

4.4 Relevance of Embeddings

Although the neural network embedding performs well in terms of classifying fonts of similar shapes, good predictive performance does not necessarily imply the relevance of the embeddings for economic analyses. To address this concern, we verify that the visual attributes captured in the resulting embeddings are relevant to economic agents' perceptions by measuring (a generalized notion of) their correlation with "perceived" attributes. For

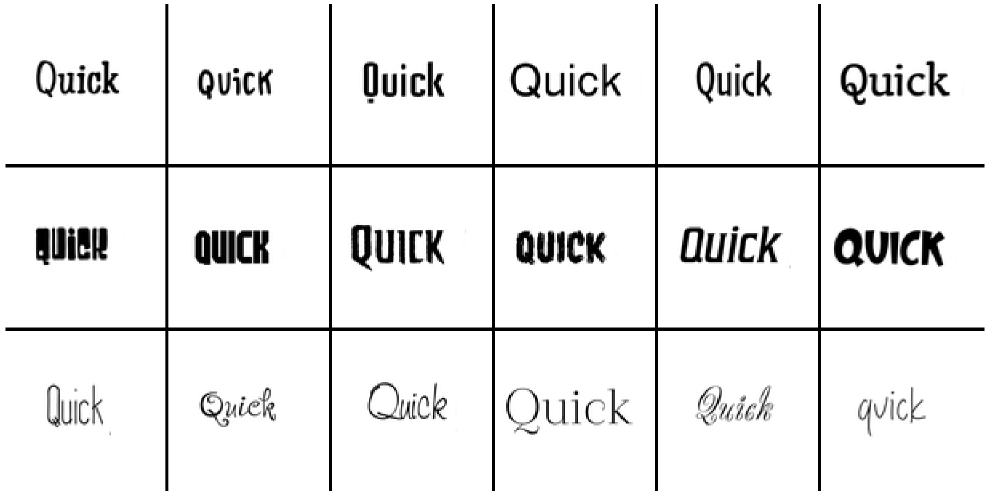


Figure 5: Nearest Neighbors (Collected in Each Row) in the 128-Dimensional Product Space

the latter, 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 the tags include nearly 30,000 different words. Therefore, we consolidate tags that are synonyms to create meta-tags that encompass many related words (e.g., "handwritten" and "cursive" would be in the same meta-tag). To this end, we create clusters of tags using a standard word embedding "Word2vec" (two-layer neural network) by Mikolov et al. (2013).

To measure the relevance between the image and word embeddings, we create clusters of the image and word embeddings, respectively, using K-means clustering.[14] Then, we measure how well the clusters of word embeddings match the clusters of image embeddings by using mutual information. Specifically, we use the normalized mutual 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 font image 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 the mutual information between F and W. The NMI can be interpreted as how informative W is in determining F. Its value ranges between 0 and 1 (the value 0 implies that W contains no information regarding F). Table 1 reports the values of

[14] To perform K-means clustering of font shape, we use the obtained 128-dimensional image embeddings directly clustered into 60 clusters using the elbow method. To perform K-means clustering of tags, we use 100-dimensional word embeddings that are first reduced to 10-dimensional embeddings and then clustered into 6 clusters using the elbow method.

F	W	$NMI(F, W)$
Image Embeddings	Word Embeddings	0.473
Industry Categories	Word Embeddings	0.261

Table 1: Normalized Mutual Information between Image Clusters and Word Clusters (first row), Compared to Baseline (second row)

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 as structured attributes instead of the image embeddings. In general, when images were treated as unobserved product attributes, then the main structured attribute available in many design products is product categories. Recall that sans serif, serif, slab serif, display, handwritten, and script are such product categories defined by the font industry.[15] In this case, we obtain $NMI(F, W) = 0.261$, which is roughly only half of the value in the first case. These results suggest that the learned image embeddings arguably capture economic agents' perceptions better than the structured attributes and can be relevant for economic analyses.

5 Effects of Merger on Design Differentiation

5.1 Background

On June 15, 2014, one of the major font foundries called FontFont was acquired by Monotype. At the time, Monotype sold fonts created by the foundries it owned as well as by third-party foundries. Before the merger, FontFont had sold its fonts through MyFonts.com as a third party. We study the causal effect of this merger on the change in the product differentiation decisions of the merging firm (i.e., FontFont foundry). In this market, the major channel for product differentiation is the design of font shapes. This creative decision of a foundry may be affected by the merger through many different channels. By merging, the firms could increase efficiency by reducing transaction costs. If costs are reduced, the merged firm might find it profitable to create more experimental products. The merged firm may also be concerned with cannibalization—that is, competition among their own products—which would increase the diversity of product design. On the other hand, if Monotype has a preemptive motive with the merger, it will crowd its products to prevent entries. In the

[15] Product categories are generally available to analysts, but it is specific to this online marketplace that tags are observed. This motivates the setup of having tags as the ground truth in this section.

presence of these opposing factors, whether the merger increases the design diversity is an empirical question. We answer this question using the embeddings we created as the main ingredient in a synthetic control method.

5.2 Design Differentiation Measures

In this merger analysis, the outcome of interest is the degree of design differentiation. We construct two measures for design differentiation based on the constructed embeddings. The first measure is the *distance to Averia*. We calculate the Euclidean distance between an individual font and a benchmark font:[16] for image x_i of font i,

$$D_i^A = \|f(x_i) - f_{averia}\|_2 \,,$$

where $f(\cdot) \in \mathbb{R}^{128}$ is the embedding and f_{averia} is the embedding of a benchmark font called "Averia," which is calculated by averaging the values of the embeddings of all the existing fonts in the market. The distance measure D_i^A is intended to capture the degree of product differentiation of font x_i.

In addition to D_i^A, we consider the following *gravity measure*: for image x_i of font i,

$$D_i^G = -\sum_{j \neq i} \frac{1}{\|f(x_i) - f(x_j)\|_2},$$

where the sum is for all other fonts j's 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 all j's and a small value when it is close to any of them. Compared to the distance to Averia, the gravity measure acknowledges the aspect of spatial competition among designers. For illustration, consider the interval $[0, 1]$ as the product space. Suppose two existing fonts are located at $\{0, 1\}$ in terms of their shapes, and the third font chooses to enter a location between one of $\{0, 1/2, 1\}$. The shape of the third font i would be most differentiated in terms of D_i^G if it is located at $\{1/2\}$. On the other hand, i would be the most differentiated product in terms of D_i^A if it is located at one of $\{0, 1\}$ (even though these points are already populated).

Finally, we aggregate each measure for all fonts created by foundry k in period t. In

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

particular, for $j \in \{A, G\}$, we construct

$$\bar{D}^j_{kt} = \frac{1}{|I_{kt}|} \sum_{i \in I_{kt}} D^j_i,$$

where I_{kt} is the set of all fonts created by foundry k in period t. In the subsequent merger analysis, we consider both $\bar{D}^A_{kt}$ and $\bar{D}^G_{kt}$ as the outcome variables, henceforth referred to as the *mean deviation* and *gravity measures*, respectively. Despite the distinct feature, both $\bar{D}^A_{kt}$ and $\bar{D}^G_{kt}$ are meant to succinctly capture each foundry's creative decision of product differentiation in terms of font design every period. We show that the subsequent analysis is robust to the choice of the measure between the two.

5.3 Data and Empirical Strategy

To estimate the effect of merger that is causal, we need a control group that would behave like the treatment group (i.e., FontFont) if it were not for the merger. The challenge is that there is only a single treated unit in the treatment group and it is difficult to find a single untreated unit that resembles the treated unit. A naive average of a control group would be a poor candidate. The synthetic control method developed by Abadie and Gardeazabal (2003) and Abadie et al. (2010) is useful to overcome this challenge. The method compares the treated unit with a "synthesized treated unit" or a synthetic control unit obtained 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 the fonts that the foundries created. 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 has not changed during the study period.

For our analysis, we construct the panel based on the dataset described in Section 3. The cross-sectional units of this panel are foundries and the time dimension is a bi-annual time series between 2002 and 2017. Table 2 shows the summary statistics for the variables (except the raw embeddings) in the panel. All the variables are constructed to be foundry-level. Glyph count is an important product specification and is considered to be related to quality.[17] The release frequency (i.e., maximum period between two releases), sales, number

[17] Glyph count is the number of characters including special characters in each font family. We calculate the average glyph count for all fonts produced by each foundry each period.

	Mean	S.D.	Min	Max
Mean Deviation	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
$N \times T$	$51 \times 32 = 1,632$			

Table 2: Summary Statistics (embeddings variables omitted)

of orders, average price, and age of foundries introduced are the other control variables we use. Our data include one treated unit (FontFont) and 50 control units.[18]

Section B in the Appendix contains a simple descriptive analysis of the supply- and demand-side trends for the shapes of fonts captured in the embeddings. The supply-side trend plots the differentiation measures constructed in the previous subsection for fonts newly introduced in the market every period. For the demand-side trend, we construct and plot analogous measures for fonts purchased every period. We find that, on average, newer fonts tend to be more differentiated than older fonts. How such differentiation decisions are affected by the change in market structure is the subject of the merger analysis. We also find that the demand-side trend is markedly stable, especially around the time of merger.[19] This supports the argument that the dramatic change in the producer behavior we find below in the merger analysis is not demand-driven.

5.4 The Effects of Merger

Figure 6 captures the main results of our causal analysis of merger using the gravity measure.[20] The results of the analysis using the mean deviation measure are presented in the Appendix. The solid line in the left panel presents the trend of the gravity measure of the fonts *newly* designed by FontFont in a given period. Around the period where FontFont was merged, which is indicated by a vertical line, the shape of FontFont's fonts substantially

[18]We may refine the control group by choosing foundries whose distribution of product categories are sufficiently different from FontFont's. This will further minimize strategic spillovers between treated and control units as discussed in Abadie (2021).

[19]In general, the stable demand-side trend is consistent with the industry norm. According to the industry experts we interviewed, font markets tend not to experience seasonality or short-term trends in consumer preferences, unlike in other design industries such as clothing.

[20]To reduce the scale, we take the logarithm of the positive part of $\bar{D}^G_{kt}$ and then put back the minus sign.

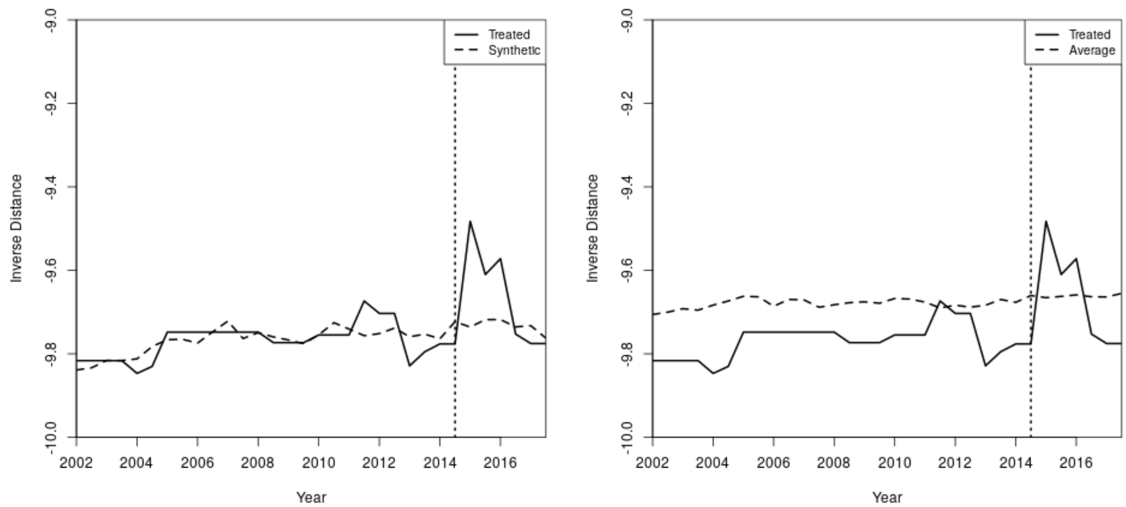


Figure 6: Trends of FontFont vs. Synthetic FontFont (left) and Naive Control Group (right)—Using Gravity Measure

Note: The solid line depicts the trend of FontFont and the dashed line depicts the trend of the synthetic control (left) or the naive average trend among all the control units (right). The vertical dotted line in each figure indicates the time of the merger.

differs from that of other fonts in the market. This before and after comparison alone cannot yield the causal effect of the merger, as other market conditions may have changed around this period. Comparing this trend with the trend of the synthetic FontFont, depicted by a dashed line, removes possible confounding factors. First, the two trends before the merger appear to be close to each other by construction. After the merger, however, FontFont tends to produce more experimental fonts (i.e., fonts that are far from others) relative to the trend of the synthetic control.[21]

To understand the virtue of our synthetic control method, we contrast the left panel with the right panel in Figure 6. 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 unit.

Finally, Table 3 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).[22] Consistent with Figure 6 (left), the short-run effects in the first and second years are sta-

[21]We also confirm that even if we backdate the period of acquisition (i.e., does not use the information of the timing), the relative increase still occurs around the same period as the vertical line.

[22]The advantage of this inferential method is that it does not assume random assignment of the policy intervention unlike earlier methods that rely on the assumption to ensure the properties of randomization tests (e.g., Abadie et al. (2010)). See also Arkhangelsky et al. (2019) for related discussions.

Years (After Merger)	2015	2016	2017
Treatment Effects	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 3: Treatment Effects Averages (by year after the merger)

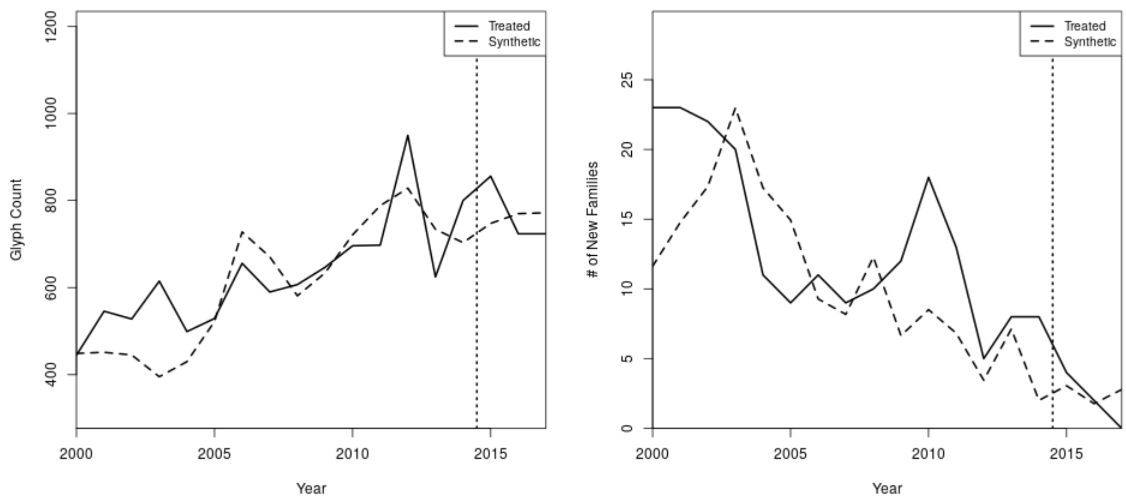


Figure 7: Trends of FontFont vs. Synthetic FontFont using Glyph Counts (left) and the Number of Products (right)

Note: The solid line depicts the trend of FontFont and the dashed line depicts the trend of the synthetic control. The vertical dotted line in each figure indicates the time of the merger.

tistically significant. To understand the magnitude of the effect, note that the standard deviation (SD) of the gravity measure in the sample is 0.16 as shown in Table 2. Therefore, the treatment effects are on average half of 1 SD. It is also informative to understand the increase in the measure (roughly from -9.8 to -9.6) relative to the distribution of the gravity measure (the left panel of Figure 14 in the Appendix). After two years of strong positive effects, the effect of merger dissipates in the third year. The placebo test results, presented in Table 5 in the Appendix, show that the treatment effects are not statistically significant before the merger.

We produce analogous results using the mean deviation measure instead of the gravity measure and find that the results are qualitatively similar. The effects are positive and statistically significant and on average larger than 1 SD; see Section C.2 in the Appendix. This suggests that our findings are robust to the choice of the measure of product differentiation as long as it is constructed based on the image embeddings. This robustness disappears if we use more traditional measures for product offerings, such as the number of products and specifications. Figure 7 shows that the merger has no effects on both the Glyph counts (which is the key specification for fonts) and the number of new fonts. Tables 9 and 11 in the Appendix confirm that the effects are not statistically significant.

Based on this analysis, we conclude that the merger caused FontFont to explore a new territory of the product space, at least temporarily.[23] That is, by being part of the parent organization, Monotype, FontFont increased the visual variety in font design. Before the merger, foundries owned by Monotype produced fonts and sold them on MyFonts.com. After the merger, Monotype may have incentives to diversify the product scope owing to the increased size and efficiency of the firm. It may also be the case that Monotype tries to avoid cannibalization by spreading apart its products, thus reducing competition amongst their own foundries. Such tendency disappears after two years, perhaps because the firm has either successfully foreclosed the market or found the strategy unprofitable.[24]

6 Conclusion

Certain policy questions are better answered by treating high-dimensional, unstructured attributes as observables and attempting to solve the resulting dimensionality problem. 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, as reflected in the mutual information between tags used to describe the fonts and the neural network embeddings. We then turn to a causal analysis to understand the effects of a merger on product differentiation. We find that the merger increases the product variety in this market. The analytical framework of this paper can be applied to various products where unstructured attributes are present.

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: construct a neural network embedding to reduce the dimensions 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. Therefore, its counterfactual predictions are

[23] Given our data frequency, the foundry's immediate increase in design differentiation after the merger is feasible because it typically takes one or two months for a foundry to design a font.

[24] According to the interview with the employees, we discovered that the company has gone through a structural change two years after the merger that is consistent with the second explanation. The details of the change cannot be publicly revealed.

not as credible as those of 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 may answer economic questions related to this market. Related to the economic analyses 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 the 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 changed the commission policy during the data collection period, which may serve as a 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 states 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, thus preventing other productions within: producing another font inside the product space is considered a 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 Neural Network Training

A.1 Details of Network Training

The training iteratively improves the parameters of the network using batches to estimate the gradient and then update the parameters accordingly (Wilson and Martinez (2003)). Each batch contains 270 cropped images of fonts, or equivalently, 90 triplets. We cropped each image based on the number of characters in the image.[25] As the gradient is evaluated at more batches, the parameters in the network are adjusted. The number of trainable parameters are $90,000 = 3 \times (3 \times 100 \times 100)$ (layers $\times$ input size). The training of the network is completed when the loss function reaches below 0.7. The learning rate is initially 0.05, and then lower to 0.01 to finalized the model. Here are the remaining hyper-parameter values: the batch

[25]For example, the first half of the pangram sentence has 20 characters; therefore, 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 such as using the white space between characters.

size is 90, epoch size 500, weight decay 1×10^{-4}, and the margin α is 0.2. The training time took approximately 24 hours with 4 GPUs (Nvidia 1080-TI).

A.2 Evaluation of Trained Network

To evaluate the neural network embeddings, we create test sets of triplets that the neural network has never seen during the training. Using the test sets, the task is to identify whether the image in a triplet is a positive or a negative. It is classified as a positive if its Euclidean 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 test set, "easy," 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 test set, "hard," 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). Table 4 presents the accuracy and validation rates for the neural network in each of these test sets with different crop sizes. The accuracy is a measure of how well the neural network embeddings perform in the classification task. Overall, it gets about 90 percent of the 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). In addition, we plot 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 8 shows the trade-off between these types of errors. There exists a steeper trade-off between precision and recall in the hard test set.

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

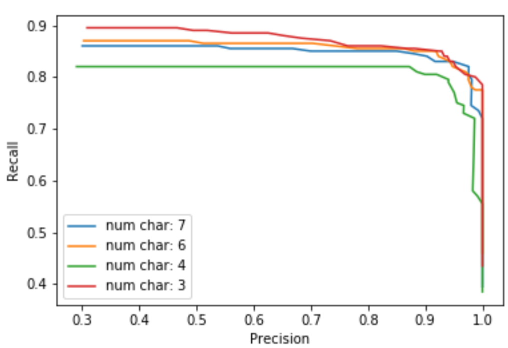
B Trend Analyses

To further illustrate the usefulness of the embeddings, we analyze the supply- and demand-side trends in font style. This also provides an additional background for the causal analysis of the merger in Section 5. On the supply side there are constant entries of new products in

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 4: Internal Validation by Accuracy, Validation Rate, and FAR

"Easy" Test Set	"Hard" Test Set



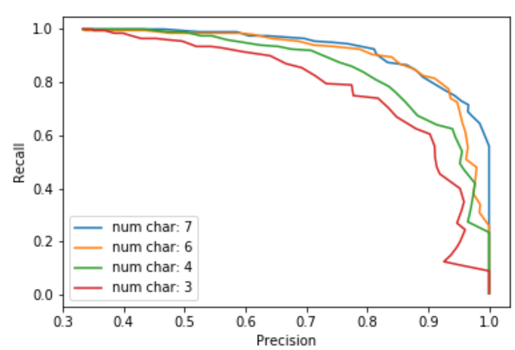


Figure 8: Precision Recall Curves

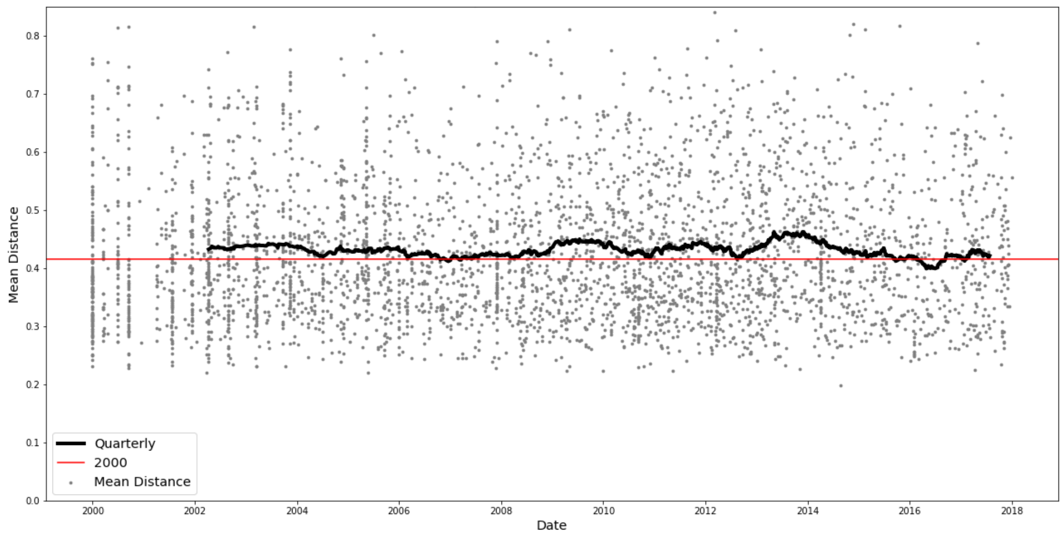


Figure 9: Trends in Mean Distance of Incumbents ($\leq$ 2000, in red) and New Entrants (2001–2017)

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

B.1 Supply-Side Trend

We analyze the supply-side trend in the Euclidean distance from Averia. We are particularly interested in how the style of fonts newly entering the market differs from the style of incumbent fonts. In Figure 9, each dot represents the daily mean distance from Averia for a range of periods. The red horizontal line is the mean distance of *all* fonts introduced before 2001, which we view as incumbents. The black line depicts the quarterly moving averages of the daily mean for fonts newly introduced since 2001 each day, which we view as entrants. Overall, we find that entrants have font shapes that are more experimental or innovative than those of incumbents, possibly to avoid competition and establish market power distant from the incumbents in the product space.

B.2 Demand-Side Trend

We now analyze the trend in the Euclidean distance between Averia and fonts purchased by consumers between 2014 and 2017.[26] To remove the supply-driven factor, we condition on

[26] The time span is shorter than that for the supply-side analysis due to the missing license type data for earlier periods.

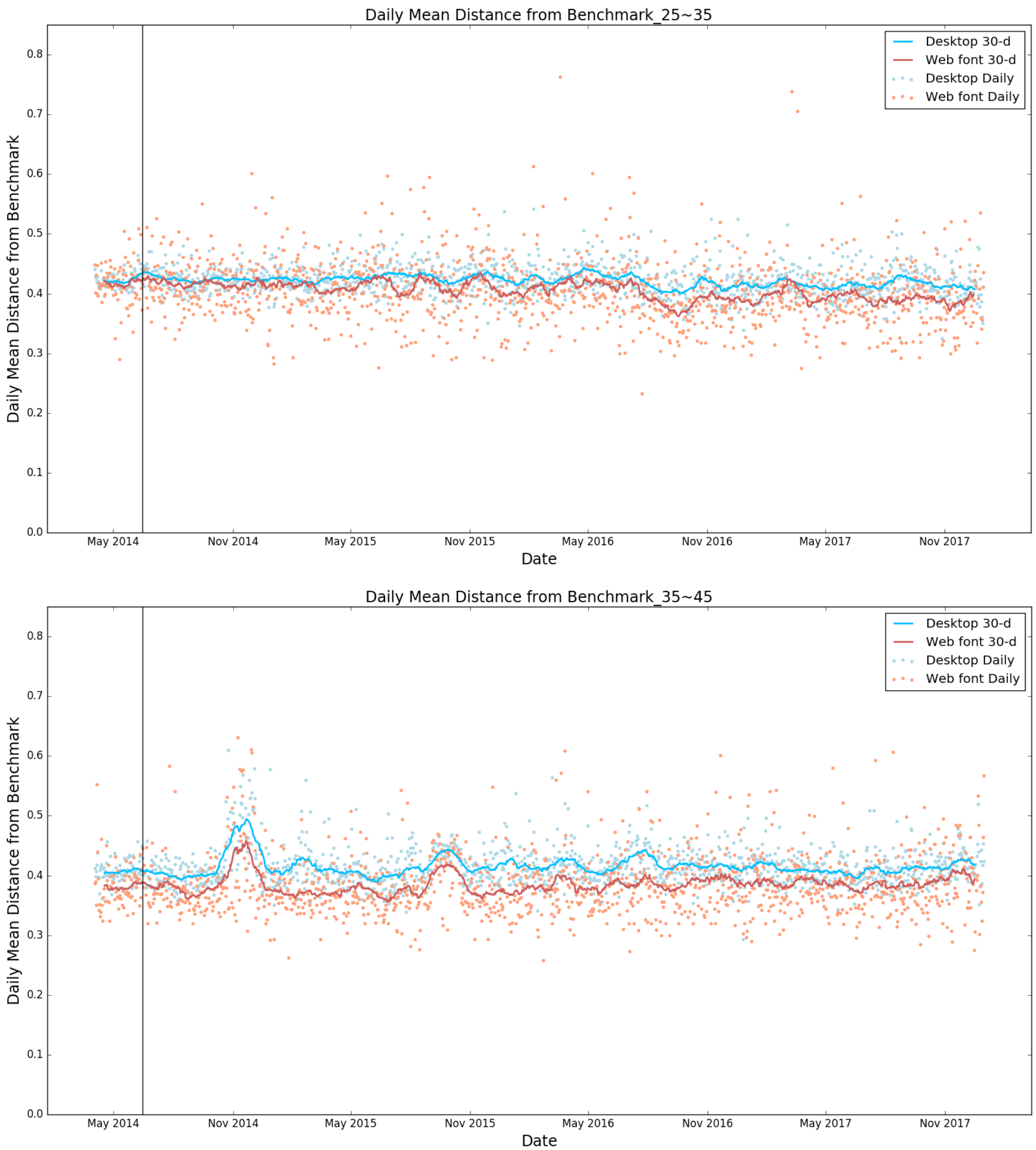


Figure 10: Trends in (Weighted) Mean Distance of Purchased Fonts (for USD 25–35 and USD 35–45, by license types)

Note: The vertical line in each figure indicates the time of the merger we analyze in Section 5.

price and focus on the price ranges of USD 25–35 and USD 35–45, where promotions are rare. These price ranges are the most common in the market. Figure 10 plots the trend conditional on the price being USD 25–35 and USD 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 sold under desktop license are more experimental (or less conservative) than fonts sold under web font license. Because both licenses are offered for most fonts, this difference cannot be attributed to the supply-side decision but is the result of consumer decisions.

B.3 Summary of Findings

Based on our analyses, we find the following stylized facts. (i) Entrants are more innovative than incumbents. (ii) Consumer preferences are stable during 2014, the year that witnessed the merger of our interest.[27] Based on this finding, we assume that the demand side has a negligible influence on the change in product differentiation decisions by the merging firm around the time of merger. (iii) Consumers have different preferences over shapes depending on the license type they purchase. Consumers prefer more conservative shapes for web font licenses and more experimental shapes for desktop licenses. Usually, web font licenses are used on webpages, where legibility is important, whereas desktop licenses are used in printed material (e.g., posters, cards), where designers (as consumers) have more control over the design environment.

[27] This feature is uniformly found in all price ranges, but we do not report the results for succinctness.

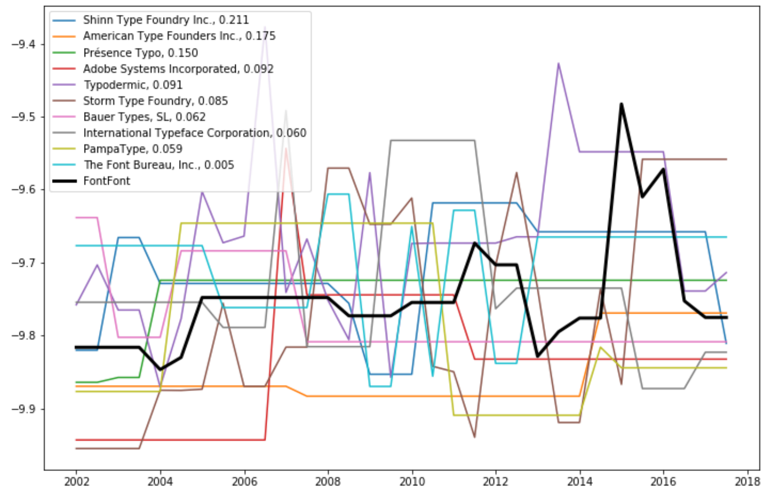


Figure 11: Trends of FontFont vs. Top 10 Control Units (Using Gravity Measure)

Note: The values next to the foundry names indicate the weights used to construct the synthetic control.

C Supplemental Findings for Merger Analysis

C.1 Placebo Test with Gravity Measure

	2003	2004	2005	2006	2007	2008
Treatment Effects	-0.0172	-0.0158	-0.0298	0.013	0.0473	0.0351
p-Value (block)	0.75	0.1667	1	0.5	0.0833	0.9286
p-Value (i.i.d.)	0.6577	0.0654	0.9672	0.5029	0.0668	0.9376
	2009	**2010**	**2011**	**2012**	**2013**	**2014**
Treatment Effects	-0.0111	-0.0288	0.012	0.0727	-0.0316	-0.0659
p-Value (block)	0.8125	0.6667	0.2	0.3636	0.1667	0.5385
p-Value (i.i.d.)	0.873	0.5351	0.1068	0.2523	0.1452	0.5335

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

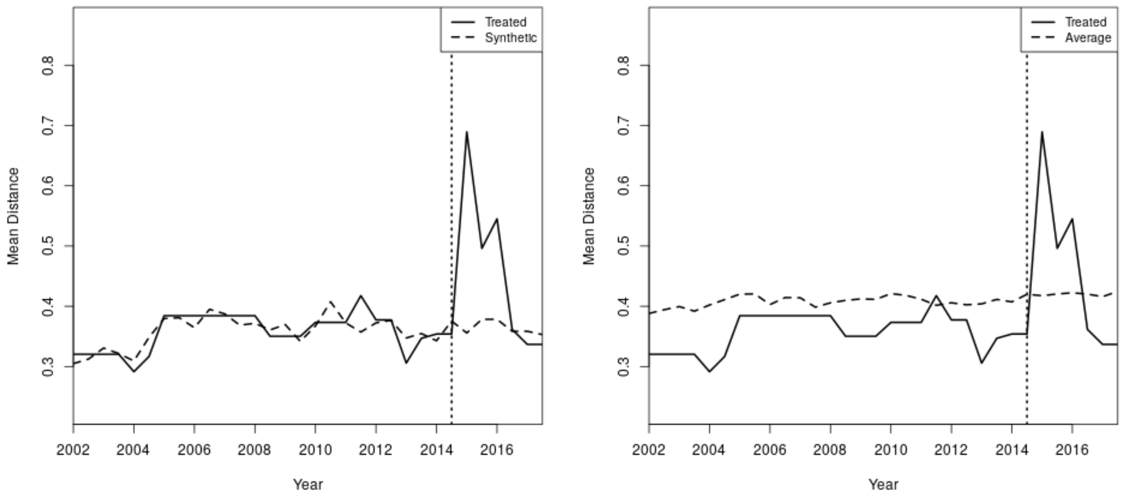


Figure 12: Trends of FontFont vs. Synthetic FontFont (left) and Naive Control Group (right)—Using Mean Deviation Measure

Note: The solid line depicts the trend of FontFont and the dashed line depicts the trend of the synthetic control (left) or the naive average trend among all the control units (right). The vertical dotted line in each figure indicates the time of the merger.

C.2 Merger Effects and Placebo Test with Mean Deviation Measure

	2003	2004	2005	2006	2007	2008
Treatment Effects	-0.0036	-0.0095	0.0177	0.0241	0.0388	0.005
p-Value (block)	0.75	0.3333	0.25	0.3	0.1667	0.7857
p-Value (i.i.d.)	0.8292	0.3391	0.1486	0.1494	0.0836	0.7552

	2009	2010	2011	2012	2013	2014
Treatment Effects	-0.0444	-0.0224	0.0371	0.0323	-0.0144	-0.0056
p-Value (block)	0.125	0.7222	0.1	0.8182	0.625	0.7308
p-Value (i.i.d.)	0.096	0.6569	0.0764	0.803	0.5563	0.7047

Table 6: Placebo Test: Treatment Effects Before Merger (Using Mean Deviation)

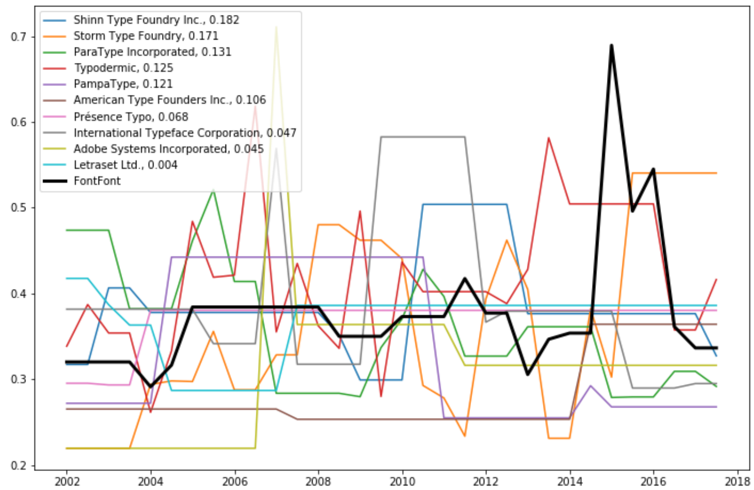


Figure 13: Trends of FontFont vs. Top 10 Control Units (Using Mean Deviation)

Note: The values next to the foundry names indicate the weights used to construct the synthetic control.

	2015	2016	2017
Treatment Effects	0.1412	0.061	-0.0305
p-Value (block)	0.037	0.0741	0.1111
p-Value (i.i.d.)	0.0034	0.0704	0.0716

Table 7: Treatment Effects After Merger (Using Mean Deviation)

C.3 Merger Effects with Traditional Measures of Product Offerings

C.3.1 Glyph Counts

	2003	2004	2005	2006	2007	2008
Treatment Effects	125.1264	48.1483	-19.9343	-14.6911	-57.5153	1.1326
p-Value (block)	0.25	1	0.8333	0.8571	0.375	0.4444
p-Value (i.i.d.)	0.2376	1	0.829	0.8534	0.3759	0.4469
	2009	2010	2011	2012	2013	2014
Treatment Effects	29.1529	-23.4679	29.4617	178.8836	99.5249	11.8745
p-Value (block)	0.7	0.5455	0.25	0.0769	0.5714	0.4
p-Value (i.i.d.)	0.6991	0.5409	0.2579	0.0786	0.5749	0.4063

Table 8: Placebo Test: Treatment Effects Before Merger (Using Glyph Counts)

	2015	2016	2017
Treatment Effects	127.9152	74.8633	73.8973
p-Value (block)	0.25	0.75	0.75
p-Value (i.i.d.)	0.2645	0.7395	0.7574

Table 9: Treatment Effects After Merger (Using Glyph Counts)

C.3.2 Number of New Fonts

	2003	2004	2005	2006	2007	2008
Treatment Effects	48.5002	1.0007	16.0004	232	248	156
p-Value (block)	1	0.8	0.6667	0.1429	0.625	0.2222
p-Value (i.i.d.)	1	0.8034	0.6677	0.1408	0.6241	0.229
	2009	2010	2011	2012	2013	2014
Treatment Effects	64.5	210.0034	376.5	144.5	164	155
p-Value (block)	0.4	0.0909	0.4167	0.6154	0.2857	0.7333
p-Value (i.i.d.)	0.3939	0.0938	0.4115	0.6137	0.2901	0.7297

Table 10: Placebo Test: Treatment Effects Before Merger (Using Number of New Fonts)

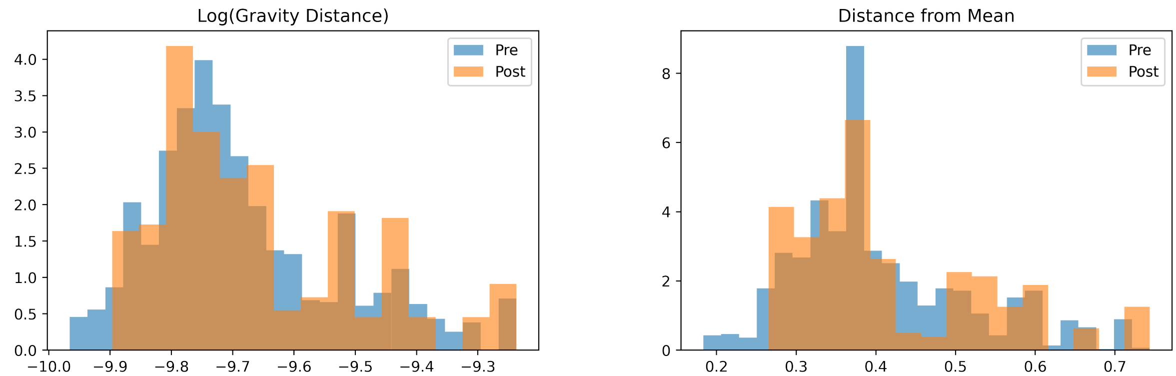


Figure 14: Histograms of the Gravity Measure (left) and Mean Deviation Measure (right), Before and After the Merger

	2015	2016	2017
Treatment Effects	73.5002	29.0002	32.0001
p-Value (block)	0.5625	0.9375	1
p-Value (i.i.d.)	0.5599	0.9426	1

Table 11: Treatment Effects After Merger (Using Number of New Fonts)

C.4 Distributions of the Measures

References

Abadie, A. (2021): "Using synthetic controls: Feasibility, data requirements, and methodological aspects," *Journal of Economic Literature*, 59, 391–425. 18

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, 5.3, 22

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, 5.3

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. 1.1

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. 6

Arkhangelsky, D., S. Athey, D. A. Hirshberg, G. W. Imbens, and S. Wager (2019): "Synthetic difference in differences," Tech. rep., National Bureau of Economic Research. 22

Ashenfelter, O. and D. Hosken (2008): "The effect of mergers on consumer prices: Evidence from five selected case studies," *National Bureau of Economic Research*. 1.1

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

Bajari, P., Z. Cen, V. Chernozhukov, M. Manukonda, J. Wang, R. Huerta, J. Li, L. Leng, G. Monokroussos, S. Vijaykunar, et al. (2021): "Hedonic prices and quality adjusted price indices powered by AI," *cemmap working paper*. 1.1

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile prices in market equilibrium," *Econometrica: Journal of the Econometric Society*, 841–890. 1

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, 1.1, 7

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. 1.1

Campbell, N. D. and J. Kautz (2014): "Learning a manifold of fonts," *ACM Transactions on Graphics (TOG)*, 33, 91. 1.1

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," . 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*. 5.4

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. 1.1

Fan, Y. (2013): "Ownership Consolidation and Product Characteristics: A Study of the US Daily Newspaper Market," *American Economic Review*, 103, 1598–1628. 1.1, 7

Foster, D. J. and V. Syrgkanis (2019): "Orthogonal statistical learning," *arXiv preprint arXiv:1901.09036*. 6

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

Galenson, D. W. and B. A. Weinberg (2000): "Age and the quality of work: The case of modern American painters," *Journal of Political Economy*, 108, 761–777. 1

——— (2001): "Creating modern art: The changing careers of painters in France from impressionism to cubism," *American Economic Review*, 91, 1063–1071. 1

Gentzkow, M., B. Kelly, and M. Taddy (2019a): "Text as Data," *Journal of Economic Literature*, 57, 535–74. 1.1

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. 1.1

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. 1.1

Goodfellow, I., Y. Bengio, and A. Courville (2016): *Deep Learning*, MIT Press, http://www.deeplearningbook.org. 1

Gross, D. P. (2016): "Creativity under fire: The effects of competition on creative production," *Review of Economics and Statistics*, 1–17. 1.1

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. 1.1

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. 1.1

Hoberg, G. and G. Phillips (2016): "Text Based Network Industries and Endogenous Product Differentiation," *Journal of Political Economy*, 124, 1423–1465. 1.1

Hotelling, H. (1929): "Stability in Competition," *Economic Journal*, 39, 41–57. 1

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. 1.1

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. 1.1

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. 1.1

Lancaster, K. (1971): *Consumer demand: A new approach*, Columbia University Press New York. 1

Lancaster, K. J. (1966): "A new approach to consumer theory," *Journal of political economy*, 74, 132–157. 1

LeCun, Y., C. Cortes, and C. Burges (2010): "MNIST handwritten digit database," *AT&T Labs*. 1.1

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. 1.1

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. 1.1

McFadden, D. (1973): *Conditional logit analysis of qualitative choice behavior*, Institute of Urban and Regional Development, University of California Berkeley. 1

Mikolov, T., K. Chen, G. Corrado, and J. Dean (2013): "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*. 4.4

Nevo, A. (2001): "Measuring market power in the ready-to-eat cereal industry," *Econometrica*, 69, 307–342. 1.1

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. 1.1, 6

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. 1.1

Parikh, D. and K. Grauman (2011): "Relative attributes," in *2011 International Conference on Computer Vision*, IEEE, 503–510. 1.1

Rosen, S. (1974): "Hedonic prices and implicit markets: product differentiation in pure competition," *Journal of political economy*, 82, 34–55. 1

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, 1.1, 4.1

Seim, K. (2006): "An Empirical Model of Firm Entry with Endogenous Product-Type Choices," *The RAND Journal of Economics*, 37, 619–640. 6

Simonyan, K. and A. Zisserman (2014): "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*. 1.1

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. 1.1, 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. 1.1, 7

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. 1.1, 4.1

Tenenbaum, J. B. and W. T. Freeman (2000): "Separating style and content with bilinear models," *Neural computation*, 12, 1247–1283. 1.1

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, 1.1

Wilson, D. R. and T. R. Martinez (2003): "The general inefficiency of batch training for gradient descent learning," *Neural networks*, 16, 1429–1451. A.1

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. 1.1

Zhang, S., D. D. Lee, P. V. Singh, and K. Srinivasan (2017): "How much is an image worth? Airbnb property demand estimation leveraging large scale image analytics," *SSRN*. 1.1