Regulating data usage by dual platforms

Malika Korganbekova
The University of Chicago
Booth School of Business

Alibek Korganbekov
Boston University
Questrom School of Business

September 27, 2024

Abstract

We examine the necessity and the design of data usage regulation in B2B markets, focusing on Amazon that functions as both a marketplace and a seller. The key issue is Amazon's alleged use of internal data to replicate top-performing third-party seller products to boost its own private label business. We train Deep Learning models, Siamese Neural Network and Sentence Autoencoder, to construct visual and textual similarity measures between more than 600 Amazon Basics and 2 million third-party seller products in ten product categories. The findings reveal significant and consistent similarities between private label and third-party products in all ten categories. Next, this paper questions the necessity and the effectiveness of recent duration-based regulation proposed in the European Union. Results suggest that Amazon typically takes around 2.5-3 years to replicate a product, while smaller sellers' products take approximately 5 years to replicate. This indicates that the duration-based data regulation settled on in the European Union provides Amazon with ample time to collect data, suggesting the need for a different data regulation approach.

1 Introduction

An increasing number of digital platforms, such as Amazon, Walmart and Target, operate in dual mode: they serve as *marketplaces* that match third-party sellers with consumers, and as *sellers* that manufacture and sell their own private label products. Platform's dual role has sparked regulatory concerns regarding potential conflicts of interest, as highlighted in the 2020 U.S. House Majority Report. Regulators are concerned that the platforms use third-party sellers' data to inform their own private label product offerings, which, in turn, may (i) hurt third-party sellers' performance, and (ii) reduce third-party sellers' incentives for innovation. These concerns led to antitrust allegations against Amazon in Europe. Amazon settled the antitrust case by agreeing to abstain from using third-party sellers' data and to commit to fair business practices for up to five-seven years.[1] Despite increasing policy concerns, there is little empirical evidence that assesses the necessity and effectiveness of different types of data regulation on dual platforms.

[1]For Prime-related concerns, Amazon committed to not engage in any anti-competitive practices for seven years, and for all other charges, they committed to five years - Reuters link.

This paper has three goals. First, we investigate the similarity between Amazon Basics and third-party sellers' products to establish that there is a cause for concern and Amazon may have imitated third-party sellers' products. Although there have been evidence of product imitation in Asian markets,[2] to the best of our knowledge, there have not been systematic evidence of product imitation in Amazon US, the largest e-commerce market in the US. To this end, we assemble a comprehensive dataset describing Amazon's private label and third-party business. The data we scraped directly from Amazon include (i) product page information, such as product introduction dates, text description, images, ratings, review counts, and (ii) a 3-year panel of prices and bestseller ranks. We concentrated on 10 major categories that contain more than 600 Amazon Basics products and 2 million third-party sellers' products. We train Deep Learning models to construct visual and textual similarity measures between Amazon and third-party sellers' products using products' images and text descriptions. We find significant similarities between Amazon Basics and third-party sellers' products with a median top cosine similarity score of 0.93 out of 1.

Second, we explore whether there is any indication suggesting that Amazon possessed exclusive knowledge about product performance, enabling them to rapidly replicate the products. However, our findings reveal that the median Amazon Basics product is introduced approximately 2.6 years (960 days) after potentially imitated third-party seller products. Interestingly, it takes twice as long to replicate smaller sellers' products. These observations indicate that Amazon may not possess private information regarding product performance, instead relying on publicly available data to inform their Amazon Basics product offerings. Nevertheless, regulatory concerns may still arise due to the significant data scale advantages held by Amazon, stemming from better data infrastructure compared to that of a smaller third-party seller.[3,4]

Third, we examine the potential efficacy of duration-based data regulation, as agreed upon by the European Union and Amazon, and provide evidence suggesting that such regulation may not be effective. Instead, it is crucial for regulators to consider the diverse impacts that Amazon Basics products have on third-party sellers.

The cornerstone of this paper, which serves as the basis for all subsequent analysis, lies in the development of visual and textual similarity measures. We train Siamese Neural Network (Bell and Bala, 2015) to extract lower-dimensional vector representation of the images, so that every image in our dataset is converted into 512×1 vector. Next, we calculate the cosine similarities between the images to infer their similarity on a scale of $[-1, 1]$.

To construct textual similarity measures, we use Universal Sentence Encoder model developed

[2]See Amazon alleged to have systematically targeted some of the most popular Indian brands to boost its private label business for further information.

[3]See Google Antitrust case that discusses the importance of data scale.

[4]Technically, any third-party seller can scrape Amazon to get information on other products' images, text descriptions, product ranks on the ranking page, and bestseller ranks to replicate products.

by Google (Yang et al., 2019). Each product description is transformed into 512×1 vector and we calculate the cosine similarity between product descriptions.

We match Amazon Basics products with the most similar third-party sellers' products using the constructed visual and textual similarity measures. To illustrate, Figure 1 shows an example output from the Siamese Neural Network model. The left figure is the Amazon Basics product and the right figure is the third-party seller's product that were identified as similar by the Deep learning model: cosine similarity between them is 0.97. Typically, products that have a cosine similarity of 0.9 and higher are visually extremely similar.

Figure 1: Example product match



Notes: The product on the left costs $132.28 and is an Amazon Basics product, while the image on the right is the third-party seller's product that costs $146.83.

Product matches play a pivotal role for multiple purposes. First, they act as a proxy for potential product imitation, allowing me to identify instances of uncanny similarities between Amazon Basics and third-party seller products. Second, by utilizing these matches, we can obtain quasi-experimental effects of Amazon Basics entry on third-party sellers. To do that, we identify third-party sellers' products that exhibit high levels of similarity to Amazon Basics products (cosine similarity of 0.9 and high). We mark these third-party sellers' products as *potentially imitated*. Next, we construct visual and textual similarity measures between potentially imitated third-party sellers' products and non-imitated other third-party seller products. This allows me to construct comparable matches between third-party seller products. Finally, we compare the performance outcomes between the potentially imitated products and *matched* non-imitated third-party sellers' products before and after Amazon Basics entry. This comparative analysis enables me to explore the impact of potential product imitation on the performance of third-party sellers. Combination of these results allows me to evaluate the extent of potential imitation of third-party seller products by Amazon, evaluate its impact on third-party seller outcomes, and assess the necessity and effectiveness of different types of data regulation.

This project contributes to the growing theoretical and empirical literature on dual mode, in which a platform competes with third-party sellers on its marketplace. Some papers, Hagiu and Spulber (2013) investigate whether a platform should offer its own products and services while serving as a marketplace. Recent work by Farronato et al. (2023) investigates self-preferencing on Amazon. Farrell and Katz (2000) and Jiang et al. (2011) analyze the trade-off that dual platform faces: on the one hand, platform might want to incentivize innovation by third-party sellers, while on the other hand, platform wants to extract short-term rents by introducing private labels. Zhu and Liu (2018) show that Amazon is more likely to compete with its marketplace sellers in the product categories that are more successful in terms of sales. While existing literature (Hagiu et al. (2022), Lee and Musolff (2021), Farronato et al. (2023)) concentrate on consumer outcomes, this project aims at investigating the effectiveness of different data use regulation that would protect third-party sellers. A closely related theoretical paper is Madsen and Vellodi (2022).

Overall, despite growing regulatory concerns over dual role of digital platforms as both the marketplace and the seller, there is a lack of empirical evidence that would credibly show the extent to which Amazon engages in product imitation and that would evaluate various data use regulations. Such evidence is critical to help design the regulation of dual platforms.

The rest of the paper is organized as follows. Section 2 describes the empirical setting, the data collection process and provides data summary. Section 3 explains the Deep Learning algorithm used to construct product similarity measures and presents matching results. Section 4 discusses the effectiveness of different data regulation policies. Section 5 concludes.

2 Empirical Setting

A private label refers to a product manufactured by one company but sold under the brand name of another company. Major offline retailers like Walmart, CVS, and Target have established multiple private label brands that cover a wide array of products across various categories. Walmart offers its private label brand called Great Value, Target carries Up&Up brand, and CVS carries CVS Health. In a similar fashion, Amazon entered private label business around 2006 with its first Amazon Basics product.[5] There are now multiple Amazon private label brands, such as Amazon Essentials, Pinzon, and Rivet.[6] Overall, retailers launch private labels because they provide affordable alternatives to branded products, fostering customer loyalty, while also offering the potential for higher profit margins due to lower manufacturing costs.

The introduction of private labels in major offline retail business dates back to Walmart's launch

[5]In our scraped dataset this product (Vacmaster 9 Inch Crevice Tool Wet/Dry Vac Accessory, 2 1/2" Diameter, V2CT , Black) is the earliest product that has Amazon Basics brand on the product page. However, the manufacturer is not Amazon Basics, but Crevo.

[6]See the article for a list of confirmed private label brands.

of its first private label brand, Ol' Roy, in 1983. Moreover, offline retailers have historically utilized consumer and product data, both internal and external, to inform their private label offerings (Dubé, 2022). However, there have been recent allegations, particularly against Amazon, regarding anti-competitive practices. Khan (2017) highlights that Amazon stands out from other retailers due to its unparalleled access to data, which it allegedly uses to favor its own private label products.

This paper focuses on examining two aspects related to Amazon's practices: product imitation of third-party sellers' products and exclusive data access that potentially aids in replicating products more efficiently. The study provides evidence that supports the first claim of product imitation by Amazon. However, it does not find empirical evidence to substantiate the assertion of exclusive data access that would enable Amazon to replicate products more efficiently. Furthermore, the paper suggests that current data regulation may not be effective and proposes that regulators should prioritize the protection of small third-party sellers. The argument is that while other retailers may imitate larger businesses that can defend themselves, Amazon potentially imitates smaller sellers who may not have the means to retaliate.

2.1 Data

In this paper, we use the data scraped from Amazon website. Specifically, we collected Amazon Basics and third-party sellers' product pages across ten product categories. The resulting dataset contains 624 Amazon Basics products and more than 2 million third-party seller products in ten product categories. Table 1 shows the number of products scraped in each category, where third-party seller products are split into FBA (Fulfilled by Amazon) and non-FBA (not Fulfilled by Amazon).[7]

Table 1: Scraped product categories

	Amazon Basics	Third-party Sellers	
		FBA	Non-FBA
Pet Supplies (Dog Beds and Mats)	122	47,306	671,358
Office Chairs	76	28,107	245,699
Cat Condos and Cat Trees	76	10,642	135,944
Dinnerware Sets	78	9,912	51,552
Dutch Ovens	75	4,022	16,527
Bookcases	55	24,644	171,670
Desks	44	43,583	337,164
Suitcases	49	3,582	9,434
End Tables	23	39,304	244,691
Ottomans	26	34,660	267,001
Total	624	245,762	2,151,040

Notes. This table reports the number of Amazon Basics and the number of Fulfilled by Amazon and non-Fulfilled by Amazon products across ten scraped product categories.

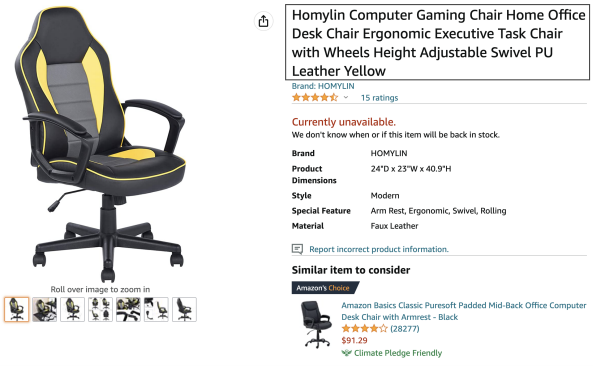
There are 624 Amazon Basics, 245,762 Fulfilled by Amazon (FBA) and 2,151,040 non-FBA third-party seller products in our sample. We specifically chose categories where products are visually different and ignored categories such as cables or bed sheets, where products are largely homogeneous.

[7] See the link for additional details on Amazon Fulfillment.

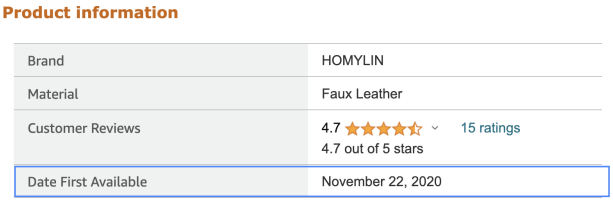
For each product, we collect the information contained on the product pages (Figure 2a): product's price, rating, number of ratings, number of product reviews, textual description, and all available product images. We also collect the information contained in the 'Product Information' part of the page. It has data on the brand, detailed specifications of the product, bestseller rank, and, most importantly, date when the product became available for the first time. The "Date First Available" is vital for establishing the product introduction sequence and ensuring accurate comparisons between Amazon Basics and *pre-existing* third-party seller products.

Figure 2: Example product page

(a) Product data



(b) Product Information part of the product page



In the first part of the paper, we investigate the similarities between Amazon Basics and third-party sellers' products. The elements of the data that are crucial for this part are product descriptions, images, and product introduction dates (Date First Available). For each product in the sample, we collected all images available in the carousel (Figure 3). The average number of images available per product is 2.6, with some products featuring only one image and others having multiple images. As a result, we train the visual similarity algorithm with more than 6 million images.

Figure 3: Image carousel of the same product on Amazon

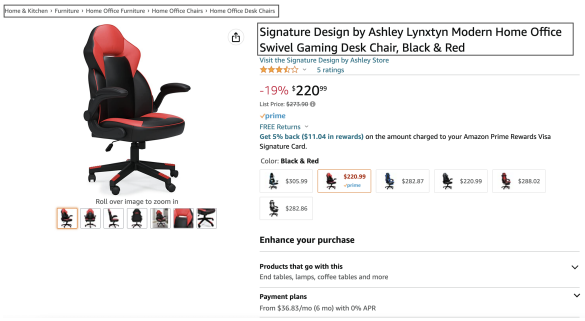


For the textual similarity algorithm, we collected data on the category taxonoour and product names, as well as the itemized product descriptions for each product (Figure 4). We use these data

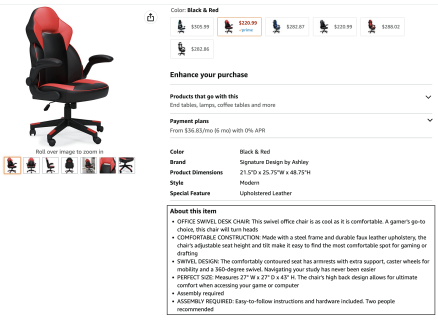
to subsequently train Deep Learning algorithms to construct textual similarity measures between products. We provide detailed explanations of textual similarity construction in the Appendix A. This version of the paper focuses on the matching results based on the image similarity algorithm.

Figure 4: Product descriptions

(a) Category and product names



(b) Itemized product description



Notes: The left-hand panel shows which part of the product page was scraped to identify the product descriptions and category taxonomy. The right-hand side boxed part shows the scraped itemized list of detailed product descriptions.

In the second part of the paper, we assess how fast Amazon was able to replicate third-party sellers' products and how the Amazon Basics entry affected the performance of the potentially imitated products. To this end, we use product introduction dates to calculate the days between the Amazon Basics entry and the third-party sellers' entry.

To assess how private labels affect third-party sellers, we collected a panel of prices, ratings, reviews, bestseller ranks, and product availability, starting January 2020 up to January 2023.[8] The panel is unbalanced with varying frequencies of snapshots, ranging from weekly to monthly, attributed to the difficulty in scraping. The panel data allows me to investigate the changes in third-party sellers' performance metrics, such as their bestseller rankings and product availability. Whenever a product is marked as 'Currently Unavailable', we view it as discontinued product which is viewed as a non-desirable outcome for the third-party seller. we report quasi-experimental results by comparing *potentially imitated* third-party seller products to the *matched* most similar *non-imitated* third-party seller products.

Finally, to assess the effectiveness of different types of regulation, we analyze heterogeneous responses of sellers to Amazon Basics entry. For each seller, we identify the number of products they carry, and we calculate the estimated weekly sales from bestseller ranks following the approach proposed in He and Hollenbeck (2020). Next, we split sellers into quintiles based on the estimated weekly sales and assess how Amazon Basics entry affected different seller quintiles.

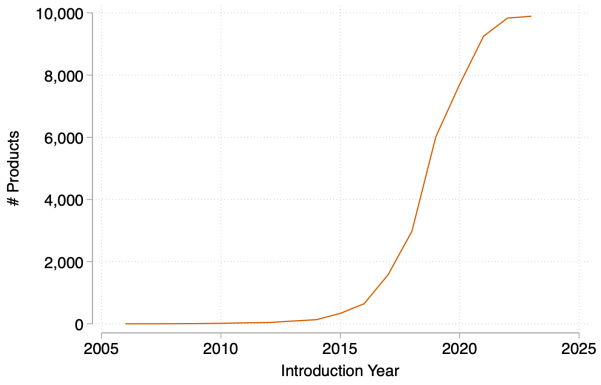
[8]The scraping process is still ongoing.

2.2 Descriptive Statistics

Amazon has been introducing Amazon Basics products since 2006 and the number of products has been increasing since 2014-2015, reaching almost 10,000 products in 2022. Figure 5a shows the cumulative number of products across *all* product categories and Figure 5b shows the number of products added by month. Starting 2021, there have been a decrease in the number of Amazon Basics products added.

Figure 5: Overall Amazon Basics entry

(a) Cumulative number of Amazon Basics products



(b) Newly added Amazon Basics products

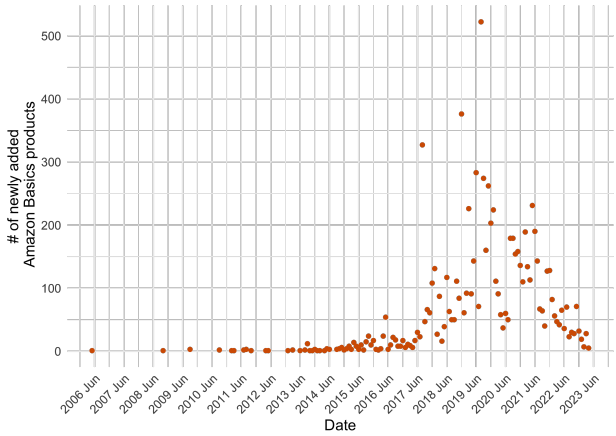


Figure 6 shows the entry in the ten main scraped categories: Figure 6a shows that the number of Amazon Basics products started to increase in 2017. Figures 6b and 6c show the entry dynamics in Fulfilled by Amazon and non-Fulfilled by Amazon business. In Appendix Figure 22, we show that the share of FBA and non-FBA products in the data is consistent with publicly available resources, where in some years the share of FBA products reaches 50%.[9] we use this as a validation check that the scraped data is sufficiently representative of the overall platform dynamics.

There are several distinct properties of Amazon Basics products compared to the third-party seller products, which we describe in Table 2. First, Amazon Basics are almost three times cheaper than the FBA products and almost twice cheaper than non-FBA products. The median prices for Amazon Basics, FBA and non-FBA products are \$64, \$169 and \$112, respectively. Second, most Amazon Basics products (96%=604/624) have at least one rating and their mean and median ratings are higher or equal to those of FBA and non-FBA products. Finally, Amazon Basics products have substantially higher number of reviews than third-party sellers. Moreover, the majority of Amazon Basics reviews are 5-star reviews (74%=2089/2818), while FBA and non-FBA products have 57% and 63% share of 5-star reviews, respectively.

Summary statistics indicate that consumers are attracted to lower prices, and purchase Amazon Basics products more. Moreover, the Amazon Basics quality is probably high which explains higher

[9] See Statista report.

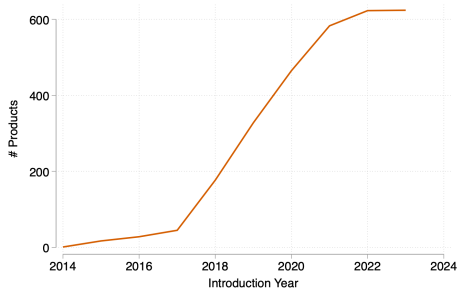
Table 2: Summary Statistics

	Observations	Min	Mean	Median	Max	SD
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Private Labels						
Sale Price	624	6.99	80.64	64.08	384.99	59.33
Star Rating (1-5)	604	3.40	4.55	4.60	5.00	0.23
# Reviews	624	0.00	10,006.26	2,443.50	216,932.00	23,815.68
# Reviews (>0)	604	4.00	10,337.59	2,818.50	216,932.00	24,136.44
5-star reviews	603	2.00	7,933.97	2,089.00	162,699.00	18,379.41
1-star reviews	603	0.00	316.71	117.00	7,010.00	751.48
Panel B: Third-party Seller (FBA)						
Sale Price	245,762	5.99	344.19	169.92	2,088.86	426.72
Star Rating (1-5)	112,361	1.00	4.35	4.50	5.00	0.79
# Reviews	245,762	0.00	914.41	0.00	220,970.00	6,919.65
# Reviews (>0)	112,227	1.00	2,002.44	21.00	220,970.00	10,132.92
5-star reviews	111,530	0.00	1,502.18	12.00	5,816,599.00	18,927.82
1-star reviews	111,530	0.00	115.82	2.00	1,812,258.00	5,449.96
Panel C: Third-party Seller (non-FBA)						
Sale Price	2,151,040	5.99	297.48	112.79	2,088.86	495.27
Star Rating (1-5)	336,964	1.00	4.33	4.60	5.00	0.88
# Reviews	2,151,040	0.00	91.84	0.00	154,434.00	1,604.55
# Reviews (>0)	336,823	1.00	586.55	11.00	154,434.00	4,018.94
5-star reviews	333,882	0.00	420.98	7.00	99,222.00	2,905.74
1-star reviews	333,882	0.00	29.63	1.00	24,709.00	207.99

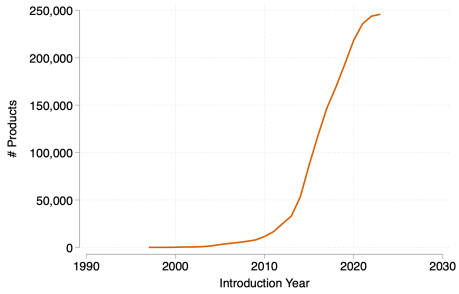
Notes: This table shows summary statistics by the product type.

Figure 6: Number of products over time

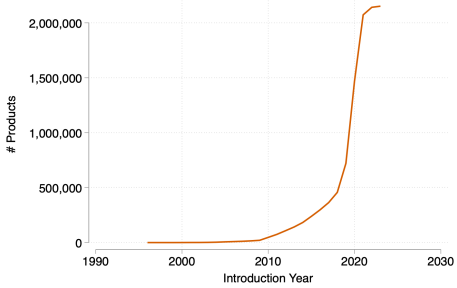
(a) Private Labels



(b) 3rd Party Sellers (Fulfilled by Amazon)



(c) 3rd Party Sellers (not Fulfilled by Amazon)



number of positive reviews and higher average rating. Therefore, it could be the case that the existence of Amazon Basics products is good for consumer welfare. However, the focus of this paper is the dual aspect of the platform. Specifically, the paper aims to evaluate the extent to which Amazon imitates third-party sellers' products, and to assess different data regulation policies. To do that, first, we build similarity measures between Amazon Basics and third-party sellers' products. Next, we use these measures to understand pricing and entry decisions, and to assess how data regulation policies will impact different sellers.

3 Product similarity and matching results

In this section, we describe the Image Embeddings algorithm and the matching results. we leave the explanation of the textual recognition algorithm to Appendix A.1.

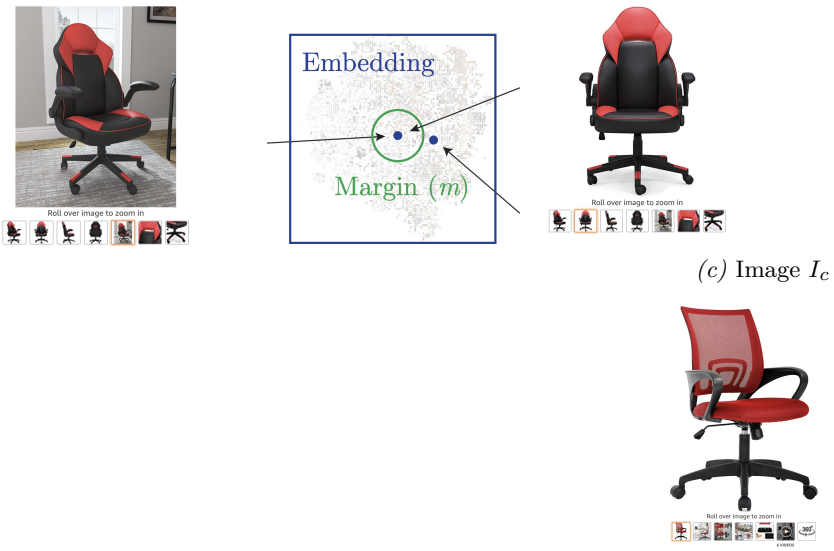
3.1 Image Embeddings

We use Deep Learning, namely Convolutional Neural Networks (CNNs), to construct image similarity. The goal of the model is to project each image into a finite-dimensional space in a way that would place similar images close to each other, while all the dissimilar images will be located at a distance. To illustrate, suppose there are three images, I_a, I_b, I_c. Images I_a and I_b are two different images of the same object (Panels (a) and (b) of Figure 7), while image I_b and I_c correspond to two completely different objects (Panels (b) and (c) of Figure 7). The goal of the model is to project

each image into a finite-dimensional space such that image I_a and I_b are located close to each other in that space, while dissimilar image I_c is more distant in the vector space.

Figure 7: Illustration of the Image Embedding algorithm

(a) Image I_a *(b)* Image I_b



(c) Image I_c

Notes. This figure illustrates the idea behind the image embedding algorithm. There are three images where image in Panel (a) and Panel (b) are highly similar because they feature same product from different angles. Third image in Panel (c) is also a red chair but differs in style. The goal of the algorithm is to project images into some finite-dimensional space where products from Panels (a) and (b) are closely located in that embedding space, and product in Panel (c) is located further away from them.

At a high level, the algorithm is as follows. The input to the algorithm are the 6 million scraped images, and the ground truth labels, i.e. y-variables that the algorithm tries to predict. The ground truth labels are binary: $y = 1$ if the images belong to the same product (Figure 3) and 0 otherwise. As illustrated above, the algorithm should predict high y for the distinct images of the same product, and vice-versa. As the model trains, it learns to a 512×1 vector representation of each image, i.e. *image embedding*. Next, we can calculate the distances (similarities) between different products using standard distance metrics, such as cosine similarity. we identify closely matched products based on the level of pairwise cosine similarity between products.

Next, we provide a more formal description of the trained Convolutional Neural Network (CNN) for image similarity (see Rusu et al. (2016) and Bell and Bala (2015) for more details). A convolutional neural network is a function f that maps each image I into an embedding position x, given parameters θ: $f_\theta : we \rightarrow x$. One could think of embedding as a numerical vector representation of the image. The goal of CNN is to solve for the parameter vector θ such that the produced embeddings x are such that similar images are placed nearby, whereas dissimilar images are more distant.

In the example above (Figure 7), images are mapped into the embedding vectors x_a, x_b, x_c. If the model is trained well and produces 'good' embeddings, then the correct (positive) pair of images represented by vectors x_a and x_b should be closely located in vector space. Meanwhile the incorrect

(negative) pair of dissimilar images represented by vectors x_b and x_c should be more distant in vector space, i.e. separated by at least a margin m in the embedding space.

The formal objective function for this type of task is called contrastive loss and was proposed by Jia et al. (2021). we minimize the loss function defined as

$$L(\theta) = \sum_{x_a, x_b} \underbrace{L_b(x_a, x_b)}_{\substack{\textit{Penalty if similar images} \\ \textit{are far away}}} + \sum_{x_b, x_c} \underbrace{L_c(x_b, x_c)}_{\substack{\textit{Penalty if dissimilar images} \\ \textit{are nearby}}} \tag{1}$$

where $L_b(x_a, x_b) = ||x_a - x_b||_2^2$ and $L_c(x_b, x_c) = max\{0, m^2 - ||x_b - x_c||_2^2\}$. The objective (loss) function consists of two parts: $L_b(x_a, x_b)$ penalizes a positive pair (x_a, x_b) if the embeddings of the images of the same product are too far apart, and $L_c(x_b, x_c)$ penalizes a negative pair (x_b, x_c) if the images of two different items are closer than the margin m.

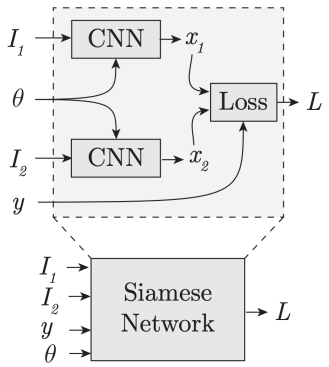
We minimize the objective function in Equation 1 with respect to parameters θ using stochastic gradient descent with momentum (Krizhevsky et al., 2017). Parameters θ govern how well the model learns the representation of the images in a vector space.

$$v^{(t+1)} \leftarrow \mu \cdot v^{(t)} - \alpha \cdot \frac{\partial L}{\partial \theta}(\theta^{(t)}) \tag{2}$$

$$\theta^{(t+1)} \leftarrow \theta^{(t)} + v^{(t+1)} \tag{3}$$

where θ is the parameters to be estimated, v is the momentum sequence, $\mu \in [0, 1)$ is the momentum and $\alpha \in [0, \infty)$ is the learning rate. To efficiently compute the loss function L and the gradient $\frac{\partial L}{\partial \theta}$, we follow Hadsell et al. (2006). Their approach is to construct a siamese network, which is the two copies of the CNN that share the same set of parameters θ. The network takes as an input two images I_1, I_2, θ and the indicator variable for whether the images are a positive pair $(y = 1)$ or a negative pair $(y = 0)$, and outputs the loss function L (Figure 8). The computational efficiency arises from the two copied networks.

Figure 8: Siamese network



Notes. This figure illustrates the architecture of the Deep Neural Network that we used to create image embeddings. Source: Bell and Bala (2015).

As was mentioned above, we use 6 million images collected in ten product categories to train the network. The output of the algorithm is the 512×1 vector representation of each image, which we use to calculate pairwise cosine similarities between products.

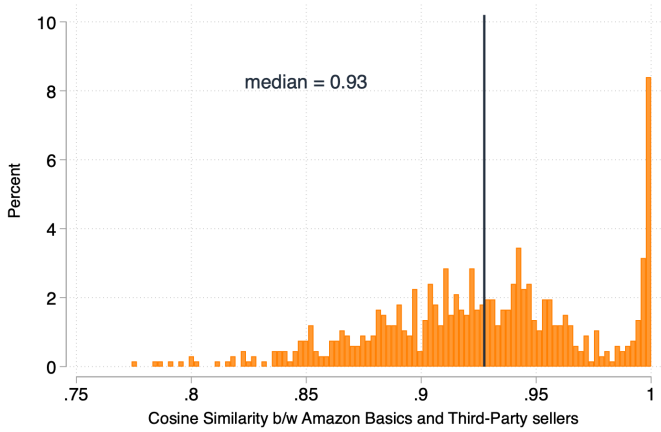
3.2 Extent of product similarity

After training the algorithm and obtaining vector representation of the images, we calculate the pairwise cosine similarities between products and identify the matches of Amazon Basics and third-party seller products.

Figure 10 shows some examples of the case where Amazon Basics products' pictures were highly similar to the pre-existing third-party seller products. Appendix B provides corresponding product page URLs that feature these products. Additional examples of highly similar products in all scraped categories are provided in Figures 25 in the Appendix.

To generalize the results, for each of the 624 Amazon Basics products, we identify the top ten most similar third-party seller products. Figure 9 plots the distribution of the top ten cosine similarities for each Amazon Basics products. we find that the median similarity is 0.93 which means that the products are highly similar. Typically, products with cosine similarity of 0.9 and higher are visually very similar as is also shown in Appendix 25.

Figure 9: Distribution of top cosine similarities



Notes. This figure illustrates the distribution of top ten cosine similarities between each Amazon Basics products and third-party sellers.

Figure 10: Examples of highly similar products. See Appendix B for product page URLs.

(a) Office chairs

$151.53
amazon basics

$79.9
qmei

Cosine similarity: 0.94



(b) Dutch Ovens

$46.28
amazon basics

$422.99
weiwei

Cosine similarity: 1.00



(c) Dining Sets

$57.75
amazon basics

$138.97
yangyang

Cosine similarity: 1.00



(d) End Tables

$39.74
amazon basics

$204.57
jymbk

Cosine similarity: 1.00



(e) Suitcases

$84.36
amazon basics

$134.4
samsonite

Cosine similarity: 0.92



(f) Dog Beds

$37.98
amazon basics

$29.99
kaiwen

Cosine similarity: 0.96



To illustrate the extent to which products are similar, out of 624 Amazon Basics products, 592 had at least one product with cosine similarity of 0.9 and higher. Note that in this part we compare each Amazon Basics products with the *pre-existing* products introduced before Amazon Basics. This evidence suggests that indeed Amazon Basics products have uncanny similarities to the third-party sellers' products. To be cautious, we call this 'potential imitation'.

We show descriptive evidence that Amazon potentially imitates products that have higher sales ranks using the followng specification:

$$cosine\ similarity_j = \alpha + \beta log(sales\ rank)_j + \varepsilon_j \tag{4}$$

where $cosine\ similarity_j$ is the maximum cosine similarity of the third-party seller's product j with an Amazon Basics products, $log(sales\ rank)_j$ is the sales rank of the third-party seller product j in some category. we focus on the results from the office chairs category. Office chairs can be categorized in multiple taxonoour levels: Home Office Desk Chairs, Office Products, Home & Kitchen or a more granular Video Game Chairs. Each product has a sales rank in this taxonomies.

Table 3 shows that there is a negative association between log sales rank of the product and the maximum cosine similarity of the corresponding third-party seller products. Thus, products with higher sales ranks (with 1 being the top) are less likely to have high cosine similarity to the Amazon Basics products. The correlational result is that an increase in sales rank is associated with a decrease of 0.5%-1.7% in maximum cosine similarity. Other research has also found that Amazon tends to focus on products with higher sales ranks (see Zhu and Liu, 2018).

Table 3: Relationship between log sales rank and cosine similarities of the third-party seller products.

	Cosine Similarity within Category...			
	(1)	(2)	(3)	(4)
	Home Office Desk Chairs	Office Products	Video Game Chairs	Home & Kitchen
log(sales rank)	-0.012***	-0.011***	-0.017***	-0.005***
	(0.001)	(0.000)	(0.000)	(0.001)
Intercept	0.516***	0.539***	0.535***	0.457***
	(0.007)	(0.003)	(0.003)	(0.014)
Observations	457,591	1,789,285	241,444	52,465

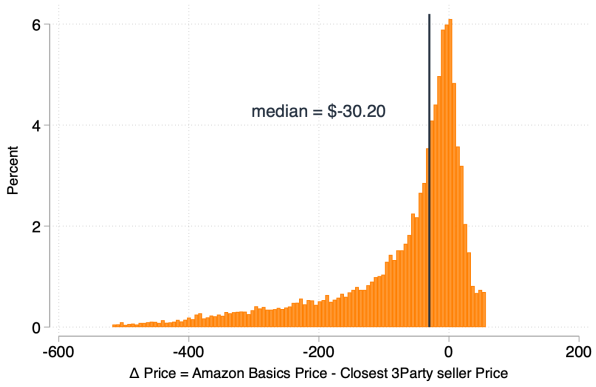
Notes. This table reports the estimation results of Equation 4. Robust standard errors in parentheses. Significance levels: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Also note that the potentially imitated products are not necessarily larger sellers' products. For instance, in Office Chairs category, among top potentially imitated third-party sellers only two are top sellers overall in the category (see Tables 5 and 6 in Appendix C). The rest are smaller sellers' products which should raise a concern, given that smaller sellers cannot retaliate to the potential imitation as they depend on Amazon as a marketplace.

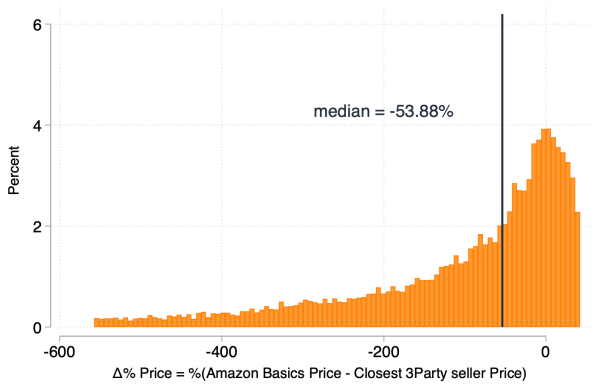
Next, we investigate the price differences between Amazon Basics and potentially imitated third-party sellers' products. we calculate absolute and relative price differences among products that had

Figure 11: Matching results

(a) Distribution of cosine similarities



(b) Price differences



Notes. This figure reports the absolute and relative price differences between Amazon Basics and matched most similar third-party seller products.

cosine similarity of 0.9 and higher. Figure 11a shows the absolute differences in prices and reports that a median imitated Amazon Basics product was priced at $30 less than the matched third-party seller product. Figure 11b shows that the median Amazon Basics product is 53% cheaper than the matched most similar third-party seller product.

To summarize, after matching Amazon Basics with the most similar third-party seller product, we find that the majority of existing Amazon Basics products in the ten scraped categories have at least one highly similar pre-existing third-party seller product. The decision to potentially imitate the products seems to be negatively correlated with the sales rank of the products. Third-party sellers' products with better sales rank (lower sale rank) have 0.5-1.7% higher cosine similarity to the Amazon Basics products manufactured later. Producing private label products is usually cheaper which may explain 53% lower Amazon Basics prices compared to the matched products.

There are two additional observations worth discussing. First, there are heterogeneities across product categories when it comes to the probability of imitating a product. we estimate the following descriptive logit specification:

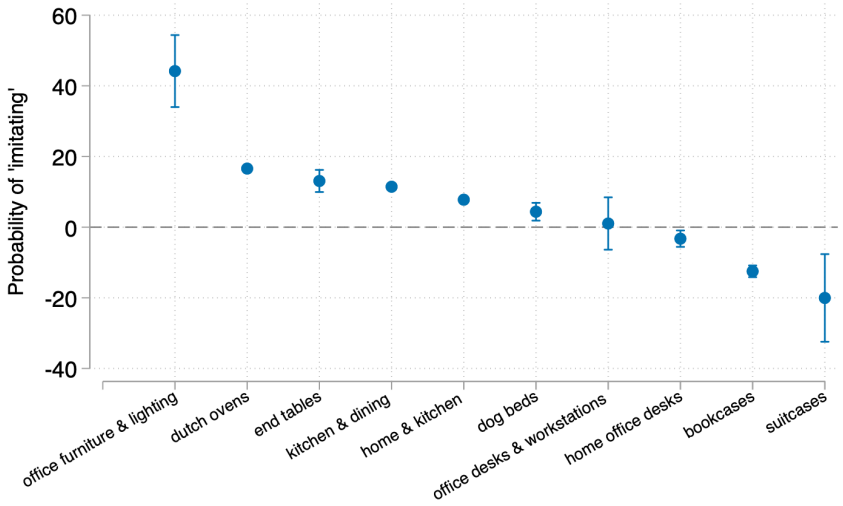
$$P(cosine \geq 0.9)_j = \alpha + \beta log(sales\ rank)_j + \varepsilon_j \tag{5}$$

where $P(cosine \geq 0.9)_j$ is the probability that a third-party product was imitated, i.e. had an Amazon Basics product with cosine similarity of ≥ 0.9, and $log(sales\ rank)_j$ is the log of the bestseller rank of third-party seller product j.

Figure 12 plots the negative β coefficients[10] for each product category. The categories on the x-axis are ordered in the descending order of the probability of imitation. Most categories have a positive or null relationship between the product's bestseller status and the probability of imitation

[10]We plot negative β's because higher sales rank number is worse.

Figure 12: Probability of potential imitation by product category



Notes. This figure illustrates the probability of imitation by different product categories.

with the exception of bookcases and suitcases that have negative correlation. This could be explained by the fact that in suitcases, there are some major brands the products of which may be harder to imitate, such as Samsonite. Overall, it seems that Amazon follows different rules in different categories when it comes to potential product imitation.

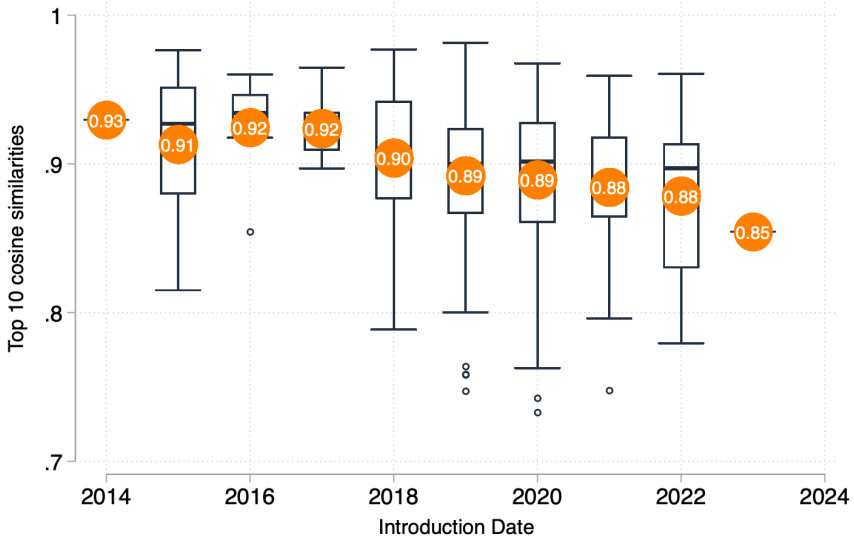
Next observation worth noting is that over the years, Amazon seems to be doing less imitation or at least it tries to explore other product ideas. Figure 13 plots the distribution of top cosine similarities over the years when Amazon Basics product was introduced. There seems to be a decreasing trend in mean cosine similarity levels of the products. For instance, in 2015, the mean similarity was 0.91 while in 2022 it is 0.88 and there are more products that are below 0.9 cosine similarity. This result suggests that over time Amazon started producing more of unique products.

To summarize, this subsection presented results that suggest systematic potential product imitation across ten scraped product categories. Both large and small sellers' product can be replicated and Amazon is able to set lower prices on the private label products. Next section explores the effect of Amazon Basics entry on third-party sellers.

3.3 Matching: Effect of Amazon Basics entry on third-party sellers

In this subsection, we investigate how the entry of Amazon Basics products affected third-party sellers products. To do that we proceed in several steps. First, we take the set of potentially imitated third-party seller products identified in the previous subsection. Next, in a similar fashion as described in Section 3, we create pairwise cosine similarities among all third-party seller products. Next, for each potentially imitated third-party seller product we identify most similar products among other

Figure 13: Cosine similarities over time



Notes. This figure illustrates the distribution of cosine similarities between Amazon Basics and most similar third-party seller products over time.

non-imitated third-party seller products ($cosine\ similarity \geq 0.9$). Thus, each potentially imitated product is allocated into a group together with the matched non-imitated product. Using scraped panel data of prices, sales ranks and product availability, we construct a panel of potentially imitated and non-imitated products with the group indicator. Next, we create an indicator variable that equals one after Amazon entered, i.e. imitated third-party seller product, and equals zero before Amazon entry. we use two-way fixed effect specification to estimate a quasi-experimental effect of Amazon entry on the resulting sales rank of the products, log price and the probability of the product being discontinued. Note that in the data we observe when a product gets discontinued because the scraped product pages have 'Currently Unavailable'. we estimate the following two-way fixed effect specification:

$$y_{jt} = \beta after_{jt} + \alpha_j + \delta_{g(j),t} + \varepsilon_{jt} \tag{6}$$

where $after_{jt}$ is the indicator for the after Amazon Basics entry period, α_j is the product fixed effect, $\delta_{g(j),t}$ is the group-year×week fixed effect.

Table 4 shows the results of the estimation. First, we find that Amazon entry moves potentially imitated third-party sellers' products down in the sales ranks by 4-4.5% (recall that higher rank is worse). The difference between Columns (1) and (2) is the granularity of the product category taxonomy. Column (1) corresponds to a more aggregate category, i.e. Office Furniture, and Column (2) corresponds to a more granular taxonomy, i.e. Office Chairs. The results are similar in both columns.

Next, Column (3) shows that after Amazon entry potentially imitated third-party sellers decrease their retail prices by 1.1% compared to the other sellers. Moreover, they are more likely to discontinue

Table 4: Effect of Amazon entry on third-party sellers

	(1) Log(sales rank)	(2) Log(sales rank)-Granular	(3) Log(Price)	(4) Discontinued
After Amazon entry	0.045***	0.040***	-0.011***	0.003**
	(0.010)	(0.009)	(0.001)	(0.001)
Intercept	13.337***	7.470***	5.126***	0.001**
	(0.005)	(0.005)	(0.000)	(0.001)
Product FE	✓	✓	✓	✓
Group-Year× Week	✓	✓	✓	✓
Clusters	1,831	1,831	4,141	4,141
N	108,446	108,444	292,897	292,897

Notes. This table reports the effect of Amazon Basics entry on third-party sellers. The comparison groups are potentially imitated third-party seller products and non-imitated third-party sellers that are most similar to the imitated ones. we used visual similarity measure. Clustered standard errors in parentheses. Significance levels: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

the products with the effect being quite large 0.003 versus the intercept of 0.001.

Thus, the entry of Amazon Basics did hurt potentially imitated third-party sellers, both through the worse sales rank and higher probability of discontinuing the products.

The findings that the entry of private labels hurt third-party sellers on Amazon may not be any different than the standard offline retailer cases. However, the main regulatory concern is that Amazon has unprecedented data access and scale to create these products. Currently, private label business is not very large on Amazon. However, there could be concerns over potential growth of the private label business and its consequences on the marketplace. Since regulators, such as FTC, consider different types of data regulation in this area, it is important to explore which regulatory measures will be most effective.

4 Data Regulation

In this section, first, we discuss the current duration-based regulation that was proposed and settled on in the European Union. In particular, Amazon has agreed to not engage in anti-competitive practices up to five years.[11] It is important from the regulatory perspective to assess the efficiency of the duration-based regulation and understand which types of data regulation could be effective.

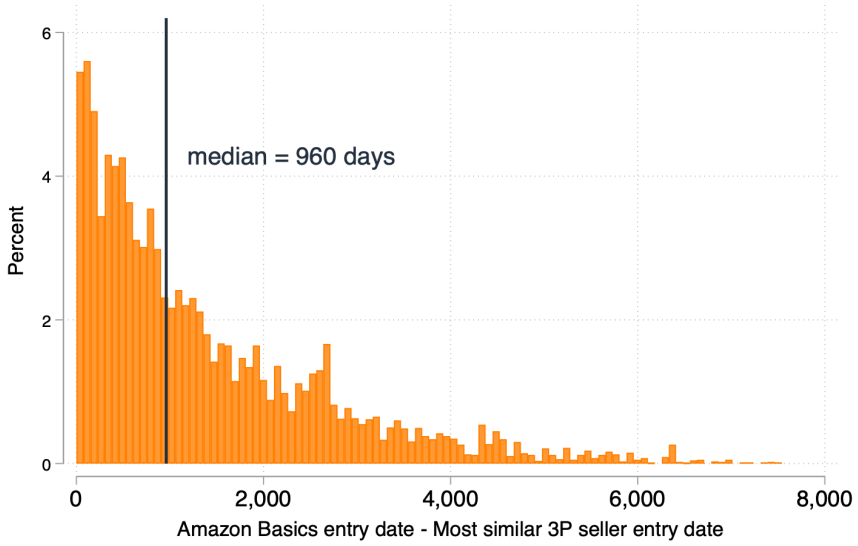
4.1 Current Regulation

First, we explore the potential effectiveness of the duration-based regulation. To do that, we calculate the number of days that Amazon takes to potentially imitate the third-party seller products. This measure is calculated using the introduction dates available for each product. Figure 14 shows the corresponding distribution of the number of days. Note that we compare Amazon Basics products to most similar pre-existing products. Figure 14 shows that a median number of days equal 960 days

[11] For Prime-related concerns, Amazon committed to not engage in any anti-competitive practices for seven years, and for all other charges, they committed to five years - Reuters link.

(or 2.6 years) with the maximum of 8,000 days.

Figure 14: Timeline of Days Between Amazon Basics Entry and Third-Party Seller Entry



Notes. This figure illustrates the number of days after which a third-party seller product gets potentially imitated by Amazon.

Next, we calculate how long it takes to potentially imitate different types of sellers. To do that, we construct a weekly estimated sales measure from the sales rank data using the approach suggested in He and Hollenbeck (2020). Their approach is to infer the weekly estimated sales as follows:

$$\hat{Q} = exp\left(\frac{\hat{c} - log(Rank - 1)}{\hat{\theta}}\right) \tag{7}$$

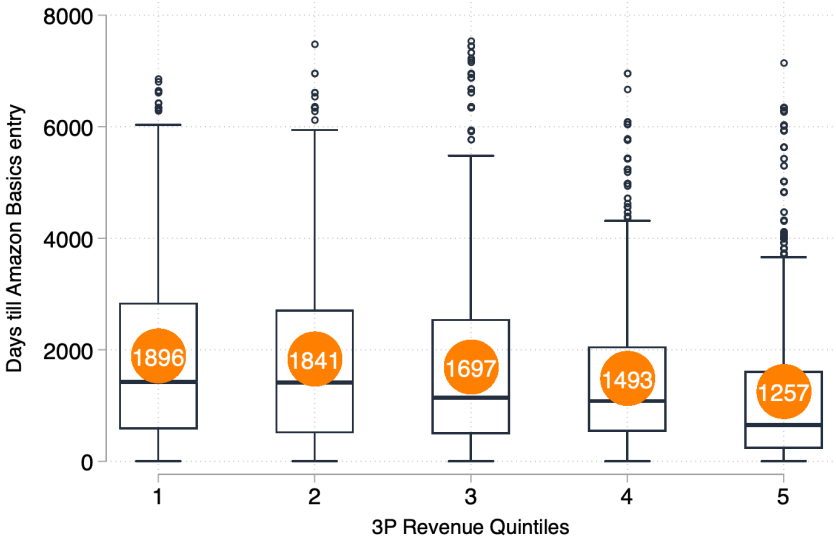
where $Rank$ is the bestseller rank that we scraped from the website, Q is the estimated weekly sales, $\hat{c}$ and $\hat{\theta}$ are parameters that He and Hollenbeck (2020) estimated using true sales data. We use their estimated coefficients, which differ by product category.

Next, we split sellers into quintiles based on the estimated weekly sales. For each quintile, we plot the distribution of cosine similarities in Figure 15. The median number of days to imitate a small sellers' (first quintile) product is 1,896 (5 years) days and the same measure is equal to 1,257 (3.4 years) for large sellers. Moreover, there is a decreasing trend in the number of days needed to potentially imitate the product as one moves from quintile one to five.

There are several important implications of these results. First, the fact that it takes Amazon 2.6 years to imitate a median product, and 5 years to imitate a small seller product could mean that Amazon does not have any private information about product performance.

Had Amazon had any private information about particular sellers, it would have replicated the products faster. Thus, Amazon may have been using the data that is close to being publicly available and may just benefit from more efficient data access. From the regulatory perspective, there may still be a concern because of the data scale even if Amazon does not use any private information.

Figure 15: Timeline of Amazon Basics entry by third-party seller revenue



Notes. This figure illustrates the relationship between the days till Amazon Basics entry by the third-party seller revenue quintiles. we calculate the estimated weekly revenue from the sales rank using the approach from He and Hollenbeck (2020).

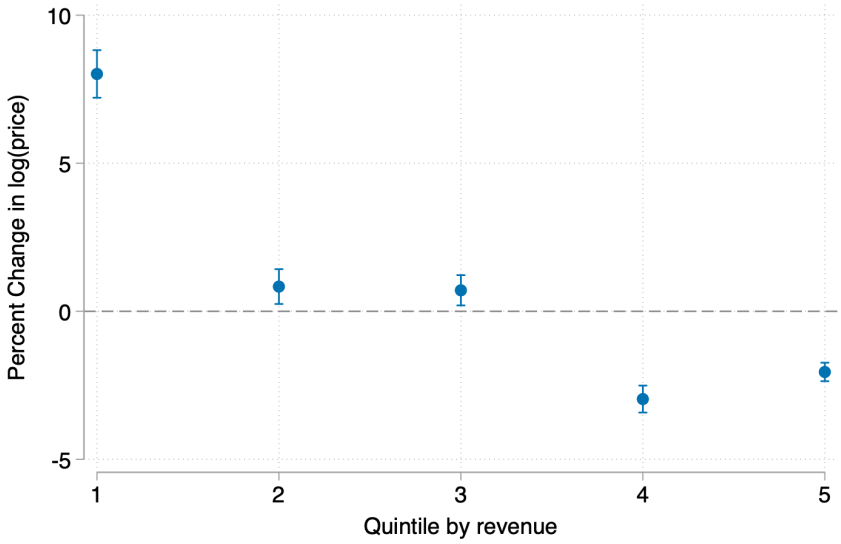
In the European Union, such regulatory concern has been settled by requiring Amazon to not engage in anti-competitive practices in private label business during five years. However, simply by looking at the days it takes Amazon to replicate the products, it becomes clear that five-year data restrictions may protect larger sellers but do not product small sellers. In fact, it takes Amazon five years to collect necessary data to replicate a smaller seller's product, which is exactly the duration of the regulation. Thus, this type of duration-based regulation will likely not be effective to protect small sellers. Next, we explore how small sellers may respond to the Amazon Basics entry and discuss a possibility of alternative data regulation.

4.2 Other Regulation

We follow the same approach as described above and split sellers into revenue quintiles based on the average weekly sales. Next, we use the two-way fixed effect specification within each sales quintile, and investigate the differences in pricing strategies after Amazon entry. The specification is the same as in Equation 6. Figure 16 plots the β coefficients.

We find that larger sellers (quintiles four and five) can afford to decrease their prices after the entry of Amazon Basics products. However, smaller sellers are increasing their prices by 0.2-7.8%. One mechanism the latter could happen is if smaller sellers signal their quality by setting a higher price. Alternatively, they may think that there is a selection in the set of consumers who purchase them, which is why they want to skim the cream from the high willingness-to-pay consumers. In any case, larger sellers can decrease their prices, thereby attracting more demand, while smaller sellers do not do that.

Figure 16: % change in price by seller revenue quintiles



Notes. This figure illustrates the relationship between seller sales quintile and the percent change in log price. O calculate the estimated weekly revenue from the sales rank using the approach from He and Hollenbeck (2020).

The fact that smaller sellers increase prices may also hurt consumers. Therefore, we think that one type of regulation may target smaller sellers by protecting their products from being replicated. In that case, they will not face a fierce price competition from Amazon Basics products that can afford to set 53% lower prices on average. The regulators could use the methodology proposed in this paper to identify the similarities between products, and protect smaller sellers.

5 Conclusion

Despite growing regulatory concerns over dual role of digital platforms as both the marketplace and the seller, there is little understanding of the extent of product imitation and the effectiveness of the potential data regulation policies. In this paper, first, we train Deep Learning models to create visual and textual similarity measures between the products. we show that out of 624 Amazon Basics products that exist in ten scraped categories, 592 products exhibit high cosine similarity to the pre-existing third-party seller products. This suggests that there is indeed systematic evidence of potential product imitation. Next, we show that Amazon can set $30 or 53% lower prices on the private label version of the potentially imitated products. As a result of Amazon Basics entry, third-party sellers that were potentially imitated experience worse sales ranks and higher probability of discontinuing the product compared to the matched non-imitated third-party seller products. These results raise concerns regarding the impact of Amazon private labels on third-party sellers and questions how should the dual platforms be regulated.

We evaluate the current duration-based data regulation that reached settlement between Amazon

and the European Union. we show that the duration-based regulation, where Amazon commits not to engage in anti-competitive practices for five years, may not be an effective solution. The main reason is that it takes Amazon 2.6 years to replicate a median product, and 5 years to replicate a small sellers' products. Thus, the duration-based regulation may protect larger sellers that are faster to replicate, but it does not protect small sellers.

Thus, there may be a need for alternative types of regulation that would specifically protect small sellers. This is also exacerbated by the fact that larger sellers can afford to set lower prices after Amazon entry, while smaller sellers increase their prices. The increased prices may hurt both sellers and consumers, and from the regulation perspective, small seller-targeted regulation can be more desirable.

References

Bell, Sean and Kavita Bala, "Learning visual similarity for product design with convolutional neural networks," *ACM transactions on graphics (TOG)*, 2015, *34* (4), 1–10. (Cited on page(s) 2, 11, 12)

Dubé, Jean-Pierre, "Amazon Private Brands: Self-Preferencing vs Traditional Retailing," *Available at SSRN 4205988*, 2022. (Cited on page(s) 5)

Farrell, Joseph and Michael L Katz, "Innovation, rent extraction, and integration in systems markets," *The journal of industrial economics*, 2000, *48* (4), 413–432. (Cited on page(s) 4)

Farronato, Chiara, Andrey Fradkin, and Alexander MacKay, "Self-Preferencing at Amazon: Evidence from Search Rankings," Working Paper 30894, National Bureau of Economic Research January 2023. (Cited on page(s) 4)

Hagiu, Andrei and Daniel Spulber, "First-Party Content and Coordination in Two-Sided Markets," *Management Science*, 2013, *59* (4), 933–949. (Cited on page(s) 4)

_ , Tat-How Teh, and Julian Wright, "Should platforms be allowed to sell on their own marketplaces?," *The RAND Journal of Economics*, 2022, *53* (2), 297–327. (Cited on page(s) 4)

He, Sherry and Brett Hollenbeck, "Sales and rank on amazon. com," *Available at SSRN 3728281*, 2020. (Cited on page(s) 7, 20, 21, 22)

Jia, Chao, Yinfei Yang, Ye Xia, Yi-Ting Chen, Zarana Parekh, Hieu Pham, Quoc Le, Yun-Hsuan Sung, Zhen Li, and Tom Duerig, "Scaling up visual and vision-language representation learning with noisy text supervision," in "International Conference on Machine Learning" PMLR 2021, pp. 4904–4916. (Cited on page(s) 12)

Jiang, Baojun, Kinshuk Jerath, and Kannan Srinivasan, "Firm strategies in the "mid tail" of platform-based retailing," *Marketing Science*, 2011, *30* (5), 757–775. (Cited on page(s) 4)

Khan, Lina M, "Amazon's antitrust paradox," 2017. (Cited on page(s) 5)

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, 2017, *60* (6), 84–90. (Cited on page(s) 12)

Lee, Kwok Hao and Leon Musolff, "Entry into two-sided markets shaped by platform-guided search," Technical Report, mimeo, Princeton University 2021. (Cited on page(s) 4)

Madsen, Erik and Nikhil Vellodi, "Insider imitation," *Available at SSRN 3832712*, 2022. (Cited on page(s) 4)

Rusu, Andrei A, Neil C Rabinowitz, Guillaume Desjardins, Hubert Soyer, James Kirkpatrick, Koray Kavukcuoglu, Razvan Pascanu, and Raia Hadsell, "Progressive neural networks," *arXiv preprint arXiv:1606.04671*, 2016. (Cited on page(s) 11)

Yang, Yinfei, Daniel Cer, Amin Ahmad, Mandy Guo, Jax Law, Noah Constant, Gustavo Hernandez Abrego, Steve Yuan, Chris Tar, Yun-Hsuan Sung, Brian Strope, and Ray Kurzweil, "Multilingual Universal Sentence Encoder for Semantic Retrieval," 2019. (Cited on page(s) 3, 26)

Zhu, Feng and Qihong Liu, "Competing with complementors: An empirical look at Amazon. com," *Strategic management journal*, 2018, *39* (10), 2618–2642. (Cited on page(s) 4, 15)

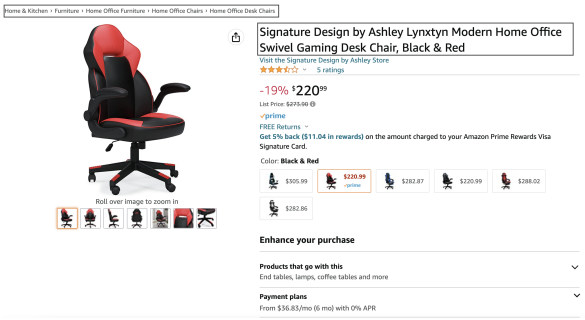
A Building product similarity measures

A.1 Textual Embeddings

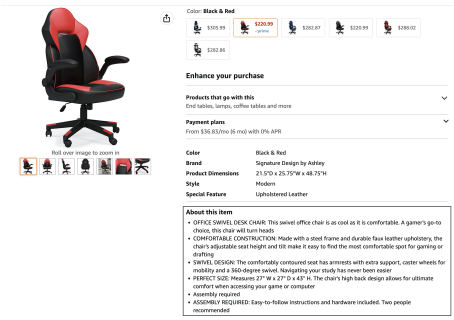
From the Amazon website, we scraped the category taxonoour of the product (e.g. Home & Office Desk Chairs), product name (Panel (a) of Figure 17) and detailed itemized product descriptions (Panel (b) of Figure 17). we use this information to create a lower-dimensional vector of the product information. The goal is to represent the unstructured textual information in an $N \times 1$ vector form to be able to calculate the cosine similarities between different products.

Figure 17: Product descriptions

(a) Category and product names



(b) Itemized product description

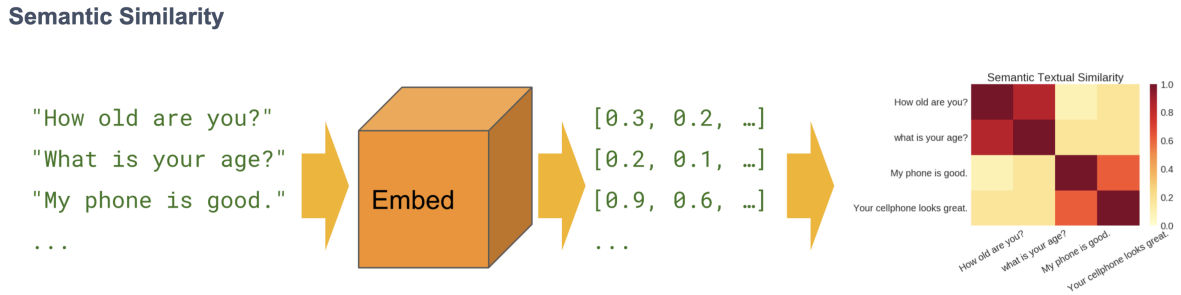


Notes: The left-hand panel shows which part of the product page was scraped to identify the product descriptions and category taxonomy. The right-hand side boxed part shows the scraped itemized list of detailed product descriptions.

To achieve that, we embed the raw text description into a single semantic space using Universal Sentence Encoder developed by Google (Yang et al. (2019). The encoder model embeds raw English text into the high-dimensional vector (512×1) that can be then used for text classification, semantic similarity, clustering and other natural language tasks. The benefit of using the Universal Sentence Encoder model compared to wide-spread word embeddings, is that encoders model the meaning of the word sequences than just individual words and are trained on multiple tasks. Figure 18 shows an example of how the semantic similarity task is performed: the goal is to understand how similar two pieces of text are. In this example, the questions "How old are you?" and "What is your age?" have high similarity, while sentence "our phone is good" is dissimilar. Given good performance of this model that was used in multiple applications, we proceed using it to construct the similarity matrix between Amazon Basics and Third-Party Sellers' products.

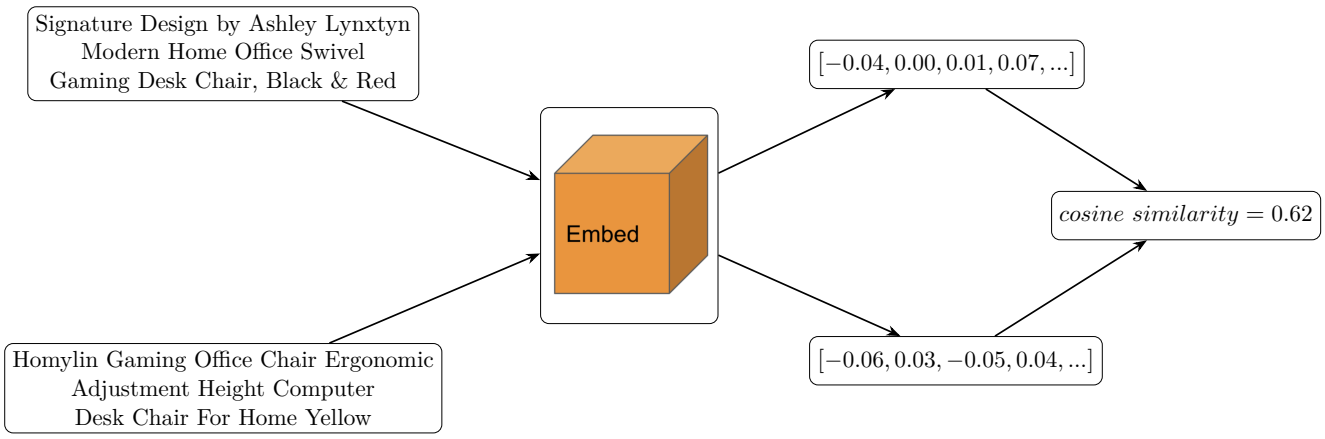
Figure 19 shows an example of how the textual embedding works. The input to the model are the product descriptions. The embedding model outputs 512×1 vector for each product's description. This allows me to calculate the cosine similarities between different products. The example uses two gaming chairs that are sold on Amazon. Both descriptions have words such as 'Home', 'Office', 'Desk Chair', 'Gaming'. According to the generated textual embeddings, the cosine similarity between the product descriptions is 0.62.

Figure 18: Example of Universal Sentence Encoder



Source: Google Universal Sentence Encoder

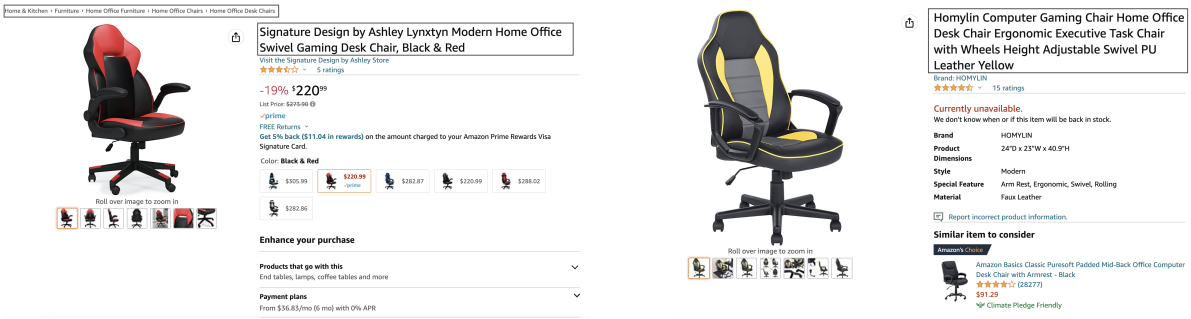
Figure 19: Example of how textual embeddings work



Notes: The input to the model are the product descriptions. The model outputs a 512×1 vector of numerical text representation. This allows to calculate the cosine similarity between these vectors.

Figure 20 shows the corresponding pictures of the two products. Indeed, chairs do look somewhat similar but differ in color and style a bit, which justifies the cosine similarity of 0.62.

Figure 20: Example of textual embedding model



Notes: The left-hand panel shows which part of the product page was scraped to identify the product descriptions and category taxonomy. The right-hand side boxed part shows the scraped itemized list of detailed product descriptions.

Figure 21: Top cosine similarity matches in Office Chairs

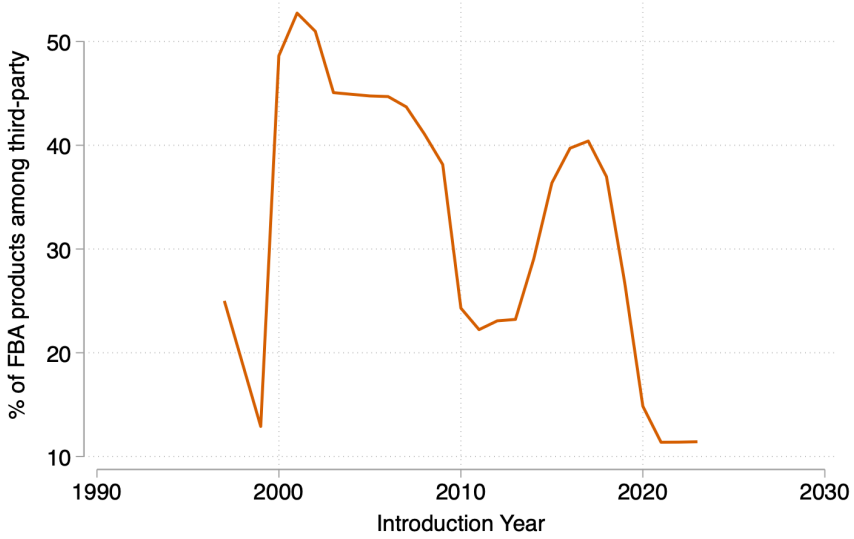


B URLs of highly similar products

- Dutch Oven Third-Party Seller
- Dutch Oven Amazon Basics
- Dog Mat Third-Party Seller
- Dog Mat Amazon Basics
- Dinnerware Set Third-Party Seller
- Dinnerware Set Amazon Basics
- End Table Third-Party Seller
- End Table Amazon Basics
- Suitcase Third-Party Seller
- Suitcase Amazon Basics
- Office Chair Third-Party Seller
- Office Chair Amazon Basics

C Additional Descriptive Statistics

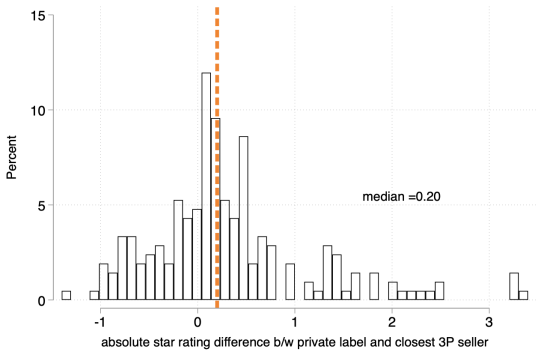
Figure 22: Share of Fulfilled by Amazon products among third-party seller products



Notes. This figure illustrates the share of Fulfilled By Amazon products among third-party sellers in ten scraped product categories over multiple years.

Figure 23: Star rating differences between private label and most similar 3P products

(a) star rating differences



(b) review count differences

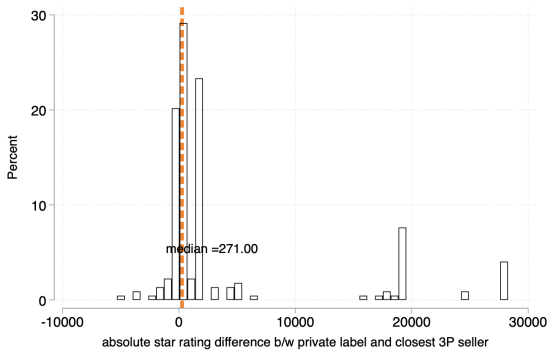
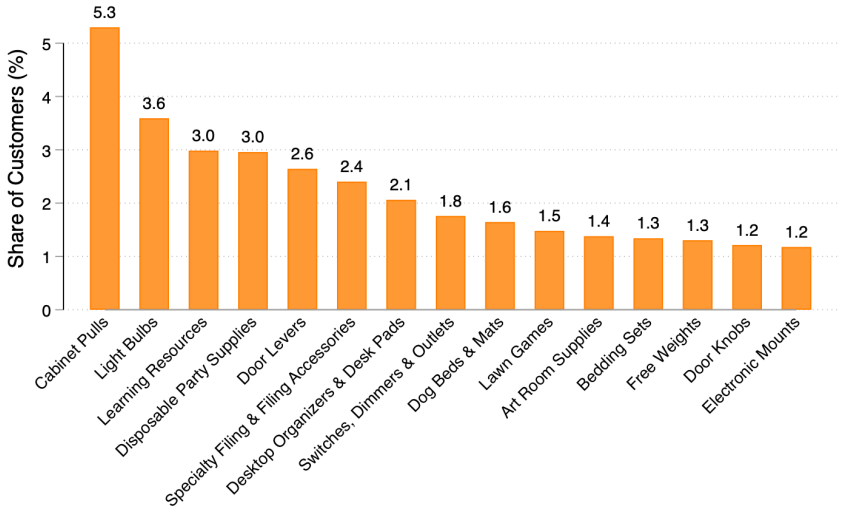


Figure 24: Top categories carrying Amazon Basics products



Notes. This figure illustrates top categories that feature Amazon Basics products (overall). The y-axis is the share of all Amazon Basics products that exist on the website attributed to a category.

Figure 25: Office Chairs

$156.99 amazon basics	$199.99 ch-air	$137.16 amazon basics	$640.85 whh	$72.0 amazon basics	$99.99 fat catalog	$82.49 amazon basics	$83.88 orveay
Cosine similarity: 1.00		Cosine similarity: 1.00		Cosine similarity: 0.97		Cosine similarity: 1.00	
							
$143.5 amazon basics	$185.32 edgemod	$132.0 amazon basics	$640.85 whh	$131.0 amazon basics	$145.99 orveay	$76.99 amazon basics	$81.15 sea
Cosine similarity: 0.94		Cosine similarity: 0.97		Cosine similarity: 1.00		Cosine similarity: 1.00	
							
$75.99 amazon basics	$69.99 orveay	$151.53 amazon basics	$79.9 qmei	$134.26 amazon basics	$555.7 lcpet	$138.48 amazon basics	$199.99 prettyshop4246
Cosine similarity: 1.00		Cosine similarity: 0.94		Cosine similarity: 0.99		Cosine similarity: 0.95	
							

Figure 26: Dog Beds

$30.08 amazon basics	$38.79 pet tek	$20.0 amazon basics	$46.99 kmmkk	$21.92 amazon basics	$46.99 kmmkk	$37.98 amazon basics	$29.99 kaiwen
Cosine similarity: 0.97		Cosine similarity: 0.98		Cosine similarity: 0.99		Cosine similarity: 0.96	
							
$32.82 amazon basics	$27.99 uheng	$24.8 amazon basics	$27.99 midwest homes for pets	$33.09 amazon basics	$34.99 jema	$38.14 amazon basics	$37.39 joejoy
Cosine similarity: 0.96		Cosine similarity: 0.95		Cosine similarity: 0.96		Cosine similarity: 0.96	
							
$71.24 amazon basics	$129.99 imfin	$21.99 amazon basics	$29.99 petsbark	$24.99 amazon basics	$102.81 deprq	$84.98 amazon basics	$141.95 snoozer
Cosine similarity: 0.99		Cosine similarity: 0.96		Cosine similarity: 0.99		Cosine similarity: 0.95	
							

Figure 27: Desks

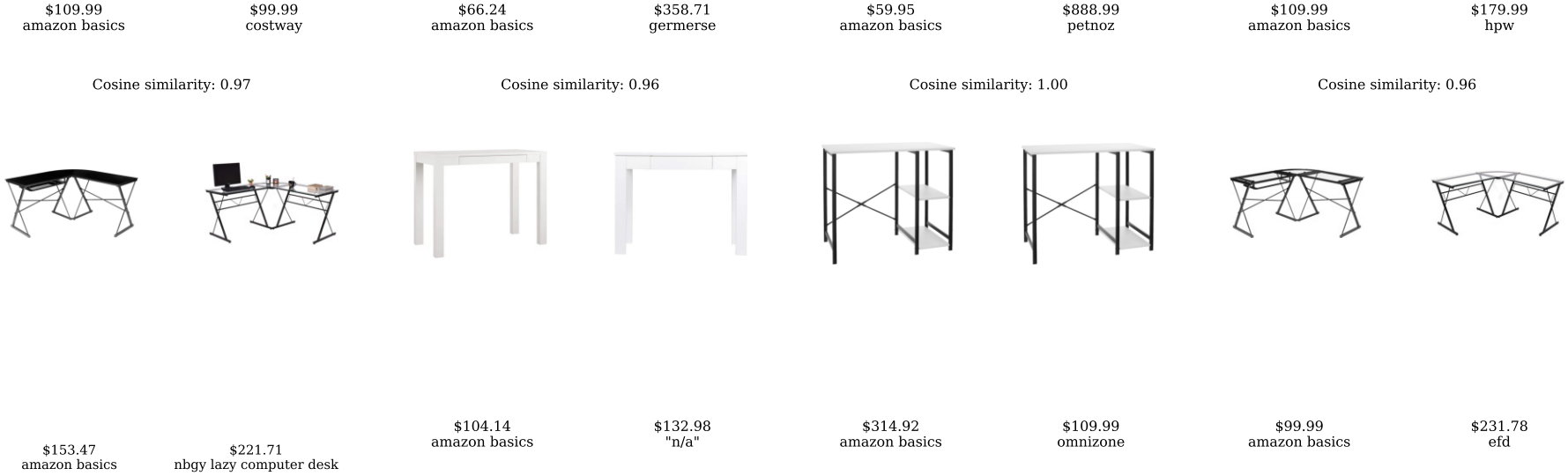
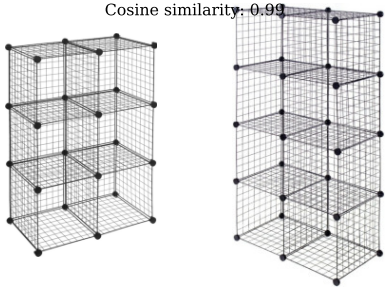


Figure 28: Bookcases

$35.17
amazon basics

$125.99
king77777

Cosine similarity: 0.99



$128.73
amazon basics

$193.14
benjara

Cosine similarity: 0.94



$50.38
amazon basics

$117.0
progiga select

Cosine similarity: 0.97



$131.8
amazon basics

$296.65
wood designs

Cosine similarity: 0.95



$98.35
amazon basics

$91.38
household essentials

Cosine similarity: 0.93



$52.49
amazon basics

$109.0
huippy

Cosine similarity: 1.00



$50.91
amazon basics

$185.99
target marketing systems

Cosine similarity: 0.93



$56.99
amazon basics

$79.99
fivegiven

Cosine similarity: 0.94



$49.87
amazon basics

$237.58
delacora

Cosine similarity: 0.96



$98.35
amazon basics

$472.89
tenozek

Cosine similarity: 0.91



$58.49
amazon basics

$67.7
furinno

Cosine similarity: 0.91



$19.35
amazon basics

$113.51
mjwlgs

Cosine similarity: 0.92

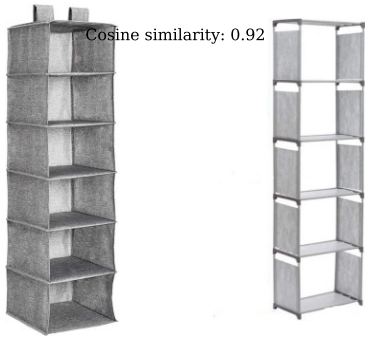


Figure 29: End Tables



Figure 30: Dutch Ovens

$46.28
amazon basics

$422.99
weiwei

Cosine similarity: 1.00



$42.64
amazon basics

$422.99
weiwei

Cosine similarity: 0.99



$63.0
amazon basics

$299.99
joenotjo

Cosine similarity: 0.95



$51.99
amazon basics

$299.99
joenotjo

Cosine similarity: 1.00



$67.31
amazon basics

$299.99
joenotjo

Cosine similarity: 0.96



$50.8
amazon basics

$299.99
joenotjo

Cosine similarity: 0.98



$49.99
amazon basics

$125.37
2 for 1 clothes

Cosine similarity: 0.96



$49.69
amazon basics

$79.46
calphalon

Cosine similarity: 0.95



$59.27
amazon basics

$524.95
le creuset

Cosine similarity: 0.92



$59.27
amazon basics

$166.04
junsg

Cosine similarity: 0.91



$63.53
amazon basics

$230.4
charon

Cosine similarity: 0.95



$52.75
amazon basics

$158.61
pentole agnelli

Cosine similarity: 0.95



Figure 31: Dinnerware Sets

$43.99
amazon basics

$36.99
super deal

Cosine similarity: 0.96



$24.41
amazon basics

$99.99
ylee

Cosine similarity: 0.94



$38.01
amazon basics

$42.66
corelle

Cosine similarity: 0.94



$24.74
amazon basics

$330.0
le cadeaux

Cosine similarity: 0.95



$37.99
amazon basics

$209.9
nicole home collection

Cosine similarity: 0.94



$57.75
amazon basics

$138.97
yangyang

Cosine similarity: 1.00



$22.55
amazon basics

$49.99
pfaltzgraff

Cosine similarity: 0.94



$54.0
amazon basics

$60.47
sea natey nates

Cosine similarity: 0.98



$43.99
amazon basics

$51.99
chenmumu

Cosine similarity: 0.99



$85.92
amazon basics

$119.52
tudor england

Cosine similarity: 0.97



$56.99
amazon basics

$52.26
sea natey nates

Cosine similarity: 0.98



$52.49
amazon basics

$75.0
better homes & gardens

Cosine similarity: 1.00



Figure 32: Suitcases



D Additional Results for Office Chairs

Table 5 shows the distribution of the bestseller manufacturers in the Office Chairs category. Table 6 shows the distribution of top manufacturers that were potentially imitated in Office Chairs category. The orange color highlights that the replicated manufacturers contain only two bestseller manufacturers, while the rest are smaller sellers.

Table 5: Bestseller manufacturers

Manufacturer	# of products	price
Steelcase	5290	1369.57
Annery	5164	35.63
Office Chair	2920	380.20
Wxf	2616	526.82
Ki	2583	796.81
Top Medical	2295	341.80
Flash Furniture	1906	229.66
Thomas Ecom	1670	379.94
Milanhome	1389	547.94
Office Star	1388	343.17
Fubas-Office Chair	1336	316.56
Ch-Air	1323	411.52
File Cabinets	1124	730.39
Household Products	1112	774.33
Herman Miller	1027	1141.28

Table 6: Most similar manufacturers

Manufacturer	# of products	price
Flash Furniture	65	182.61
Office Star	59	315.53
Scranton & Co	35	196.11
StarSun Depot	34	173.00
Delacora	31	182.19
PRO&Family	31	276.85
KLS14	29	379.12
iHome Studio	29	220.68
F&F Furniture Group	24	321.51
A Line Furniture	23	263.12
our Friendly Office	22	253.31
OFM	22	220.13
Parkside	21	183.99
Lorell	20	277.98
Offex	17	182.21