Are Exposure Bucks Worth It? Evidence from Instagram*

Andrew T. Ching

Johns Hopkins University

Carey Business School

Email: andrew.ching@jhu.edu

Daniel Goetz

University of Toronto

Rotman School of Management

Email: daniel.goetz@utoronto.ca

July 28, 2025

Abstract

Using Instagram's recently introduced collaborative post feature, which lets brands credit creators in shared posts seen by the brand's followers, this study quantifies the value of exposure to a brand's audience for small creators. We collect a new dataset which consists of all collaborative posts between 260 large beauty/clothing brands and small creators, as well as a daily panel of follower counts for creators. Leveraging the high frequency of the panel, we identify changes in creators' follower counts after the collaborative post in a difference-in-differences analysis. We address creator and post content selection through propensity score matching, using creator characteristics and LLM-derived visual content embeddings to match treatment and control posts. Our estimates suggest that exposure to a brand's audience through a collaborative post increases a creator's cumulative followers by 0.47%, or roughly one week's worth of growth. Although larger brand audiences lead to more growth all else equal, larger brands tend to have less engaged followers and lower quality collaborative posts. This leads to stable treatment effects from collaboration with brands that have between 0.7m and 8.6m followers. The top 10% of collaborations average 2.1% follower growth for the creator. Our average conversion rate from brand followers to creator followers is similar to the effect of online display ads estimated in the literature.

Keywords: Collaboration, Exposure Benefit, Content Creation, Creator Economy, Influencer Marketing, Owned Social Media, Large Language Models, Image Processing

*This work was funded by a grant from the TD Management and Data Analytics Lab at Rotman School of Management, and a grant from Johns Hopkins Carey Business School.

1 Introduction

> In exchange for your artwork, we will publish it on the birthday of the player assigned to you, with a mention on Instagram (8,800,000 followers) [. . .] We will tag you in the photo itself and in the caption.
>
> *– Solicitation of artwork by AFC Ajax, the Netherlands' largest football club, Sept. 2024*

> Exposure is something you die from, not something you get paid with.
>
> *– User ZombieButch, post on reddit.com/r/ArtistLounge, Sept. 2023*

For brands, sourcing visual content from their users has become increasingly important (Forbes, 2023). When adding media to webpages, populating social media feeds, and crafting advertisements, many brands now use user-generated content (UGC) supplied by independent creators (Influee, 2025; Social Media Strategies Summit, 2025). A key question for brands is how to incentivize such contributions; a common solution is to offer exposure to their social media audience as a reward, i.e., so-called "exposure bucks."[1] For example, in the leading quote, a football club asks for club-related art for their Instagram page, and in exchange, the artist will be visibly credited to the millions of followers of the football club. Small brands routinely offer exposure as part of creators' compensation;[2] however, even for large brands such as Apple and Nike, offering visibility for the authors of contributed content is typical.[3]

Despite brands' use of social media exposure to motivate creative contributions, very little is known about the degree to which exposure can actually benefit creators. In particular, creators tend to have mixed opinions on the value of exposure, with some referring to the practice disparagingly (Somerfield, 2023).[4] Perhaps this is not surprising because it is difficult to gauge the benefits. On the one hand, brands have large reach on social media, with fashion and beauty

[1]The practice is also common outside of social media. In print media, see , "The Paradox of Art as Work", *The New York Times*, May 2014, "Magazines that once paid by the word and employed healthy rosters of staff writers are now recruiting readers to contribute articles, offering to compensate them with exposure and 'prestige." In a high-profile example from entertainment, see "No, Performers Don't Get Paid for the Super Bowl Halftime Show — Here's Why," *People Magazine*, April 2024.

[2]See "27 Brilliant User-Generated Content Examples," (Convertcart, 2024) and "10 User Generated Content Campaigns that Actually Worked (Taylor, 2023).

[3]See "Connect Deeper, Sell Smarter: 7 UGC Campaigns That Transformed Brand-Customer Relationships", *PepperContent*, January 1, 2024.

[4]For positive creator impressions of being credited by brands, see The State of User-Generated Video, 2024-05, "For exposure and awareness, it's good to get your name around and get your brand across." For negative impressions, see https://www.helloraderco.com/sacrificing-time-money-exposure-bucks/, accessed April 20, 2025, and https://amandajspedding.com/2019/05/27/exposing-exposure-bucks-hint-its-not-a-real-thing/, accessed April 20, 2025. See (LinkedIn Contributors, 2023) for social media managers' positive opinions of exposure.

brands often followed by millions or tens of millions of consumers, which may present a rare opportunity for a small creator to get noticed and build their audience.[5] On the other hand, brands' large audiences may not be very engaged or attentive (Wies et al., 2023), and the content made for a brand may not be a good representation of the creator's typical content (Arsel et al., 2024; Cheng and Zhang, 2025).

In this paper, we aim to empirically quantify the benefit to creators of exposure to a brand's audience. To do so, we need to overcome four obstacles. First, it is challenging to observe a large number of credited contributions by small creators for large brands, and to also be able to measure their benefit to creators. Second, there is a selection problem: creators may be chosen for contributions because of their rising popularity, which would be confounded with exposure. Third, the creator's contributed content could be endogenous—for instance, because the brand may have some influence over it, or the creator may put in more effort. Finally, the treatment effect could exhibit complex dynamics.

To address the first obstacle, we take advantage of the recent collaborative post feature introduced by Instagram in late 2021. This feature allows multiple accounts to jointly collaborate on a post, which will appear to followers of all collaborating accounts simultaneously—see Figure 1 for an example. Leveraging this feature, we consider collaborative posts between brand accounts and creator accounts as treatment events, where a creator's contribution is publicly credited by the brand to its Instagram followers. We collect a new, unique, large-scale dataset consisting of the entire post history of 260 top fashion and beauty brands on Instagram, from which we are able to identify almost 1,000 collaborative posts with relatively small independent creators. Building a follower base is the primary objective for influencers on social media platforms (Abidin, 2016; Smith and Fischer, 2021; Libai et al., 2025; Campbell et al., 2025). The Instagram environment therefore offers a natural way to quantify exposure benefits. Specifically, we track daily follower counts for creators engaged in brand collaborations, providing direct measures of follower growth attributable to these partnerships.

To address the second obstacle of creator selection on pre-trends, we drop creators with unusually high spikes in follower counts pre-treatment, and present evidence of no pre-trends for treated creators in an event study analysis. We also leverage the high (daily) frequency of the follower panel to identify discrete jumps in followers immediately after a collaborative post (e.g., Liaukonyte et al., 2015; Li et al., 2025).

[5] See e.g. https://starngage.com/plus/en/brand/ranking.

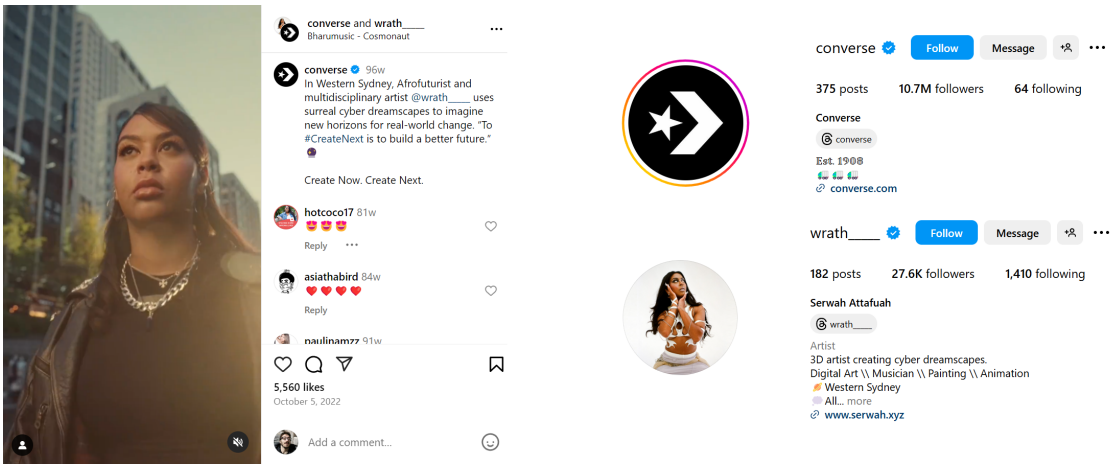


Figure 1: Example of a collaborative post

Note: The left image is the collaborative post, which is credited to both the brand (converse) and the creator (wrath_____). At time of posting, it is simultaneously displayed to the followers of both the brand and the creator. From the profile information, this implies an audience of over 10 million converse followers for the creator's post.

Third, we address the obstacle of endogenous post content, where collaborative posts might have performed well even without exposure to the brand's audience simply because of their particular features. To disentangle the effect of content versus exposure, we propose a novel matching approach using the computer vision capabilities of the ChatGPT 4o Large Language Model to extract textual descriptions of images and video from creators' posts. We convert these descriptions to numerical embeddings, and match collaborative posts by treated creators to non-collaborative posts by control creators with similar content.

Finally, we find visual evidence that the treatment effect is short-lived and exhibits non-monotonic decay over time. To capture these dynamics in follower growth, we modify the standard two-way fixed effects framework by modeling a treatment effect decay parameter that evolves according to the Poisson distribution. We then estimate this non-linear model using the Generalized Method of Moments.

Our results show that there is some value to creators from exposure to a brand's social media audience on average. We estimate a total increase in the creator's follower count of 0.47% after the collaborative post, or roughly 115-277 new followers, which corresponds to approximately one week of average growth. Given the average brand audience size of 7.6 million followers, the number of new followers for the creator implies a conversion rate of 0.002-0.004%, which is in line with the performance of online display ads estimated in the literature (Gordon et al., 2019).

We then turn to study when exposure can be most valuable for the creator. We hypothesize

that brand audience size, the attentiveness of the brand's audience, the quality of the collaborative post, and whether the content is an authentic representation of the creator will play key moderating roles. We estimate heterogeneous treatment effects, allowing it to vary by brand, post and creator attributes, and find that all else equal, collaborating with brands with larger follower counts benefits the creator substantially more. However, larger brands with more followers also tend to have less engaged followers and fewer likes on their collaborative posts, and their collaborative posts tend to differ significantly from creators' typical content. We estimate that these attributes all reduce the effectiveness of exposure. After accounting for their negative impact, we find that collaborating with brands ranging from 0.7 million to 8.6 million followers yields a roughly constant increase in the creator's follower count.

In our final analysis, we estimate the full distribution of treatment effects. We find that the top 10% of collaborations lead to a 2.1% lift in followers on average. Our results suggest that working with large, highly engaged brands who allow the creator freedom in how they craft their content can be valuable, but that it is still unlikely for creators to enjoy "viral" follower growth even in the right tail of the distribution of treatment effects.

1.1 Literature review

Our paper lies at the intersection of two growing literatures. First, this paper contributes to a recent literature on influencer incentives. Early theory papers sought to explain the behavior of influencers, and coalesced around the idea that influencers build an organic following by offering entertaining content or informative product reviews (Fainmesser and Galeotti, 2021; Mitchell, 2021; Pei and Mayzlin, 2022), and then monetize this following by creating sponsored content for their followers. Recent empirical papers have found support for the idea that creating sponsored content "harvests" the goodwill and credibility developed by influencers (Cheng and Zhang, 2025; Bairathi and Lambrecht, 2023). We add to this literature by empirically exploring the possibility that influencers can leverage brand audiences to build their own followings by supplying UGC to brands.

Second, there is a long literature on how brands can motivate UGC. We directly connect to a series of qualitative papers that ask why creators make UGC directly for brands. Roberts et al. (2014) find that individuals who co-create with video game companies are driven by the prospect of opportunities or goals to do more work in the video game industry, Berthon et al. (2008) identify self-promotion as a goal of creating unpaid ads for companies, and Somerfield (2023) show that

gaining exposure motivates UGC, especially from creators who themselves work inside the social media industry. However, these papers have not tried to quantify those benefits, which is the goal of our paper.

From the perspective of platforms, motivating UGC is important, and there has been substantial work on understanding incentives to create UGC. There are two types of incentives: extrinsic rewards and intrinsic rewards. When it comes to extrinsic rewards, visible status or hierarchy markers on an online platform have been shown to be highly effective at generating more UGC in contexts including the creation of news by citizen journalists (Sen et al., 2024), contributions on a knowledge sharing website (Goes et al., 2016), editing articles on Wikipedia (Gallus, 2017), writing comments on Reddit (Burtch et al., 2022), and posting on an online discussion platform (Lu et al., 2023b). Meanwhile, monetary payments have a mixed record of success, as paying people to contribute reviews may decrease new reviews (Sun et al., 2017) or increase review quantity but reduce review informativeness (Khernamnuai et al., 2018). Being closely related to the motivation of our research, Ghosh and McAfee (2011) show that it is theoretically possible to incentivize high-quality UGC from users who are eager for attention, by eliciting visible measures of UGC quality from platform participants.

We also connect to ongoing work on the role of brands in the social media ecosystem. Libai et al. (2025) conceptualize brands as using influencers to maximize customer equity, and we propose that brands may also leverage their customers' attention to attract influencers. Brands themselves also actively create content on social media, motivated by competition with rivals (Lysyakov et al., 2025), shaping brand perceptions (Nam et al., 2017), generating sales and engagement (Kumar et al., 2016; Lu et al., 2023a; Liadeli et al., 2023), and even informing investors (Lacka et al., 2022). Our paper spotlights how brands can leverage their audience's attention to incentivize contributions of collaborative content for their social media accounts.

Finally, our paper connects to a small literature that asks whether the creators of content can benefit from showing it on a larger platform. In a natural experiment where Google News was shut down in Spain in 2014, both Calzada and Gil (2020) and Athey et al. (2021) provide evidence that news aggregators increase page visits to news outlets, with the former showing that the aggregators increase consumers' awareness of news outlets' contents.

The rest of the paper is organized as follows. Section 2 discusses the data and institutional background. Section 3 presents our econometric model, explains the identification issues, and discusses the baseline results. Section 4 studies heterogeneous treatment effects and sheds light

on the situations under which creators can benefit more from exposure to a brand's audience. Section 5 discusses the implications of our findings for practice, and we conclude in Section 6.

2 Data and Context

2.1 Instagram and collaborative posts

Instagram is a platform for photo and video sharing with over 2 billion active monthly users, allowing users to follow accounts whose content they enjoy consuming, and, in turn, be followed by other users.[6] Instagram serves as a significant hub for social media marketing, where brands share content to their followers to promote engagement with the brand, a channel referred to as "owned" social media (Stephen and Galak, 2012). Thousands of brands are active on Instagram, with many brands pursuing a "social media first" marketing strategy and building a loyal following through content on their brand account that encourages customer awareness, engagement, and even sales (Kumar et al., 2016; Liadeli et al., 2023; Lysyakov et al., 2025).[7]

Collaborative posts were introduced by Instagram in late 2021, allowing two accounts to collaborate on a post and share it with both of their follower bases simultaneously.[8] From the creator's perspective, collaborative posts can increase their visibility, as the creator's account name clearly appears alongside the brand's name (see Figure 1). From the brand's perspective, the collaborative post helps supply their social media page with fresh content, and may be perceived as more authentic or community-building.[9]

Collaborative posts are not the only way that brands can promote creators on Instagram: brands can also tag creators in posts the brand creates, and brands can pay to "boost" sponsored posts the creator makes that feature the brand. In Appendix A.1, we discuss why collaborative posts have several advantages for a study of the value of exposure compared to tagging or sponsored posts.

[6] See "60+ Instagram Statistics Marketers Need to Know in 2024," *SocialPilot*, September 10, 2024.
[7] See also "Launching Consumer Brands in the Age of Social Media," *Meltwater*, November 11, 2018.
[8] For details, see "Here's What to Know About Instagram Collab Posts in 2025," *Later.com*, March 25, 2025.
[9] See "How to Use Instagram Collab Posts to Get More Engagement," *Hootsuite*, August 14, 2023.

2.2 Data collection

Brands. We start by identifying the top brands in beauty, clothing, shoes, and fashion that are active on Instagram. We first use *FashionUnited.com* to collect the top 100 fashion brand Instagram accounts, and then parse the top 5000 most-followed Instagram accounts in the "business" category on *StarNGage.com* to collect all brand accounts tagged with "Fashion", "Beauty", "Clothing", or "Footwear."[10] Finally, we search Wikipedia for the top U.S.-based fashion and beauty brands, and collect Instagram accounts for brands with at least 100,000 followers.[11] After removing duplicated accounts and dropping brands' auxiliary accounts, we gather 260 brand accounts in total, primarily based in the United States.[12] The distribution of the number of followers across brands is shown in Figure 2a (note that the y-axis of Figure 2a is in exponential scale); the largest brand is Nike with over 300,000,000 followers, and the median brand has 2,670,000 followers.

Collaborative Posts and Collaborating Creators. For each brand, we use a web-scraper to pull the entire post history of that brand on Instagram back to January 2021. Using post metadata, we recover 15,340 collaborative posts and extract the complete set of 4,320 collaborator accounts.

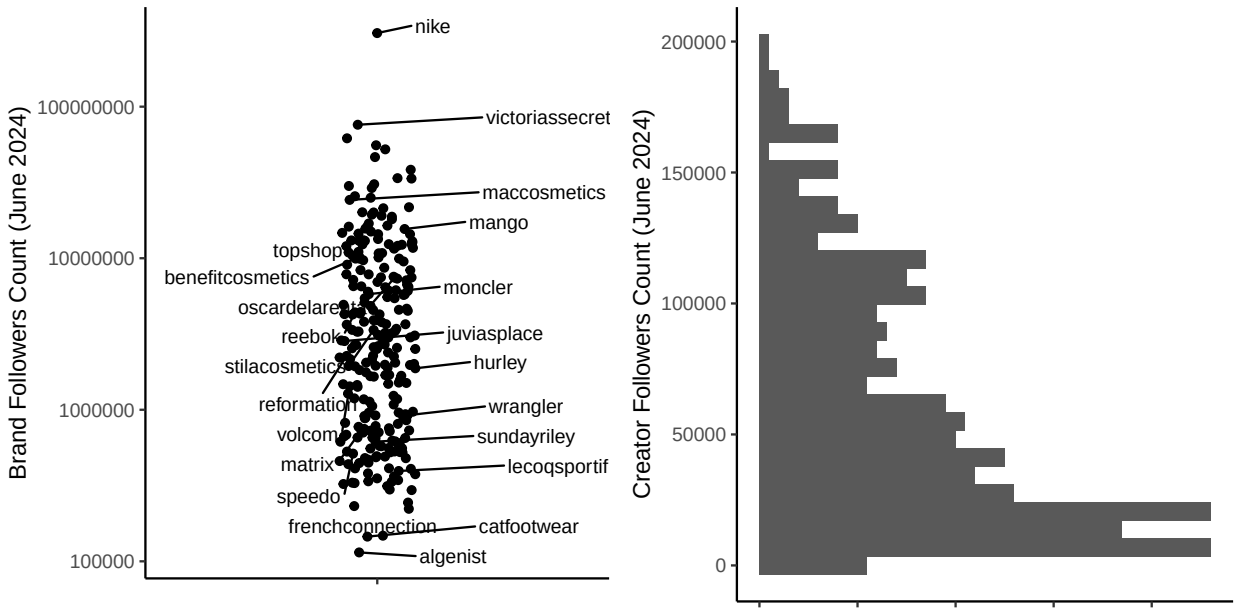
We trim this set of collaborator accounts in four steps to reach our final sample of creators. First, Instagram accounts self-identify as either "business," "personal," or "creator," and we drop all business accounts. Second, for each of our 260 brands we compile a list of auxiliary Instagram accounts managed by that brand, and drop those accounts. For instance, @nikewomen would be dropped as a potential creator account, as they are owned and operated by Nike. Third, of the remaining 1,720 accounts, we drop all accounts with more than 200,000 followers, yielding 953 collaborator accounts. This selection aligns with our research objective: to measure the effect of collaborations between relatively large brands and smaller creators, where we expect that the exposure value to be strongest.

Finally, we drop all creators who have more than one collaborative post with the same brand. This ensures a cleaner quasi-experimental setup for measuring dynamic treatment effects from

[10] See https://fashionunited.com/i/instagram-fashion-index and https://starngage.com/plus/en/brand/ranking, accessed on 10/10/2024. StarNGage includes all types of business Instagram accounts in its ranking, the majority of which are not fashion, beauty, clothing, or footwear related.

[11] We search on Wikipedia for Category: Clothing brands of the United States and Category: Cosmetic companies of the United States.

[12] Some brands have multiple accounts, e.g. @nike and @nikerunning, or have different accounts for different geographic regions, e.g. @thenorthface and @thenorthfaceuk. We only keep the account for each brand with the most followers.



(a) Brand follower counts (exponential scale)

(b) Creator follower counts

Figure 2: Distribution of follower counts on Instagram

Note: In the left figure, each point is a particular brand, and labels are the brand's Instagram account name. Points are jittered horizontally for readability, and the y axis has an exponential scale for easier visualization. In the right figure, we plot the histogram of follower counts for creators. The distribution of creator counts has an interquartile range of [20,000, 96,000].

exposure, as multiple treatment events in quick succession could otherwise contaminate the results. We are left with 692 treated creators who only have a single collaborative post with one of the 260 brands. We plot the distribution of these creators' follower numbers (in August 2024) in Figure 2b.

Note that we do not observe the contract between the brand and the creator that governs the collaboration, and it is possible that the creator may be paid in some instances. Insofar as compensation leads to different types of collaborative content, we describe how we control for this content in Section 2.3.

Follower data. The key outcome of our study is how the number of followers for each creator changes after they collaborate with a brand. We collect the complete followership history for each collaborating creator account from SocialBlade. SocialBlade is a social media marketing company that helps brands identify influencers for partnerships. Importantly for our purposes, they collect daily follower count information for tens of millions of accounts. Since the number of followers of each account tend to increase over time and is therefore non-stationary, we work primarily with the follower growth rate or the change in followers, to make the data amenable to fixed effects analysis.

We exclude creators whose posts are preceded by an unusually high follower growth, to remove collaborations between brands and creators that may be sparked by a creator's sharply rising popularity. Specifically, we drop any creator whose follower growth rate in the 30 days before their collaborative post falls within the top 1% of their historical daily growth rates. After this filtering step, we are left with 648 creators in the sample.

Posting data. We gather the complete posting history for all 648 treated creators from January 2021 through August 2024. In addition to post level metrics such as likes, comments, and video views, we extract the complete set of media elements including the caption, images, and videos associated with the post.

Pool of potential control creators and posts. In our difference-in-differences analysis, for each treated creator's collaborative post, we would ideally want to find a control creator whose follower growth satisfies the parallel trend assumption and who makes a post that closely resemble the collaborative post. To find such control creators and posts, we turn to two sources: (a)

posts by treated creators pre-treatment, and (b) posts by never-treated creators (Callaway and Sant'Anna, 2021; Roth et al., 2023).

For posts by treated creators pre-treatment, we restrict to non-collaborative posts created at least 90 days prior to their collaborative post. This criterion is designed to avoid overlap in our observations of follower growth, as we analyze its evolution within a three-week pre-treatment and six-week post-treatment window. This yields approximately 210,000 control posts by treated creators pre-treatment.

For never-treated creators, we choose them from the set of "similar accounts" for each treated account determined by Instagram (see Appendix A.2). We apply the following selection criteria: (1) the accounts must be in the same category (e.g. artist, athlete, etc.) and have a similar follower count as the treated account; (2) SocialBlade must have follower data for those accounts. This leads to 625 never-treated creators who collectively post 220,000 times during our sample period.

The 430,000 control posts by treated creators pre-treatment and never-treated creators comprise our potential control pool which we use in our baseline analysis. One may worry that there might be spillovers from the treated creators to the never-treated creators because the never-treated "similar accounts" are recommended by Instagram when browsing the treated creator accounts. In Section 3.4, we discuss why spillovers are unlikely due to how we select the matched posts and details of the Instagram interface. In a robustness check, we also re-estimate the treatment effect using only pre-treatment posts by treated creators as our control pool (i.e., dropping never-treated creators), and the results remain unchanged.

2.3 Controlling for post content

To estimate the effect of collaborative posts on follower growth, one would ideally compare two identical creators with identical posts, one of which was posted by both brand and creator accounts simultaneously (i.e., the collaborative post), and the other which was only posted by creator account (i.e., creator's solo post). This type of comparison would allow us to control for the effect of post content on follower growth. To approximate this ideal experiment, we pursue a strategy of matching treated creator-posts to control creator-posts, to create a matched sample of 648 solo posts by control creators that closely resemble the 648 collaborative posts by treated creators on average.

To achieve this matching, we take the following three steps. First, we convert the rich, complex visual content of the images and videos in the 648 treatment and 430,000 control posts into a

numeric embedding space. Second, we reduce the dimension of these embeddings in a way that preserves the attributes of the content, but is also interpretable by humans. Third, we estimate a matching function that predicts treatment status based on creator and post features, and validate that the function yields control creator-posts that are observably similar to treated creator-posts. Combined with the assumption that the content features we extract are sufficient to solve the selection problem, we can then isolate the effect of exposure to the brand's followers.

2.3.1 Embedding post content

The literature has taken two approaches to capture visual data in studies of social media content: (i) hand-selecting a set of features such as facial emotions and color hue and then using machine learning libraries to extract those features for each image (Li and Xie, 2020; Zhang et al., 2022; Beichert et al., 2024; Cheng and Zhang, 2025), or (ii) using pre-trained image-recognition models such as CLIP or VGG-19 (Simonyan, 2014) to embed visual content as numeric matrices before conducting dimension reduction (Tian et al., 2024; Yang et al., 2025; Rajaram and Manchanda, 2020).

Our approach presents a more advanced and automated data processing flow that extends the latter method. We start by using the vision capabilities of ChatGPT 4o to recover textual descriptions of the visual content of a post. ChatGPT 4o is a Large Language Model (LLM) with substantially improved image recognition capabilities relative to previous models. Unlike alternatives such as CLIP or VGG-19 that mainly focus on object detection, ChatGPT retains more information about the relationship between individuals and objects (e.g., distinguishing "a person playing with a ball," from just "a person and a ball") and contextual details (e.g., recognizing "exercising outdoors" versus "exercising in a gym"). For each image in a post, we prompt ChatGPT 4o with "Describe this image," and recover the textual description. For each video, we take snapshots every 3 seconds for the first 15 seconds and perform the same prompt on each image. In aggregate, combining both images and video screenshots, we have over 2 million images requiring description.[13] In Appendix A.3, Table A.1 we provide examples of ChatGPT 4o supplied descriptions of images in our dataset.

We next convert each image description that we recovered to a dense vector embedding. We

[13]Before prompting ChatGPT, we downscale images to have a maximum size of 300 x 450 pixels. We choose 300 x 450 pixels since there is a tradeoff between the cost of the API and the accuracy of ChatGPT 4o when faced with lower-resolution images; we find that ChatGPT 4o is able to describe 300 x 450 pixel images with similar levels of accuracy as images in their original resolutions.

use *paraphrase-mpnet-base-v2* from the Sentence Transformers library, a pre-trained BERT-based algorithm with 109M parameters to contextualize information in an entire paragraph as opposed to single sentences at a time (Reimers, 2019). We take the pre-trained embeddings as a starting point, but add a fully connected layer to SBERT that predicts whether a post is a collaborative post or not. We unfreeze the top 4 layers of SBERT to allow the embedding model to re-weight parameters—a supervised learning process known as fine-tuning. This process will recover embedding vectors that are most relevant to our goal of predicting treatment status. The fine-tuned SBERT model yields a 768-dimensional vector embedding that captures the semantic meaning of the description for each image.

2.3.2 Creating topic clusters from content embeddings

In theory one could match control posts to treatment posts using the 768-dimensional vector of SBERT embeddings (aggregated across images for posts with multiple images or video content). We pursue dimension reduction and topic clustering to improve robustness to overfitting on noise, increase interpretability of the matching variables, and conserve computational resources.

For each image, we reduce the SBERT embedding to 200 dimensions using Principal Components Analysis. We then use the Uniform Manifold Approximation and Projection (UMAP) algorithm to reduce each PCA vector to 5-dimensions, which preserves both local and global features of the embedded vectors. Finally, we use a simple k-means clustering algorithm to group vectors, and use the elbow method to select the optimal number of clusters, which is 50. Figure 3 details the complete dimension reduction and clustering pipeline for images in the dataset.

We showcase representative images from 4 of the 50 clusters in Table 1. The images are chosen at random from within a fixed radius of the centroid of each cluster. The "Topic" column is a post-hoc analysis after clustering using the TF-IDF formula to pick out words that are especially prevalent within the image descriptions in each topic cluster relative to the corpus of descriptions as a whole, and the "Topic Size" column indicates how many images are associated with the cluster.

The comparison between the second and third rows showcases a benefit of our approach of using the rich contextual data supplied by the LLM descriptions. While all six images in these two topics feature individuals working out in similar attire, the LLM identified that images in the second row feature individuals focused on running on an outdoor track, while images in the third row feature individuals lifting weights or doing pliometric exercises in a gym.

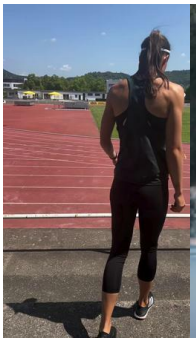
Topic	Topic Size	Image 1	Image 2	Image 3
salon, hair, stylist, cape	67154			
running, track, athletic, race	64153			
gym, exercise, performing, equipment	48624			
selfie, mirror, taking, room	23240			

Table 1: A sample of topics resulting from embedding and clustering

Note: Each row is a distinct cluster in the data, created through k-means clustering on the dimension-reduced SBERT embeddings. We present 4 clusters (out of 50 total) to provide face validity of the algorithm's ability to cluster images meaningfully. "Topic" words are recovered after clustering, by applying TF-IDF on the image descriptions to determine which words appear much more frequently in the image descriptions in that topic cluster relative to the entire corpus of image descriptions. "Topic Size" indicates how many images belong to that topic cluster. Image 1, 2, and 3 are example images drawn from that topic cluster, for each cluster.

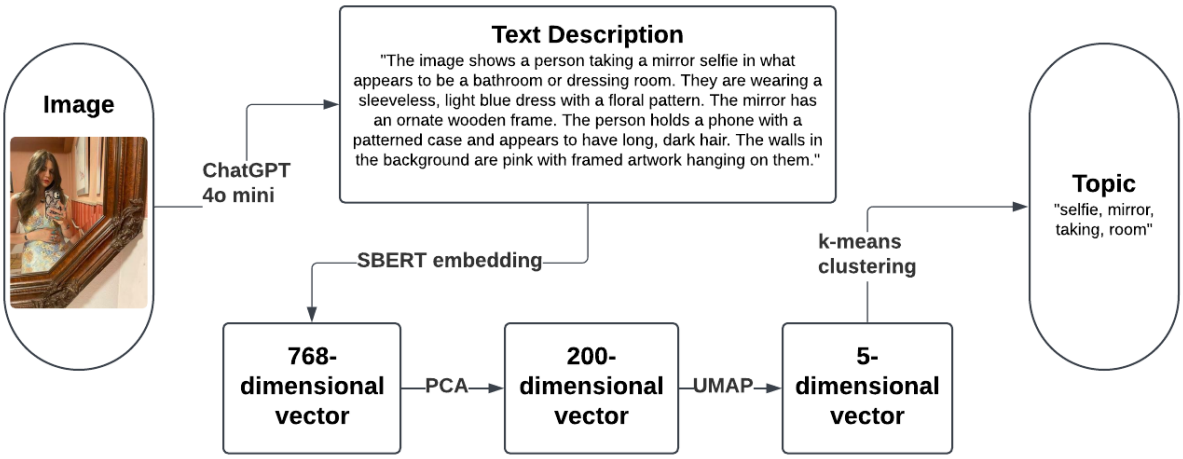


Figure 3: Topic creation flow

Note: ChatGPT 4o is accessed using the OpenAI API, using the prompt "Describe this image." SBERT is the sentence transformer model paraphrase-mpnet-base-v2, PCA is principal component analysis, UMAP is the uniform manifold approximation and projection, and k-means clustering is performed with $k = 50$. Topic text is included for descriptive purposes; the topic number itself (from 1-50) is included as an image's contribution to post-level content features.

2.3.3 Selecting matched control creator-posts

Our objective is to select a set of matched solo posts by control creators that are similar to the treated creator-posts. We proceed by converting our image-level topics into post-level content attributes, adding several hand-selected features of posts and creators, and then detailing our matching function.

To convert our image-level topics to post-level attributes, we take the average of image topics over all images within a post to arrive at a post-level topic description. For example, for a post that includes both a video and an image, we sum the topic vectors for the image and each video snapshot, and then divide by the total number of images plus the number of snapshots. This 50-dimensional vector captures the topic weights within each post and sums to 1.

For our hand-selected features, we include several post-level attributes that are likely relevant to a post's performance, including the precise timing of a post (year, month-of-year, day-of-week, and hour-of-day), and the type of post (Video, Image, or Sidecar, the latter of which denotes a post that contains multiple images, videos, or both). For each post, we also include the lagged value of the total number of creator's followers 7 days prior to the collaborative post.

To select our set of matched control posts, we apply Propensity Score Matching (PSM). PSM models the propensity to be treated as a function of observable characteristics, and for each

treated post, selects a control post with the most similar propensity score. We use the post and creator-level attributes described above and estimate the propensity score using XGBoost, a flexible machine learning model based on decision trees that minimizes overfitting. Note that PSM is essentially a further dimension reduction, where a post's only relevant attribute becomes its propensity to be treated. We use greedy pairwise matching to select control posts sequentially in the following way: starting with a random treatment post, we select the matched control post from the pool of 430,000 with the closest propensity to be treated, and then remove from the pool all candidate control posts by the same creator 90 days before and 90 days after the selected matched control. We drop these posts because in our empirical strategy, we will track follower growth from 3 weeks (21 days) before each post to 6 weeks (42 days) after, and this trimming step ensures no overlap in observations. We highlight that when a control post is selected, its creator is also selected as a control creator.

Appendix A.4 reports the standardized mean differences between covariates, and shows that the PSM model achieves a substantial balance improvement relative to the complete set of control posts. In robustness checks in Section 3.4, we consider an alternative PSM-matched sample that only includes control posts from the treated creators pre-treatment.

2.4 Controlling for cross-posting

An important confound for identifying the effect of collaborative posting on Instagram is the tendency of brands to cross-post content to multiple social media platforms (Tian et al., 2024). It is possible that brands may simultaneously post about a creator on Instagram, Facebook, and Twitter/X. If our analysis does not control for such cross-posting activities, our estimated exposure effect maybe biased upwards. To address this issue, we control directly for brand cross-posting after gathering cross-platform data.

We start by collecting each brand's account name on TikTok, Facebook, and Twitter/X by hand, and then scrape each brand's entire posting history on each platform. To code cross-posting, we look at a brand's posts on other platforms within a two-week window of one of their collaborative posts, and check if they mention the collaborating creator in these posts. We check for mentions by conducting fuzzy matching between the collaborating creator's Instagram handle and any mentioned accounts in the posts on other platforms, and manually validate potential matches.

We find that the propensity to engage in cross-posting varies substantially across platforms,

but in general is infrequent. We document 88 instances of a brand creating a post that mentions the collaborating creator on TikTok, 44 on Facebook, and only 10 on Twitter/X. Of these, Facebook exhibits the highest proportion of same day posts—37 out of 44—which likely reflects the ease of cross-posting between two platforms owned by Meta. Seven of 10 Twitter/X posts are also simultaneous, while only 37 TikTok posts are simultaneous. In Section 3.1 when we present the model, we discuss how we operationalize our controls for cross-posting. In Section 3.4, we conduct a robustness check without using any cross-posts, and our findings remain largely unchanged.

2.5 Summary statistics and model-free evidence

Summary statistics. Our final sample consists of 648 collaborative posts, 648 control posts, and the follower growth rates for the creators making the collaborative posts and the creators making the control posts. In Table 2, we document the average rates of follower growth in our data around the time of the treated and control posts.

Table 2: Outcome variable summary statistics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Follower Growth	49,142	0.0008	0.0034	−0.0545	−0.0001	0.0006	0.2270

Note: follower growth is daily growth rate. The summary stats are computed using a window of 21 days before to 42 days after the collaborative posts (or matched control posts).

Model-free evidence. To visualize the patterns of follower growth in the data, we plot their average values (across creators, within day) by treatment and control before and after the collaborative posts and matched control posts in Figure 4. We plot the averages for 21 days before and 42 days after the collaborative post (or the matched control post for the control creators). We choose our post-treatment time window conservatively: Cheng and Zhang (2025) find that follower counts change within a few days after a sponsored post on YouTube, Huang and Morozov (2025) find that Twitch streams lead to changes in viewers' product engagement for streamed video games that decay within 24 hours, and Li et al. (2025) find that YouTube videos about video games lead to changes in engagement that decay within seven days.

The raw data suggests that prior to treatment, follower growth for treated and control creators

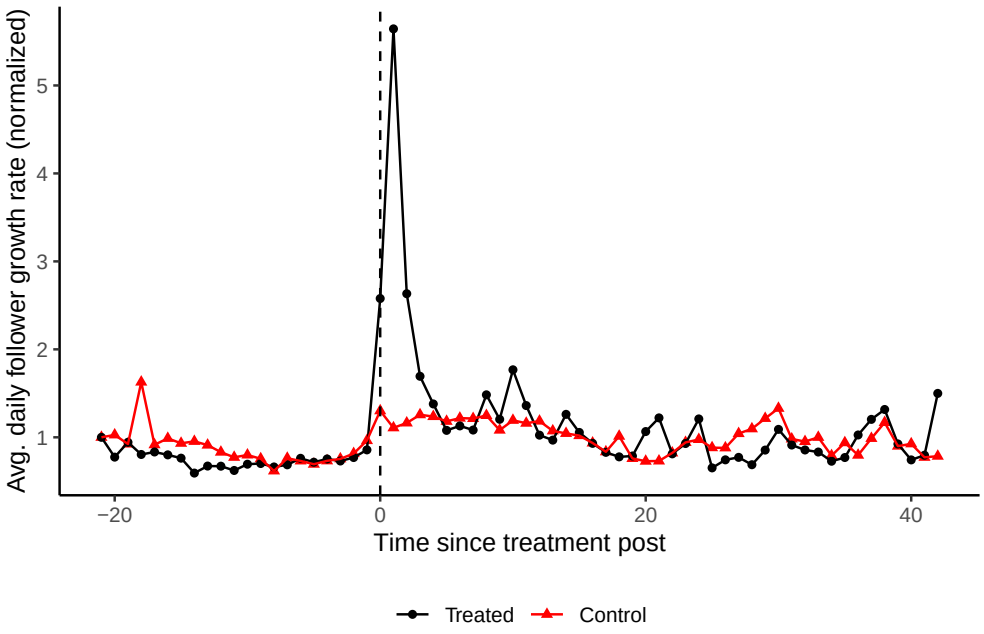


Figure 4: Normalized average follower growth rates around time of treatment

Note: The black circles denote the average outcome for treated creators, divided by its value 21 days prior to the collaborative post. The red triangles denote the average outcome for control creators, divided by its value 21 days prior to the matched control post. For treated creators the treatment post is the collaborative post. For control creators, the "treatment" post is the control post selected through the propensity score matching model in Section 2.3.

behaves similarly, with a sharp divergence on the day of posting, peaking at a 5.5x higher growth rate for treated creators on the day after posting. This effect decays quickly, with treated creators largely returning to the same trend as control creators by 4 days after posting. Visually, for control creators there does seem to be a slight uptick in the daily follower growth rate at the time of the control post, which decays by 20 days after the post. Thus, both types of post seem to lead to growth, but with a large initial spike for collaborative posts before returning to the baseline growth trend. The short duration of this spike is consistent with the nature of the Instagram algorithm, which tends to prioritize showing recently posted content (Vassio et al., 2021). In Section 3, we formalize this analysis in an econometric model, and argue that we can use this variation to causally identify the effect of exposure on follower growth.

3 Model and Results

3.1 Model

Our analysis aims to identify the Average Treatment Effect on the Treated (ATT) for the subset of creators who are selected to collaborate on posts with brands. We estimate the effect of collaboration on follower growth rate, using a modification of the standard two-way fixed effects (TWFE) model:

$$\text{FollowerGrowth}_{jt} = \beta \cdot f(t - t^*(j), \theta) \cdot \text{Treat}_{jt} + \boldsymbol{\gamma}' \mathbf{X}_{jt} + \alpha_j + \alpha_t + \alpha_{t-t^*(j)} + u_{jt} \quad (1)$$

Our fixed effects α_j and α_t are the standard entity and time fixed effects, and control for time-invariant differences in follower growth rates across creators and Instagram-wide shocks on particular calendar dates, respectively. $\alpha_{t-t^*(j)}$ differs from the typical TWFE model: this is a fixed effect for the relative time-to-post, where $t^*(j)$ is the calendar date on which treated creator j makes their collaborative post, or on which control creator j makes their matched post. Intuitively, the fixed effect $\alpha_{t-t^*(j)}$ flexibly captures the baseline average follower growth from the matched control posts with similar content around the time of posting (the red triangles in Figure 4), so that we can isolate the treatment effect of exposure due to collaboration.

Our treatment variable Treat_{jt} equals one for a treated unit post-treatment, as in a standard TWFE model. The function, $f(t - t^*(j), \theta)$, which is parameterized by θ, captures the decay of the effect of collaboration over time relative to the day of posting, $t^*(j)$. As seen in the model-free evidence in Figure 4, the treatment effect does not peak on the day of treatment and then uniformly drop off, but rather peaks a day after the post before subsiding over the next few days. We therefore use a Poisson probability density function with $f(s, \theta) = \frac{\theta^s e^{-\theta}}{s!}$ for $s \geq 0$, which is parsimonious and still flexible enough to capture the non-monotonic decay. For $s < 0$, we set $f(s, \theta) = 0$.

The parameter β captures the effect of treatment on the outcome. To interpret β, note that since $\sum_{s=0}^{T} f(s, \theta) \approx 1$ for large enough T, β is the *cumulative* effect of treatment over time, and $f(s, \theta)$ captures the share of the total treatment effect that manifests on day s post-treatment.

Finally, the vector of controls $\boldsymbol{X}_{jt}$ includes dummy variables that capture cross-posting by brand j on other social media platforms. Specifically, we construct a vector of three dummy variables, each of which is equal to one for creator j at time t if the brand they collaborate with

on Instagram posts about them on TikTok, Facebook, or Twitter/X respectively. We then include lags of these dummy variables up to seven days prior, which yields a 24-dimensional vector of control dummy variables for each creator that captures contemporaneous and lagged posts about the creator by collaborating brands on other platforms.

In the next subsection, we discuss our identification assumptions. Readers who are interested in the results can skip to Section 3.3.

3.2 Identification

Equation (1) is designed to identify the causal effect of a collaborative post on a creator's follower growth. Our identification assumption is that conditional on the creator fixed effects, date fixed effects, time-to-post fixed effects, and controls, the treatment variable Treat_{jt} is orthogonal to the unobserved error u_{jt}.

Having controlled for post content through matching and cross-posting through our vector of cross-posting dummies, the key remaining concern is that the timing of treatment may coincide with unobserved shocks that also drive interest in the creator. There are two distinct issues related to timing: brands may select creators based on trending follower growth, and brands may release collaborative content to coincide with creator-specific shocks.

Ruling out selection based on pre-trends

To rule out that brands select creators based on trending popularity, we note that in Section 2.3, we dropped treated creators with unusually high growth spikes in the 30 days prior to treatment. However, creators may still be selected based on a strong upwards growth trend even without unusually large spikes in growth. Our high-frequency daily data allows us to observe whether a pre-trend in creator growth drives the treatment effect, or whether there is indeed a spike in growth concurrent with treatment (e.g., Liaukonyte et al., 2015; Huang and Morozov, 2025). We conduct a formal event study analysis in Appendix B.1 and find no evidence that treated creators exhibit positive trends in follower growth in the days prior to their collaborative posts. We also show that there is no differential linear pre-trend for treated creators in Appendix B.2. These tests suggest that creators are not selected to contribute content based on trending popularity. This finding accords with institutional evidence that brands have complex internal approval processes for posting content that frequently lead to delays when working with creators, and

keeps them from acting quickly on trends.[14]

Ruling out co-ordinated shocks

The second timing issue arises if a brand coordinates the release of a creator's content with an external popularity shock. For instance, if a brand is working with a visual artist, they may post the collaborative content on the day of a gallery opening; if they are working with an athlete, they may post the content on the day of a high-profile competition.[15] Identification of Equation (1) requires us to assume that treatment is orthogonal to such transitory creator-specific demand shocks.

To check this assumption, we gather a proxy variable for creator-specific demand shocks and correlate it with the timing of collaborative posts. Our proxy is weekly global search volumes scraped from Google Trends for the collaborating creator's actual name. We successfully link 83% of treated creators to their actual name using metadata from their Instagram account, and then collect global search volume data by week for each name. We regress Treat_{jt} on search volume with creator fixed effects to test whether treatment timing is related to changes in search volume, and find an insignificant partial correlation of 0.0001 ($p = 0.77$), suggesting no relationship between treatment timing and external demand shocks.[16] We also perform a visual timing test: we first identify the week with peak search volume for a creator that occurs within six weeks of the collaborative post for that creator, and then plot the propensity of the brand to create the collaborative post around the period of peak interest.[17] If the timing of collaborative posts is exogenous with respect to surges of search volume for the creator, then we should find that the propensity to post does not rise during the period of peak interest, or in any period around the period of peak interest. As shown in Figure 5, we find no evidence that brands are more likely to post during the week of peak interest compared to other weeks surrounding the collaborative post.[18] In Appendix B.3, we check the robustness of this exogeneity result to different timing

[14]See "6 steps for an efficient marketing approval process", *Ziflow*, March 27, 2023, as well as "How to Speed Up the Content Approval Process in a Large Organization", *Kuno Creative*, February 3, 2017.

[15]Brands are designated as the post owner in 100% of our posts, and therefore have control over timing. This arrangement is typical and is meant to manage reputation risk, as only the owner has the ability to remove or modify the collaborative post. See https://www.facebook.com/help/instagram/291200585956732.

[16]Since treatment is observed at the daily level and search volume is at the weekly level, we impute the daily volume as the weekly volume in this regression.

[17]Appendix B.3 provides visual examples of the size and timing of local peaks for several creators.

[18]If the brand's collaborative post leads to Google search interest in the creator, then the timing of the post and Google search volumes may coincide. Since we find no relationship between post timing and search volume, this mechanism is also ruled out.

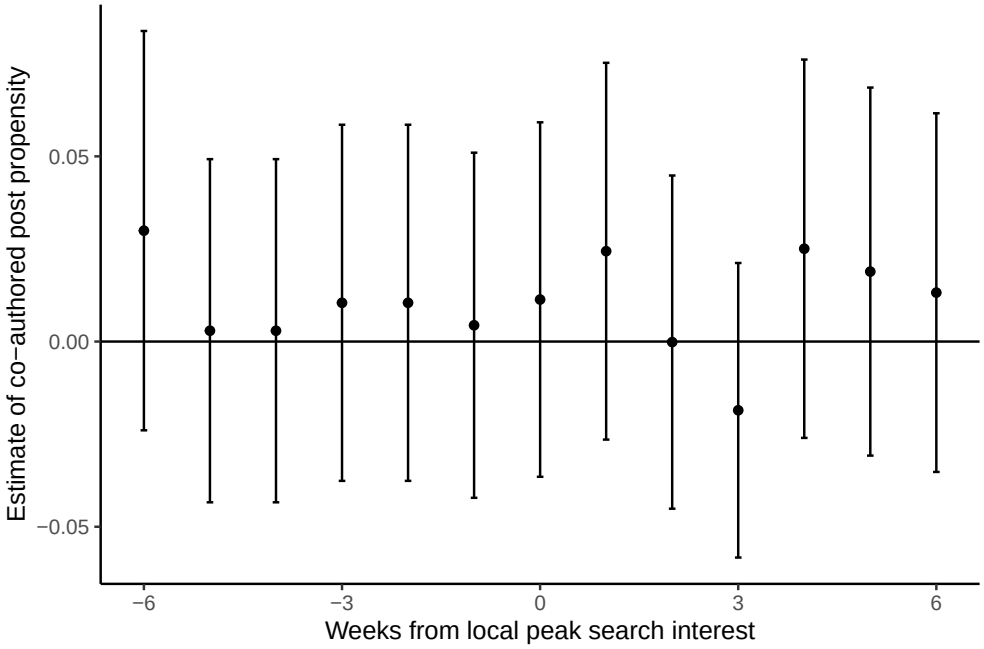


Figure 5: Collaborative post propensity around times of peak search interest in creator

Note: This figure presents coefficients from a regression at the creator-week level, where the dependent variable is whether the creator's collaborative post on Instagram was posted that week, and the independent variables are a set of leads and lags around the week of highest search volume for the creator's actual name. If brands time collaborative posts on Instagram with external demand shocks for the creator, there should be a positive propensity to post in the week before, during, and after the external demand shock, which we do not find. Error bars are 95% confidence intervals clustered at the creator level. See Appendix B.3 for more details.

windows, and our findings remain unchanged.

3.3 Results

We estimate Equation (1) using the Generalized Method of Moments (details in Appendix B.4). We report the estimated effect of a collaborative post on the follower growth rate for creators in Column 1 of Table 3 without controls, and in Column 2 controlling for brand cross-posting.

The β parameter is positive and highly statistically significant in both specifications, as expected given the patterns in Figure 4. The estimate of β in Column 2 implies that the collaborative post leads to a *cumulative* 0.47% bump in followers. The β coefficient in Column 2 shrinks slightly compared to Column 1 after including the controls. We find that this reduction only occurs after including the Facebook posting controls, which may be because cross-posting between Facebook and Instagram is relatively easy as both are owned by Meta.

For the Poisson distribution parameter θ, the estimate in Column 2 suggests that 26% of the

cumulative follower growth is experienced on day 0, 62% by day 1, and 85% by day 2. This pattern accords closely with the visual evidence in Figure 4, where the majority of follower growth is earned the day after the post.

In Columns 3 and 4, we replicate our analysis and estimation using the change in followers as the dependent variable instead of the growth rate in followers.[19] After including controls, we estimate a cumulative increase in followers of 114.6 on average, with a Poisson distribution parameter whose magnitude implies most growth is earned on the day after the post.

The average creator in our data has 59,000 followers prior to collaborating, and the average brand has 7.6 million followers, an audience over 120 times larger than the creator's. From Column 2, creators earn a 0.47% cumulative bump implying 277 new followers, while estimating the direct gain in Column 4 implies only 114.6 new followers.[20] Given that the average daily growth rate is 0.08%, the gain from exposure represents earning approximately an extra week's worth of growth with a single post.

To benchmark the magnitude of the treatment effect, we draw on the literature measuring the effectiveness of online advertisements. The settings have some similarities: a potential audience of known size is exposed to a piece of content, and a successful conversion involves engaging with that content, in our case by following the creator, and in advertising by signing up for an app/email list or purchasing a product. With an average audience of 7.6 million and 115-277 new followers, we have an implied conversion rate of 0.002-0.004%. Comparing this magnitude to intention-to-treat (ITT) estimates of conversion rates of display ads from (Gordon et al., 2019), our finding is within the range of these estimates, but closer to the lower end.[21] This small average effect may reflect that these posts are primarily designed to increase engagement with the brand, with audience conversion (from brand followers to creator followers) operating as a secondary benefit. In our heterogeneity analysis, we explore whether collaborative posts that better align with the creator's content have a larger treatment effect.

[19] We replicate the event study plot using the growth in followers (not growth rate) in Appendix B.5.

[20] In general, $E[Y_t - Y_{t-1}|X_t] \neq E[(Y_t - Y_{t-1})/Y_{t-1}|X_t] \cdot Y_{t-1}$. The discrepancy in estimates is driven by the skewness of the follower growth variable.

[21] Gordon et al. (2019) estimate (significant) ITT lift between 0.003%-0.059% for RCTs where conversion is purchase, and 0.003%-0.492% for RCTs where conversion is registration. ITT is the appropriate comparison because we do not know which members of a brand's audience actually see the post. The actual number of brand followers who see the post depends on whether they log in while the post is recent, and whether they frequently engage with the brand's posts, which are unobserved.

Table 3: Average baseline effects of collaborative posts

	Dependent variable:			
	Follower Growth Rate		Follower Growth	
	(1)	(2)	(3)	(4)
β	0.0053	0.0047	129.9	114.6
	(0.0013)	(0.0012)	(44.7)	(44.6)
	[p=0.0000]	[p=0.0001]	[p=0.0037]	[p=0.0102]
θ	1.3562	1.3291	1.3060	1.2304
	(0.1224)	(0.1224)	(0.9021)	(0.9020)
	[p=0.0000]	[p=0.0000]	[p=0.1477]	[p=0.1725]
Creator FE	✓	✓	✓	✓
Date FE	✓	✓	✓	✓
Time-To-Post FE	✓	✓	✓	✓
Cross-Posting Controls		✓		✓
Avg. Value of Dep. Var.	0.0008	0.0008	42.2	42.2
Observations	47,524	47,524	47,521	47,521
R^2	0.1907	0.1932	0.2760	0.2770

Note: This table presents estimates from Equation (1). In Columns 1 and 2, the dependent variable is the daily follower growth rate, and the estimated β coefficient, when multiplied by 100, is the cumulative percent increase in followers (relative to just before treatment) due to the collaborative post. In Columns 3 and 4, the dependent variable is follower growth, and the estimated β coefficient is the cumulative increase in the number of followers (relative to just before treatment) due to the collaborative post. θ is the Poisson parameter and governs the shape of the treatment effect decay over time. Cross-posting controls are a vector of dummy variables capturing if a brand posted about the creator to TikTok, Facebook, or Twitter/X, with 7 days of lagged dummies included for each platform. Standard errors are clustered at the creator level and reported in (.) below the estimated coefficients. Exact p values are reported in [.] below the standard errors.

3.4 Robustness

Dropping Collaborative Posts with Cross-posting

Our approach to handling unobserved exposure generated by brands cross-posting to other platforms is to include a vector of cross-posting controls. As an alternative to controlling explicitly for cross-posting, we also try dropping all treated creators where we detect cross-posting. Estimates in Appendix C.1 show that dropping these events has little effect on the size or significance of the treatment effect.

Control Creator-Posts

To create our matched sample in Section 2.3, we draw on control posts from two sources: treated creators pre-treatment, and never-treated creators. Because never-treated creators are

"similar accounts" for treated creators, one might worry that users who examine the treated creator's profile may also view these other creators, and this could lead to positive spillovers in attention from the treated creators to control creators. We believe this issue is unlikely to be severe: (1) in Appendix A.2, we document that these "similar" accounts are not salient and it requires effort to reveal them, minimizing the chance of spillovers; (2) in Table 3 we find that our treatment effect decays rapidly, and since our matched posts are not matched on calendar dates, the chance of a spillover actually affecting the matched post growth rate around the time we observe a matched post is very low. Nevertheless, we replicate Table 3 using only non-collaborative, pre-treatment posts from treated creators as the control pool, after re-estimating the PSM function and selecting new matching control posts from the more limited pool. Results reported in Appendix C.1 suggest that the estimated treatment effect is unchanged.

Dropping Creators Who May Cross-Post

Some creators are active across multiple social media platforms, and may attempt to publicize their brand collaboration to followers on these other platforms, potentially leading to greater follower growth on Instagram if there are cross-platform externalities. We estimate event studies for treatment effects after dropping creators who have large followings on other social media platforms in Appendix C.2, and find that omitting these creators has no impact on treatment timing or magnitude.

Dropping Creators with Immediate Follow-on Posts

Finally, our sample includes creators who post immediately after the collaborative post. Since the largest follower growth does not manifest until a day after the collaborative post, one may worry that immediate follow-on posts are driving the treatment effect. We therefore estimate Equation (1) using only those creators who do not create any new posts immediately after the collaborative post. We find no significant difference in the cumulative increase in the number of followers in Appendix C.3, and a slight increase in the Poisson distribution parameter compared to the baseline estimates.

4 Heterogeneity Analysis

In this section, we focus on heterogeneous treatment effects. We articulate our hypotheses in Section 4.1, specify the econometric model specifications and heterogeneity variables in Section 4.2, and present results in Section 4.3. Section 4.4 discusses whether it is more beneficial to work with brands with more followers. In Section 4.5, we report the empirical distribution of follower increases from exposure.

4.1 When do creators benefit more?

Our first set of heterogeneity hypotheses focuses on brand-level attributes. We hypothesize that brand size will positively moderate the effect of exposure, as a larger audience provides more opportunities for the creator to be seen and followed.[22] Moreover, we hypothesize that the engagement level of the brand's audience will also have a positive moderating effect, increasing the likelihood that followers notice the collaborative post (Smith and Fischer, 2021; McGranaghan et al., 2022). Finally, we hypothesize that higher posting frequency by the brand posts will negatively impact exposure effectiveness, all else equal. This is because the collaborative post will be replaced more quickly in followers' content feeds by new posts from the brand (Vassio et al., 2021; SocialBee, 2025), leading to a shorter exposure window.

Our second set of heterogeneity hypotheses are based on post content. If the content of a collaborative post is very different from the creator's typical content, the brand's followers may be less likely to follow the creator even if they like the collaborative post.[23] We thus hypothesize that the difference between the post content and the creator's typical content will dampen the benefit from exposure. Meanwhile, a collaborative post whose content is very different from a brand's usual content may grab users' attention if it is perceived as creative and novel (Kanuri et al., 2024; Villanueva et al., 2008; Somerfield et al., 2018), or it may be mismatched with the brand audience's average tastes, leading to negative attention. The combination of these two factors makes the overall prediction regarding the difference between the collaborative post and

[22] Creators tend to express beliefs that only working with large brands can be valuable (e.g., "anybody big enough to actually give you exposure is big enough to pay you for your work," see "Artwork for Exposure - when is it a good deal?", on https://www.reddit.com/r/ArtistLounge, April 8 2019.) Social media managers suggest that even small brands can successfully incentivize creators through exposure (Somerfield, 2023).

[23] If brand followers have heterogeneous tastes over content and take the collaborative post as a signal of the creator's typical content, then individuals who inspect the creator's account after seeing the collaborative post will be less likely to follow the more horizontally differentiated the typical content is from the initial post.

the brand's typical post ambiguous; however, the empirical finding will inform us which factor dominates. Lastly, we hypothesize that a higher quality collaborative post will also lead to a greater benefit for the creator.

4.2 Model specification and heterogeneity variables

We modify our baseline model from Section 3.1 to allow for treatment effect heterogeneity through $\boldsymbol{Z}_j$:

$$\text{FollowerGrowth}_{jt} = \boldsymbol{\beta}'\mathbf{Z_j} \cdot f(t - t^*(j), \theta) \cdot \text{Treat}_{j,t-s} + \boldsymbol{\gamma}'\mathbf{X_{jt}} + \alpha_j + \alpha_t + \alpha_{t-t^*(j)} + u_{jt} \quad (2)$$

where $\mathbf{Z}_j$ is a vector of variables that capture brand, post, and creator attributes associated with the collaborative post of creator j.

Our brand-level hypotheses involve (i) the brand audience size, (ii) brand audience engagement, and (iii) brand posting frequency. For brand audience size, we use the number of followers of the brand at the time of data collection. For brand audience engagement, we use the likes per post divided by followers, averaged over all brand posts made in the 90 days prior to the collaborative post.[24] For brand posting frequency, we use the average number of posts-per-day made by the official brand account in the 90 days prior to the collaborative post.

Our post-level hypotheses concern (i) the difference between the post content and the creator's typical content (PostCreatorDiff), (ii) the difference between the post content and the brand's typical content (BrandPostDiff), and (iii) the post quality (PostRelLikes). To operationalize the difference between the post content and the creator's content, we compute the average cosine distance between the SBERT embeddings for the collaborative post and embeddings for the brand's posts in the 90 days prior to the collaborative post. Similarly, for the difference between the post content and the brand's typical content, we take the average cosine distance between the collaborative post embeddings and the embeddings for the creator's posts in the 90 days prior to collaborative post. For post-level quality, we compute a measure of within-brand relative likes for the collaborative content. For each brand, we take the logged, Z-transformed distribution of likes for all of the brand's posts in the 90 days prior to the collaborative post. We then perform the same Z-transformation on the logged likes for the brand's collaborative post, which yields

[24] We measure post likes and brand followers in August 2024. While a post could gain new likes over time, in practice most posts on Instagram accrue over 95% of their lifetime likes within a week (Vassio et al., 2021).

the relative performance of the collavoative post compared to the brand's typical content.[25]

Finally, we also include two creator-level attributes as additional controls. Note that if brands pick creators systematically, creator attributes are likely correlated with brand and post attributes, and this in turn may affect the exposure benefits. We therefore control for creator's size (CreatorSize) and engagement rate (CreatorEng). To capture creator size, we use the number of followers 7 days prior to the collaborative post. For creator engagement, we use the likes per post divided by number of followers for non-collaborative posts in the 90 days prior to the collaborative post.

Each variable included in $\mathbf{Z}_j$ (except for relative likes, which we explain above) is transformed by first taking the logs, and then normalized by subtracting the mean and dividing by the standard deviation of that variable. The interpretation of coefficients in Equation (2) will be that a one-standard deviation in the k^{th} element of $\mathbf{Z}_j$ changes the cumulative increase in followers by $100 \cdot \beta_k\%$ relative to the baseline effect.

4.3 Heterogeneity results

In Table 4, we present the results from estimating Equation (2). The specification in Column 1 uses the daily growth rate in followers as the dependent variable, while Column 2 uses daily change in number of followers as the dependent variable.

The average estimated baseline lift in Row 1 of Column 1 is similar to that in Row 1, Column 2 of Table 3. Since almost all interaction variables are mean zero by construction and we include controls in Table 4, the baseline lift in this specification continues to capture the average effect. The Poisson distribution parameter that governs decay remains robust (see Column 1), as it is not statistically significantly different from its estimate in the baseline regression.

With respect to brand attributes, we first find that a one standard deviation increase in the log number of brand followers leads to a 0.53% greater increase in cumulative follower growth from collaborating with the brand, all else equal. As expected, larger audiences translate into a larger benefit for the creator from working with the brand. Similarly to brand audience size, brand audience engagement also positively moderates treatment effects, which reflects the benefits to the creator of an attentive audience. Brand post frequency has a negative but insignificant effect.

With respect to post-level hypotheses, the estimate of PostBrandDiff is very small and in-

[25]Note that if collaborative posts tend to receive very high engagement, then this variable will not be mean zero across brands, since the normalization happens within-brand, not across brands. In our sample, we find that this variable is still centered approximately at zero.

Table 4: Treatment effect heterogeneity of the collaborative posts

	Dependent variable:	
	Follower Growth Rate	Follower Growth
	(1)	(2)
Treat	0.0046	100.2238
	(0.0012)	(54.3973)
	[p = 0.0001]	[p = 0.0654]
Treat×BrandSize	0.0053	177.4497
	(0.0021)	(59.8018)
	[p = 0.0133]	[p = 0.0030]
Treat×BrandEng	0.0056	145.1590
	(0.0020)	(56.6232)
	[p = 0.0051]	[p = 0.0104]
Treat×BrandPostFreq	-0.0017	-79.7337
	(0.0024)	(82.2380)
	[p = 0.4787]	[p = 0.3323]
Treat×PostBrandDiff	0.0005	8.0461
	(0.0014)	(55.3964)
	[p = 0.7210]	[p = 0.8845]
Treat×PostCreatorDiff	-0.0032	-109.2352
	(0.0016)	(46.5947)
	[p = 0.0455]	[p = 0.0191]
Treat×PostRelLikes	0.0044	159.2320
	(0.0011)	(48.3917)
	[p = 0.0001]	[p = 0.0010]
Treat×CreatorSize	-0.0072	-71.3467
	(0.0019)	(34.8342)
	[p = 0.0002]	[p = 0.0405]
Treat×CreatorEng	-0.0020	-56.9748
	(0.0012)	(42.1404)
	[p = 0.0956]	[p = 0.1764]
θ	1.4518	1.6246
	(0.3073)	(0.2942)
	[p = 0.0000]	[p = 0.0000]
Creator FE	✓	✓
Date FE	✓	✓
Time-To-Post FE	✓	✓
Cross-Posting Controls	✓	✓
Avg. Value of Dep. Var.	0.0008	44.253
Observations	49,097	49,097
R^2	0.2064	0.2537

Note: This table presents estimates from Equation (2). In Column 1, the dependent variable is the daily follower growth rate. The estimated Treat coefficient, when multiplied by 100, is the cumulative percent increase in followers (relative to just before treatment) due to the collaborative post for a post with all attributes at their normalized averages of zero. In Column 2, the dependent variable is follower growth. The estimated Treat coefficient is the cumulative increase in the number of followers (relative to just before treatment) due to the collaborative post for a post with all attributes at their normalized averages of zero. Interaction coefficients are the change in the treatment effect relative to the average post for a one standard deviation increase in the given heterogeneity variable, holding all other heterogeneity variables constant, see Section 4.2 for definitions. θ is the Poisson parameter and governs the shape of the treatment effect decay over time. Cross-posting controls are a vector of dummy variables capturing if a brand posted about the creator to TikTok, Facebook, or Twitter/X, with 7 days of lagged dummies included for each platform. Standard errors are clustered at the creator level and reported in (.) below the estimated coefficients. Exact p values are reported in [.] below the standard errors.

significant. This may occur because while posts that diverge from a brand's typical content style can attract attention, they may fail to resonate with the target audience. Meanwhile, PostCreatorDiff, which measures the dissimilarity between the collaborative post and creator's typical posts, leads to lower lift. This finding enriches our earlier observation that collaborative posts may lead to low conversions since they are not optimized to promote the creator. If a collaborative post fails to spotlight the creator's content, then potential followers cannot discern the creator's identity and value proposition; moreover, the creator's existing followers may perceive the content as inauthentic (Cheng and Zhang, 2025). Finally, PostRelLikes is associated with higher lift. Intuitively, high engagement captures residual attributes of the collaborative post that the audience values, such as quality or trendiness, which may reflect positively on the creator.[26]

Turning to the creator controls, we find that CreatorSize strongly negatively moderates the benefit from exposure in Column 1. Note that this negative moderating effect could be mechanical because we are looking at the percentage change in followers. However, in Column 2 where we use "follower growth" as the dependent variable, the relationship remains negative, and the magnitude implies that a creator two standard deviations larger than the average earns essentially no new followers from a collaborative post. This may arise because of diminishing returns of marginal exposure benefits in the creator's follower size; for example, because the creator is already relatively well-known. Interestingly, the estimate for Treat$\times$CreatorEng is negative, suggesting that more engaging creators may benefit less from exposure. It is possible that when a creator's followers are very engaged, they may especially dislike any brand collaboration and perceive those posts as inauthentic, leading to a greater backlash from working with a brand.

4.4 Unconditional effect of brand size

Although all else equal we find that larger brands lead to more exposure benefits, our estimates in Table 4 suggest that there may be countervailing forces to size. For instance, Instagram accounts with more followers tend to have lower engagement—Nike, with 300 million followers, only has a 0.03% engagement rate—and we find that engagement is positively related to demand lift.[27] We thus explore what treatment effect a creator can actually expect from working with brands of different sizes, if we allow brand and post level attributes to co-move with brand size as they do

[26] High engagement can also lead a post to remain in brand followers' feeds for longer (SocialBee, 2025), which creates more opportunities to view and follow the creator.

[27] See https://starngage.com/app/global/influencer/ranking/all, accessed 2025-05-07.

Table 5: Correlation matrix: heterogeneity variables

	Brand Size	Brand Eng	Brand PostFreq	Post BrandDiff	Post CreatorDiff	Post RelLikes	Creator Size	Creator Eng
BrandSize	1							
BrandEng	-0.357	1						
BrandPostFreq	0.168	-0.292	1					
PostBrandDiff	0.028	0.087	-0.129	1				
PostCreatorDiff	0.016	0.071	-0.118	0.387	1			
PostRelLikes	-0.245	-0.151	0.066	0.005	-0.064	1		
CreatorSize	0.181	-0.022	-0.044	0.052	-0.025	0.130	1	
CreatorEng	0.069	0.184	-0.085	-0.024	-0.019	0.121	-0.117	1

Note: BrandSize is the log normalized number of brand followers, BrandEng is the log normalized ratio of likes to followers for brand posts in the 90 days prior to the collaborative post, BrandPostFreq is the log normalized daily frequency of brand posting in the 90 days prior to the collaborative post, PostBrandDiff is the normalized difference between the collaborative post content and brand post content in the 90 days prior to the collaborative post, PostCreatorDiff is the normalized difference between the collaborative post content and creator post content in the 90 days prior to the collaborative post, PostRelLikes is the engagement of the collaborative post normalized relative to the brand's non-collaborative posts in the 90 days prior to the collaborative post, CreatorSize is the log normalized number of creator followers, and CreatorEng is the log normalized ratio of likes to followers for creator posts in the 90 days prior to the collaborative post. Normalization is Z-transformation across collaborative posts.

in the sample.

In Table 5, we check the correlation between brand size and the other heterogeneity attributes. In the first column, it is evident that brands with larger follower counts indeed tend to have lower average engagement rates (BrandEng), with a correlation of -0.36. They also tend to post more frequently (BrandPostFreq), with a correlation of 0.17, likely reflecting their greater resources. Collaborative posts with larger brands also tend to be more different to the brand's usual content (PostBrandDiff) and to the creator's usual content (PostCreatorDiff) although both correlations are small in magnitude.[28] Interestingly, collaborative posts perform worse on engagement metrics for large brands than for small brands: the PostRelLikes variable, which captures the relative performance (in likes) for a brand's collaborative posts compared to its recent non-collaborative posts, has a negative correlation of -0.24 with brand size. This correlation suggests that large brands may be systematically worse at translating collaborations with creators into engaging content for their audience. We also find that larger brands tend to select larger, more engaging creators for collaborations, with positive correlations between brand size and creator size (CreatorSize, 0.181), and brand size and creator engagement (CreatorEng, 0.069).

[28]The positive correlation between BrandSize and PostCreatorDiff suggests that larger brands may exercise more restraints on creators than smaller brands (Libai et al., 2025).

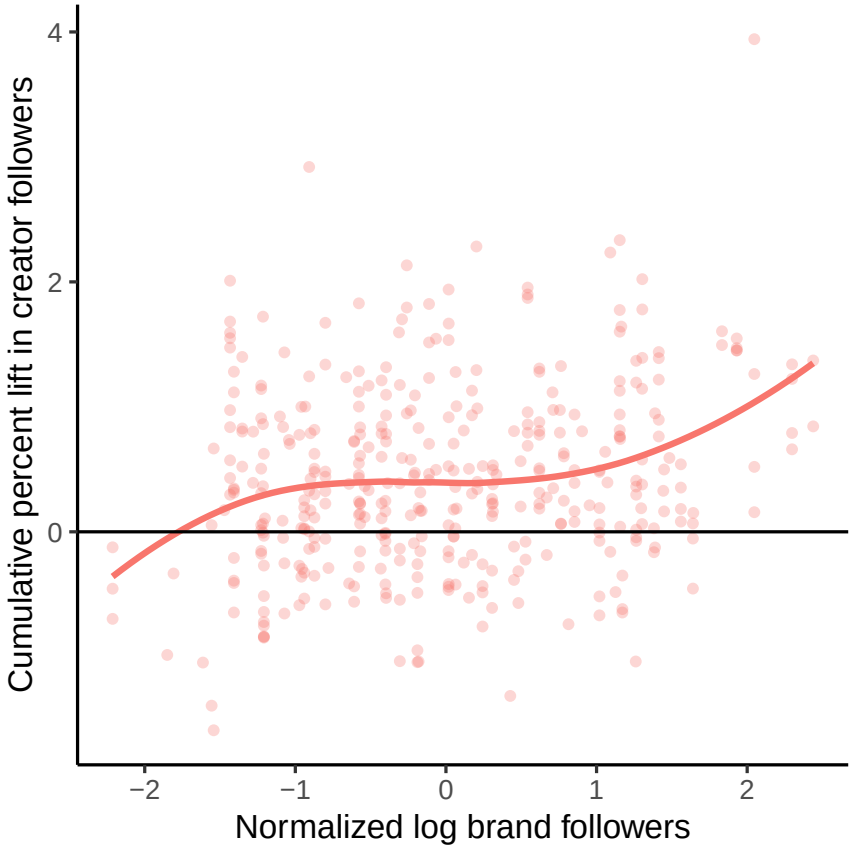


Figure 6: Treatment effects by brand size

Note: Each point is the estimated empirical treatment effect for a brand-post multiplied by 100, which we compute by combining each vector of actual brand-post attributes in the data with the estimated heterogeneous treatment effects in Column 1 of Table 4, holding CreatorSize and CreatorEng fixed at their average values of zero (given the Z-transformation). The smooth curve is a local linear approximation to these estimated treatment effects. This plot shows that for brands between -1 and +1 standard deviation of log followers in our data, the predicted percent increase in cumulative followers from a collaborative post is flat for the average creator.

In Figure 6 we plot the estimated treatment effect for each collaborative post in the data holding creator attributes fixed. That is, each point is obtained by plugging in an observed vector of brand-post attributes into the estimated $\hat{\boldsymbol{\beta}}'\mathbf{Z}_j$, while keeping the normalized creator size and engagement values at their average (zero). We then fit a smoothed local linear curve to these points. We find that for a large part of the distribution of brand sizes, from roughly 1 standard deviation below to 1 standard deviation above the mean log followers, the average relationship between brand followers and the exposure benefit to the creator is essentially flat at 0.43% cumulative lift. Translating the normalized brand size values back into brand follower counts, we find that working with brands between 0.7m followers and 8.6m followers leads to a roughly constant exposure benefit, although the largest (smallest) brands still drive higher (lower) follower growth.

4.5 Distribution of follower growth

In this section, we present the full empirical distribution of treatment effects. This analysis is intended to assist creators, influencer agencies, and brands in negotiating and evaluating compensation packages.

Using our estimates from Table 4, we plot the distribution of the increase in the number of followers in Figure 7 by substituting the observed values of the $\mathbf{Z}_j$ heterogeneity vector into $\hat{\boldsymbol{\beta}}'\mathrm{Z}_j$ for each collaborative post and then plotting a histogram of the resulting treatment effects.

A key insight from Figure 7 is that we do not estimate a skewed right tail of large growth events. Even for the top 10% of collaborations, the average increase in followers is only around 2%. This finding is important: many creators may not be overly concerned with the average effect of collaboration, but are focused on tail events that would lead to explosive growth. However, we find that even in the right tail of the distribution there may not be outsize gains from exposure.

To convert this distribution of follower growth into a distribution of monetary gains would require assumptions on how creators monetize their followers. We leave this for future work, as there are many ways for creators to value their followers; for example, creators may post sponsored ads for them to view, sell them their creative services (such as artwork), or creators may simply gain utility from the attention they receive. Our follower growth estimates can be combined with creators' subjective estimates of their valuation of followers, so that creators make more informed decisions about collaboration opportunities with brands and negotiate transaction terms that reflect their specific circumstances.

5 Implications for Practice

Our work offers three practical implications. First, it sheds light on an ongoing debate about the value of social media exposure to creators. More specifically, creators are divided on their preferred form of compensation: some favor direct payment over exposure,[29] while others, along with many social media marketers, believe the exposure brands offer holds significant value (Somerfield, 2023; LinkedIn Contributors, 2023). Our findings suggest that these two sets of beliefs potentially arise from selective attention to different parts of the distribution of treatment

[29] See e.g. https://www.helloraderco.com/sacrificing-time-money-exposure-bucks/, accessed April 20, 2025, and https://amandajspedding.com/2019/05/27/exposing-exposure-bucks-hint-its-not-a-real-thing/, accessed April 20, 2025.

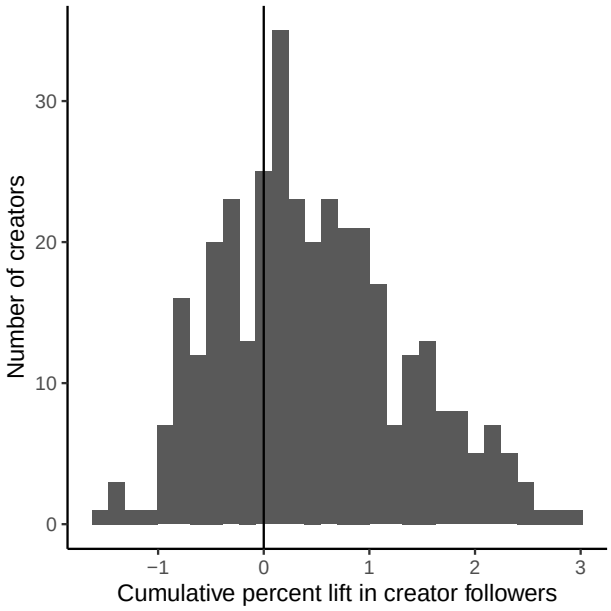


Figure 7: Distribution of treatment effects of collaboration

Note: Each observation is the estimated empirical treatment effect for a brand-post-creator, which we compute by combining each vector of actual brand-post-creator attributes in the data with the estimated heterogeneous treatment effects in Column 1 of Table 4. We then create the histogram of these observations after multiplying the estimated treatment effects by 100.

effects. It is possible that, as with influencer marketing (Huang and Morozov, 2025), brands may only consider examples from the upper tail of the distribution of treatment effects, leading them to have a false sense of how valuable exposure is. On average, for the set of creators and content we consider in this paper, exposure has a positive but small benefit for creators. Larger brand audiences are better for the creator all else equal, but because brands with large audiences have systematically lower engagement rates and collaborative post quality, we find that follower growth from exposure is essentially flat for the creator over a substantial range of intermediate brand audience sizes.

Second, we provide novel insights for influencers—or influencer agencies who negotiate contracts on behalf of influencers—on how they should value the offer of collaborations from brands. Influencers who hope to experience explosive growth in followers from collaborating with brands should moderate their expectations: even the top 10% of follower gains in our estimated distribution in Figure 7 only lead to a cumulative bump of 2.1% new follower growth. Our findings also suggest that when negotiating with brands, creators should strive to keep their content close to what they typically create if they want to enjoy a larger benefit from exposure. Moreover, given that larger brands tend to have collaborative posts that underperform in lifting follower

growth for creators relative to their typical posts, creators may wish to require higher monetary compensation in their contracts with large brands.

Finally, our study contributes novel empirical findings to the literature on the theory of influencer marketing. While prior work conceptualizes influencers as first building audiences through organic content before monetizing via sponsorships (Fainmesser and Galeotti, 2021; Mitchell, 2021), we demonstrate an alternative pathway: collaborations can directly generate follower growth through brand audience exposure. We find that the viability of this approach depends strongly on the creator's size, since relatively large creators gain almost no new followers from collaborations. Our findings suggest potentially different dynamic incentives for creators to establish themselves, with smaller and newer creators having more to gain from working with brands.

6 Conclusion

Soliciting content from creators in exchange for exposure is a common business tactic, but the value of exposure has been difficult to quantify. In this paper, we tackle this problem in the context of creators on Instagram. We collect a new, novel dataset that covers collaborations between the largest fashion and beauty brands and relatively small creators, where the creators' work is shown to the brands' audiences in a one-off collaborative post. We find that as a result of the exposure from this collaborative post, creators gain followers, but the magnitude does not vary much across most of the distribution of brand follower counts. The average conversion rate for brand audience members to start following the featured creator is in line with estimates of conversion rates in display advertising.

Our research raises several questions for future work. We show that the effect of exposure depends strongly on audience attention—it is thus possible that in live or televised settings with more attentive audiences, exposure may have much larger effects. Moreover, a key next step is to estimate how creators value larger audiences, in order to understand the dollar value of "exposure bucks" and compare it to alternatives like a direct payment to supply content. While we have quantified the value of exposure in terms of the lift in followers on Instagram, it remains an open question as to how valuable exposure is relative to other elements of being recognized by brands like prestige or feelings of community. Understanding which factors are most effective in motivating creators can help brands craft their messages for soliciting creative content.

References

Abidin, Crystal, "Visibility labour: Engaging with Influencers' fashion brands and# OOTD advertorial campaigns on Instagram," *Media International Australia*, 2016, *161* (1), 86–100.

Angrist, Joshua D and Alan B Krueger, "Empirical Strategies in Labor Economics. Handbook of Labor Economics, Volume 3A. o. Ashenfelter and D. Card, eds," 1999.

Arsel, Zeynep, Maria Carolina Zanette, and Carolina da Rocha Melo, "Sponsored Content as an Epistemic Market Object: How Platformization of Brand-Creator Partnerships Disrupts Valuation, Co-production, and the Relationship between Market Actors," *Journal of Marketing*, 2024.

Athey, Susan, Markus Mobius, and Jeno Pal, "The impact of aggregators on internet news consumption," Technical Report, National Bureau of Economic Research 2021.

Bairathi, Mimansa and Anja Lambrecht, "Influencer Marketing: Sponsorship Disclosure and Authenticity," *Available at SSRN 4679776*, 2023.

Beichert, Maximilian, Andreas Bayerl, Jacob Goldenberg, and Andreas Lanz, "Revenue generation through influencer marketing," *Journal of Marketing*, 2024, *88* (4), 40–63.

Bertaglia, Thales, Catalina Goanta, Gerasimos Spanakis, and Adriana Iamnitchi, "Influencer self-disclosure practices on Instagram: A multi-country longitudinal study," *Online Social Networks and Media*, 2025, *45*, 100298.

Berthon, Pierre, Leyland Pitt, and Colin Campbell, "Ad lib: When customers create the ad," *California Management Review*, 2008, *50* (4), 6–30.

Bronnenberg, Bart J, Jean-Pierre Dubé, and Robert E Sanders, "Consumer misinformation and the brand premium: A private label blind taste test," *Marketing Science*, 2020, *39* (2), 382–406.

Burtch, Gordon, Qinglai He, Yili Hong, and Dokyun Lee, "How do peer awards motivate creative content? Experimental evidence from Reddit," *Management Science*, 2022, *68* (5), 3488–3506.

Callaway, Brantly and Pedro HC Sant'Anna, "Difference-in-differences with multiple time periods," *Journal of Econometrics*, 2021, *225* (2), 200–230.

Calzada, Joan and Ricard Gil, "What do news aggregators do? Evidence from Google News in Spain and Germany," *Marketing Science*, 2020, *39* (1), 134–167.

Campbell, Colin, Jeffrey M Girard, Daniel McDuff, and Sara Rosengren, "EXPRESS: How Influencers Grow: An Empirical Study and Future Research Agenda," *Journal of Interactive Marketing (Forthcoming)*, 2025.

Cheng, Mengjie and Shunyuan Zhang, "Reputation burning: Analyzing the impact of brand sponsorship on social influencers," *Management Science*, 2025, *71* (7), 5910–5932.

Ching, Andrew T and Daniel Goetz, "Consumption Responses to an Unpopular Policy: Evidence from a Short-Lived Soda Tax," *Marketing Science*, 2024, *43* (6), 1204–1225.

Convertcart, "27 Brilliant User-Generated Content Examples (eCommerce)," 2024. Published: January 11, 2024. Accessed May 5, 2025.

Ershov, Daniel and Matthew Mitchell, "The effects of advertising disclosure regulations on social media: Evidence from instagram," *The RAND Journal of Economics*, 2025, *56* (1), 74–90.

_ , Yanting He, and Stephan Seiler, "Frontiers: How Much Influencer Marketing Is Undisclosed? Evidence from Twitter," *Marketing Science*, 2025, *44* (3), 505–515.

Fainmesser, Itay P and Andrea Galeotti, "The market for online influence," *American Economic Journal: Microeconomics*, 2021, *13* (4), 332–72.

Forbes, "Why User-Generated Content Is Winning," https://www.forbes.com/councils/forbesbusinesscouncil/2023/03/13/why-user-generated-content-is-winning/ 2023. Published: March 13, 2023. Accessed: April 14, 2025.

Gallus, Jana, "Fostering public good contributions with symbolic awards: A large-scale natural field experiment at Wikipedia," *Management Science*, 2017, *63* (12), 3999–4015.

Ghosh, Arpita and Preston McAfee, "Incentivizing high-quality user-generated content," in "Proceedings of the 20th international conference on World wide web" 2011, pp. 137–146.

Goes, Paulo B, Chenhui Guo, and Mingfeng Lin, "Do incentive hierarchies induce user effort? Evidence from an online knowledge exchange," *Information Systems Research*, 2016, *27* (3), 497–516.

Goodman-Bacon, Andrew, "Difference-in-differences with variation in treatment timing," *Journal of Econometrics*, 2021, *225* (2), 254–277.

Gordon, Brett R, Florian Zettelmeyer, Neha Bhargava, and Dan Chapsky, "A comparison of approaches to advertising measurement: Evidence from big field experiments at Facebook," *Marketing Science*, 2019, *38* (2), 193–225.

Hayashi, Fumio, *Econometrics*, Princeton University Press, 2011.

Huang, Yufeng and Ilya Morozov, "The Promotional Effects of Live Streams by Twitch Influencers," *Marketing Science*, February 2025. Forthcoming.

Influee, "UGC Ads: Create Ads That Work (+ Examples)," https://influee.co/blog/ugc-ads 2025. Published: June 11, 2025. Accessed: July 25, 2025.

Kanuri, Vamsi K, Christian Hughes, and Brady T Hodges, "Standing out from the crowd: When and why color complexity in social media images increases user engagement," *International Journal of Research in Marketing*, 2024, *41* (2), 174–193.

Khernamnuai, Warut, Karthik Kannan, and Hossein Ghasemkhani, "Extrinsic versus intrinsic rewards for contributing reviews in an online platform," *Information Systems Research*, 2018, *29* (4), 871–892.

Kumar, Ashish, Ram Bezawada, Rishika Rishika, Ramkumar Janakiraman, and PK Kannan, "From social to sale: The effects of firm-generated content in social media on customer behavior," *Journal of Marketing*, 2016, *80* (1), 7–25.

Lacka, Ewelina, D Eric Boyd, Gbenga Ibikunle, and PK Kannan, "Measuring the real-time stock market impact of firm-generated content," *Journal of Marketing*, 2022, *86* (5), 58–78.

Li, Nan, Avery Haviv, and Mitchell J Lovett, "Opposing influences of YouTube influencers: Purchase and usage effects in the video game industry," *Marketing Science*, 2025.

Li, Yiyi and Ying Xie, "Is a picture worth a thousand words? An empirical study of image content and social media engagement," *Journal of Marketing Research*, 2020, *57* (1), 1–19.

Liadeli, Georgia, Francesca Sotgiu, and Peeter WJ Verlegh, "A meta-analysis of the effects of brands' owned social media on social media engagement and sales," *Journal of Marketing*, 2023, *87* (3), 406–427.

Liaukonyte, Jura, Thales Teixeira, and Kenneth C Wilbur, "Television advertising and online shopping," *Marketing Science*, 2015, *34* (3), 311–330.

Libai, Barak, Ana Babić Rosario, Maximilian Beichert, Bas Donkers, Michael Haenlein, Reto Hofstetter, PK Kannan, Ralf van der Lans, Andreas Lanz, H Alice Li et al., "Influencer marketing unlocked: Understanding the value chains driving the creator economy," *Journal of the Academy of Marketing Science*, 2025, pp. 1–25.

LinkedIn Contributors, "What are some creative ways to incentivize and reward user-generated content creators?," https://www.linkedin.com/advice/1/what-some-creative-ways-incentivize-reward-user-generated 2023. Accessed: April 14, 2025.

Lu, Shijie, Isaac Dinner, and Rajdeep Grewal, "The ripple effect of firm-generated content on new movie releases," *Journal of Marketing Research*, 2023, *60* (5), 908–931.

__, Ying Xie, and Xingyu Chen, "Immediate and enduring effects of digital badges on online content consumption and generation," *International Journal of Research in Marketing*, 2023, *40* (1), 146–163.

Lysyakov, Mikhail, PK Kannan, Siva Viswanathan, and Kunpeng Zhang, "Retailer Differentiation in Social Media: An Investigation of Firm-Generated Content on Twitter," *Journal of Marketing*, 2025, *89* (4), 39–58.

McGranaghan, Matthew, Jura Liaukonyte, and Kenneth C Wilbur, "How viewer tuning, presence, and attention respond to ad content and predict brand search lift," *Marketing Science*, 2022, *41* (5), 873–895.

Mitchell, Matthew, "Free ad (vice): Internet influencers and disclosure regulation," *The RAND Journal of Economics*, 2021, *52* (1), 3–21.

Nam, Hyoryung, Yogesh V Joshi, and PK Kannan, "Harvesting brand information from social tags," *Journal of Marketing*, 2017, *81* (4), 88–108.

Pei, Amy and Dina Mayzlin, "Influencing social media influencers through affiliation," *Marketing Science*, 2022, *41* (3), 593–615.

Rajaram, Prashant and Puneet Manchanda, "Unboxing engagement in YouTube influencer videos: An attention-based approach," *arXiv preprint arXiv:2012.12311*, 2020.

Reimers, N, "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks," *arXiv preprint arXiv:1908.10084*, 2019.

Roberts, Deborah, Mathew Hughes, and Kia Kertbo, "Exploring consumers' motivations to engage in innovation through co-creation activities," *European Journal of Marketing*, 2014, *48* (1/2), 147–169.

Roth, Jonathan, Pedro HC Sant'Anna, Alyssa Bilinski, and John Poe, "What's trending in difference-in-differences? A synthesis of the recent econometrics literature," *Journal of Econometrics*, 2023.

Sen, Ananya, Tom Grad, Pedro Ferreira, and Jörg Claussen, "(How) does user-generated content impact content generated by professionals? Evidence from local news," *Management Science*, 2024, *70* (9), 6045–6068.

Simonyan, Karen, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

Smith, Andrew N and Eileen Fischer, "Pay attention, please! Person brand building in organized online attention economies," *Journal of the Academy of Marketing Science*, 2021, *49*, 258–279.

Social Media Strategies Summit, "12 Ways to Incorporate Your User-Generated Content on Different Channels," https://blog.socialmediastrategiessummit.com/4-ways-to-incorporate-your-user-generated-content-on-different-channels/ 2025. Published: September 28, 2025. Accessed: July 25, 2025.

SocialBee, "How the Instagram Algorithm Works in 2025: The Complete Guide," https://socialbee.com/blog/instagram-algorithm/ 2025. Published: January 6, 2025. Accessed: April 21, 2025.

Somerfield, Kardi, "Sharing the love: Motivation and reward in brand-related user-generated content creation for social media," *Unpublished Manuscript*, 2023.

_ , Kathleen Mortimer, and Geraint Evans, "The relevance of images in user-generated content: A mixed method study of when, and why, major brands retweet," *International Journal of Internet Marketing and Advertising*, 2018, *12* (4), 340–357.

Stephen, Andrew T and Jeff Galak, "The effects of traditional and social earned media on sales: A study of a microlending marketplace," *Journal of Marketing Research*, 2012, *49* (5), 624–639.

Sun, Yacheng, Xiaojing Dong, and Shelby McIntyre, "Motivation of user-generated content: Social connectedness moderates the effects of monetary rewards," *Marketing Science*, 2017, *36* (3), 329–337.

Taylor, Evelyn, "10 User Generated Content Campaigns That Actually Worked [+Expert Insights]," 2023. Published: November 11, 2023. Accessed: May 5, 2025.

Tian, Zijun, Ryan Dew, and Raghuram Iyengar, "Mega or micro? Influencer selection using follower elasticity," *Journal of Marketing Research*, 2024, *61* (3), 472–495.

Vassio, Luca, Michele Garetto, Carla Chiasserini, and Emilio Leonardi, "Temporal dynamics of posts and user engagement of influencers on Facebook and Instagram," in "Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining" 2021, pp. 129–133.

Villanueva, Julian, Shijin Yoo, and Dominique M Hanssens, "The impact of marketing-induced versus word-of-mouth customer acquisition on customer equity growth," *Journal of Marketing Research*, 2008, *45* (1), 48–59.

Wies, Simone, Alexander Bleier, and Alexander Edeling, "Finding goldilocks influencers: How follower count drives social media engagement," *Journal of Marketing*, 2023, *87* (3), 383–405.

Yang, Jeremy, Juanjuan Zhang, and Yuhan Zhang, "Engagement that sells: Influencer video advertising on TikTok," *Marketing Science*, 2025, *44* (2), 247–267.

Zhang, Shunyuan, Dokyun Lee, Param Vir Singh, and Kannan Srinivasan, "What makes a good image? Airbnb demand analytics leveraging interpretable image features," *Management Science*, 2022, *68* (8), 5644–5666.

Web Appendix

A Data construction

A.1 Tagging and Sponsored Posts

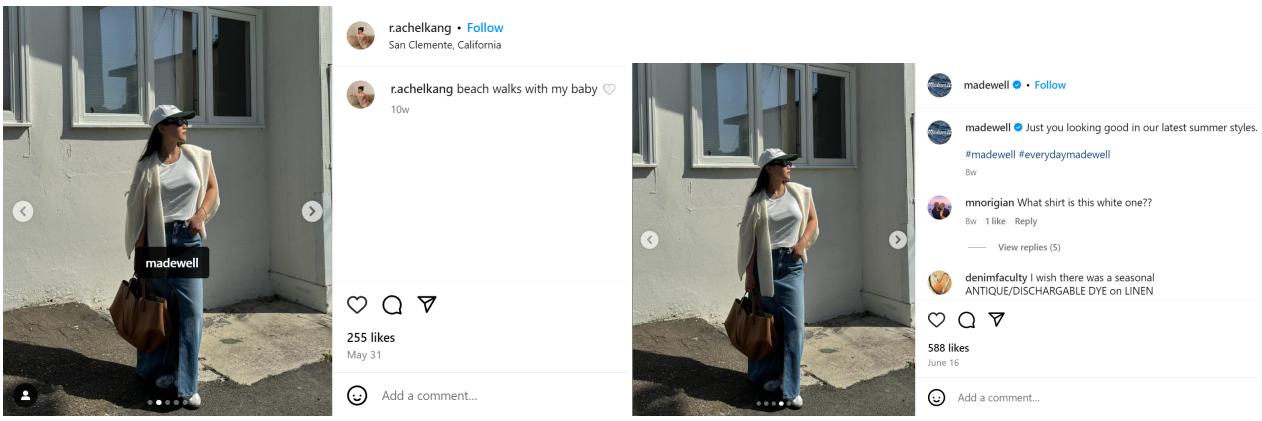
A.1.1 Tagging

Tagging is an alternative way that individuals on Instagram can credit other accounts in a post. As an example of tagging, in Figure A.1a, the content creator **rachelkang** tags the account **madewell**, which is a clothing company. The tagged company account name **madewell** is visible on this photo after a user taps it, and tapping the name will take the user directly to the brand account page.

Compared to collaborative posts, tagged posts have several disadvantages which make them less amenable to a study of the value of exposure. First, and as can be seen in Figure A.1a, tags are not always made in recognition of contributed content—**rachelkang** tagged **madewell** because she wanted to gain the brand's attention. Indeed, a tag can be made without the tagged account's consent. When examining brand posts, this property of tagging makes it very hard from a data perspective to parse which accounts that a brand tagged actually contributed content. Conversely, in collaborative posts, explicit consent of each collaborating party is required before that party's account name appears on the post, so it is much more likely that the collaborative post reflects a contribution.

Second, if a brand tags a creator in recognition of contributed content, that tagged content on the brand account often duplicates an image or video the creator posted elsewhere. For instance, in Figure A.1a, the account owner of **rachelkang** posts an image of herself wearing **madewell** jeans and tags the brand account, and the brand then posts that content to their page in Figure A.1b, likely for free but with **rachelkang**'s consent. However, the Instagram algorithm tends not to promote duplicated content, which would lead to less exposure of that content to **madewell**'s audience.[30] Compared with re-used tagged content, collaborative posts are not duplicates, but original content posted to both accounts simultaneously, which obviates this algorithmic bias.

[30] See https://creators.instagram.com/blog/recommendations-and-originality, accessed May 15, 2025.



(a) Original photo

(b) Photo with creator tag on brand account

Figure A.1: Original and tagged UGC on Instagram

Note: The left image is the original photo, posted by the creator rachelkang on her Instagram account and tagging madewell, the clothing brand. The tag brings the post to madewell's attention, and they repost the content on their brand page in the right image and tag rachelkang (not visible). Tagging is an alternative way of crediting content contributions compared to collaborations.

Third, tagged content can often be UGC submitted with a hashtag or brand tag in response to a contest. For example, in Apple's "Shot on iPhone" campaign in 2019, users were encouraged to use the hashtag #ShotOniPhone to bring photos to Apple's attention. Apple then selected photos with the hashtag to use as part of a marketing campaign (see iPhone challenge rules). If the contest hashtag is trending, then unobserved greater interest in this content will lead the algorithm to promote content connected to the hashtag—including the selected photo—which will contaminate measurement of the exposure effect. Conversely, collaborative posts tend to be one-off and not part of trending hashtags, which provides a cleaner quasi-experiment to identify the effect of exposure.

A.1.2 Sponsored posts

Sponsored posts are posts made by the creator, which tag or mention the brand, and are typically paid endorsements of the brand. A specific type of sponsored post is the "partnership ad" (renamed from branded content ad in May 2023), in which the creator can allow the brand to have access to the sponsored post for purposes of "boosting" the post. Boosting is a paid service that Instagram offers to business accounts whereby the post is promoted in the Instagram algorithm to gain additional views.

There are two disadvantages that partnership ads have relative to collaborative posts for a

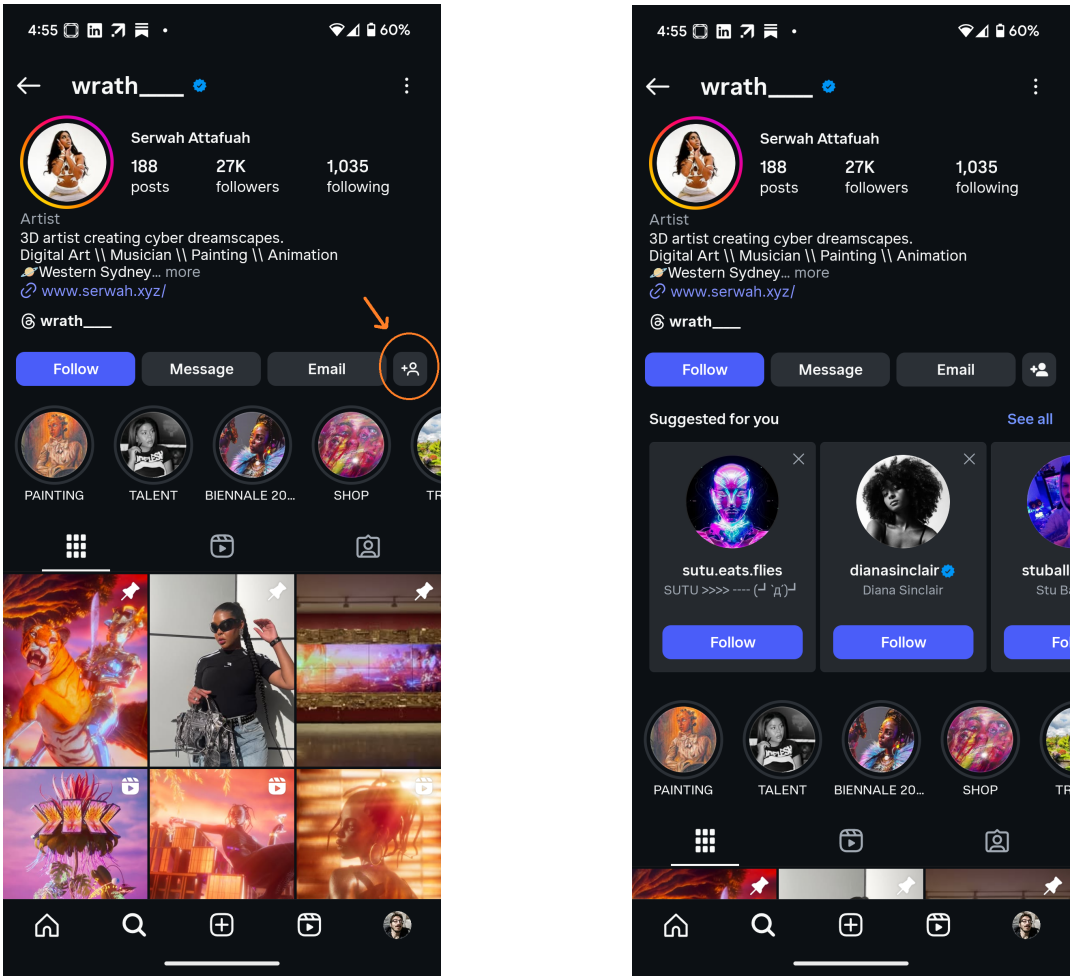
study of the benefits of exposure. First, partnership ads are only shown to the creator's followers by default, not the brand's followers, and thus a partnership ad may not provide any new exposure for the creator. If the partner brand chooses to boost the post then it will be shown to a wider audience. However, the audience depends on the amount of money the brand spends on the boost, which is unobservable. Second, evidence suggests that the vast majority of sponsored posts go undisclosed (e.g., Bertaglia et al., 2025; Ershov et al., 2025; Ershov and Mitchell, 2025), which presents an additional challenge for identifying their effect on a creator's follower count.

A.2 Control pool of posts

We source additional control posts by never-treated creators in several steps. First, we generate a list of Instagram account names. Instead of randomly sampling from the universe of Instagram accounts, we use Instagram's recommendation algorithm to suggest similar accounts for each treated creator. In Figure A.2a, we showcase what an account details page looks like for a sample creator, with the "Similar Accounts" button highlighted. In Figure A.2b, we show the same details page after having tapped the button.

The similar accounts button lists up to 80 accounts per creator. While it is difficult to source information on how Instagram generates these similar accounts, anecdotally they are similar in content to the main account. We gather 16,467 unique account names using this approach. We then drop all accounts that are categorized as businesses, yielding 10,013 non-business accounts as candidate never-treated creators.

Gathering post histories and follower counts for such a large sample of accounts is prohibitive. We therefore select two unique similar accounts per treated creator that belong to the same category (e.g. artist, athlete, creator, digital artist, etc.) and have a similar number of followers at time of data collection. Specifically, we discretize the full distribution of follower counts into 20 intervals, and select accounts that are in the same interval as the original treated account to which they are related. If there are more than two similar accounts with the same category and in the same follower interval as the original treated account, then we randomly select two. This procedure guarantees that the control posts we gather will be made by creators covering similar topics, and with a similar number of followers, as the collaborative posts by treated creators. Note that we are not explicitly matching control creators to treated creators, but rather selecting accounts that have the best chance of supplying control posts that strongly resemble the collaborative posts, by similarly sized creators.



(a) "Similar Accounts" button

(b) Displayed similar accounts

Figure A.2: Sourcing similar accounts for control creators

Note: The left image shows a screen capture from the Instagram platform after navigating to the details page for the account "wrath____", with the "Similar Accounts" button highlighted. The right image shows a screen capture of the same details page after having tapped the "Similar Accounts" button.

Of the similar accounts we select, we are able to collect followership history on SocialBlade for 625 accounts. We then pull the complete post history for these 625 creators, which yields approximately 220,000 non-collaborative candidate control posts. Combined with the pre-treatment control posts from the collaborating creators, we have approximately 430,000 non-collaborative posts in total from which to select matched control posts.

A.3 Post content

Before matching treatment posts to control posts, we must extract post content. We proceed by first using ChatGPT 4o's vision capabilities to describe images. We then embed these descriptions, perform a dimension reduction of the embedding, and then perform clustering, selecting an optimal number of clusters using the elbow method. Finally, we use TF-IDF after the clustering is complete to identify human-interpretable topics.

A.3.1 Using ChatGPT 4o to describe images

Media on Instagram comes in two forms: images and videos. For videos, we take snapshots every 3 seconds for the first 15 seconds. For both images and video snapshots, we reduce the resolution to a 300 x 450 pixel resolution. We then prompt ChatGPT 4o to analyze each image and video snapshot, using the prompt "Describe this image" in all cases.

To showcase the outcome of this process, we select five example images from the set of treatment (collaborative) posts in Table A.1 along with the recovered image description from ChatGPT. ChatGPT 4o is able to recover not just the number of individuals and their emotional state, but also highly-idiosyncratic individual features like exact details of their clothes (e.g. "white checkered sleeveless vest", "heart-shaped belt buckle", "loose-fitting jeans"), their pose ("kneeling", "sitting"), detailed elements of the backdrop ("a snowy landscape with trees", "a cloudy sky with a warm, golden hue"), and text that appears on the image. The description of the images also often provides holistic, human-interpretable text that either directly captures both valence and arousal ("calm and inspiring", "smiling and ... having a good time") or provides context clues that allow these features to be understood.

The total corpus of creator media comprises over 2.24 million images in total. Using ChatGPT 4o, in November 2024 the price of generating all descriptions was $645 USD.

A.3.2 Embedding and clustering image descriptions

Our aim is to cluster the images into topics, and to use these topics as representations of the content of posts so that treated post topics can be matched to control post topics.

The first step is to transform the text descriptions into numeric embeddings. While in theory it is possible to skip the step of recovering the text description from ChatGPT and directly query ChatGPT to provide numeric embeddings of the content of images, at time of writing these visual

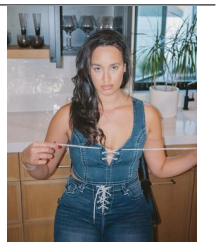
					
Post ID	CfKlh8SvyYq	Csb2EXCO62z	Ctjjv90gdyH	CuAGAqxLzO6	C0SGnBUPeeK
ChatGPT 4o description	The image features a tall stadium light pole against a backdrop of a cloudy sky with a warm, golden hue. There's a motivational quote overlaid on the image that reads, ""I've already won as long as I show up as myself."" The overall tone of the image is calm and inspiring.	The image shows two people outdoors enjoying the snow. They are smiling and appear to be having a good time, with their arms around each other. One person is wearing a hat and the other is in a T-shirt with colorful graphics. Both are wearing sunglasses, and the background features a snowy landscape with trees.	The image shows a person standing outdoors near a display of colorful flowers in pots. They are wearing sunglasses, a black and white checkered sleeveless vest, a white T-shirt, and black shorts with a heart-shaped belt buckle. The person has short, dark hair and visible tattoos on their arm and legs. The background includes various plants and a structure with a lattice roof.	The image is a black and white photograph of a person kneeling on the ground. They are wearing a cropped t-shirt and loose-fitting jeans. The individual's hands are placed behind their head. The setting appears to be minimal, with a simple, plain background.	The image shows a person with long dark hair sitting in a kitchen setting. They are wearing a denim top and jeans that both have lace-up designs. The person is holding a measuring tape. The background includes kitchen cabinets, countertops, and potted plants.

Table A.1: Example images with ChatGPT descriptions

Note: Post ID links to the post on Instagram. The link can be manually created using "https://www.instagram.com/p/" + post ID. Posts typically contain multiple images, videos, or a combination. The image shown here for each post is a random selection from within the post.

embeddings were not available to be called from the API.

To recover embeddings of the text descriptions, we use the pre-trained *paraphrase-mpnet-base-v2* sentence transformer, a transformer model with 109 million parameters also referred to as SBERT. SBERT embeddings recover contextual information at the paragraph level, unlike base BERT which analyzes text at the sentence level. This is important in our setting: for instance, in Table A.1, the final sentence in the second image starts with "Both are wearing sunglasses...", which references the "two people outdoors enjoying the snow". SBERT would correctly attribute "both" as referring to the two people outdoors since it uses the entire paragraph. For each image and each video snapshot, we recover the SBERT embedding, which is a 768-dimensional vector.

In theory we could cluster images into topics based on the SBERT embeddings. However, clustering requires holding all of the image embeddings in memory to form clusters. We have 2.24 million images, and the memory requirements are too large to perform the clustering using the 768-dimensional vector embedding. We therefore do dimension reduction. The Uniform Manifold Approximation and Projection (UMAP) is a popular technique for dimension reduction

when the end goal is to cluster data, since it works to preserve local structure. UMAP is also memory intensive, since it constructs a complete graph of the data, where each point is a node and each vector encodes local similarity between that node and other nodes, and it is infeasible to perform using the complete data set with the SBERT embeddings. We therefore first use Principal Components Analysis, a fast, linear method that identifies new axes in the data which maximize the remaining variance in the data, which helps preserve global structure by minimizing information loss. We use PCA to reduce the SBERT embeddings to 200 dimensions, then use UMAP on the 200-dimensional PCA vectors to reduce the embeddings to 5 dimensions.

A.4 Propensity Score Matching

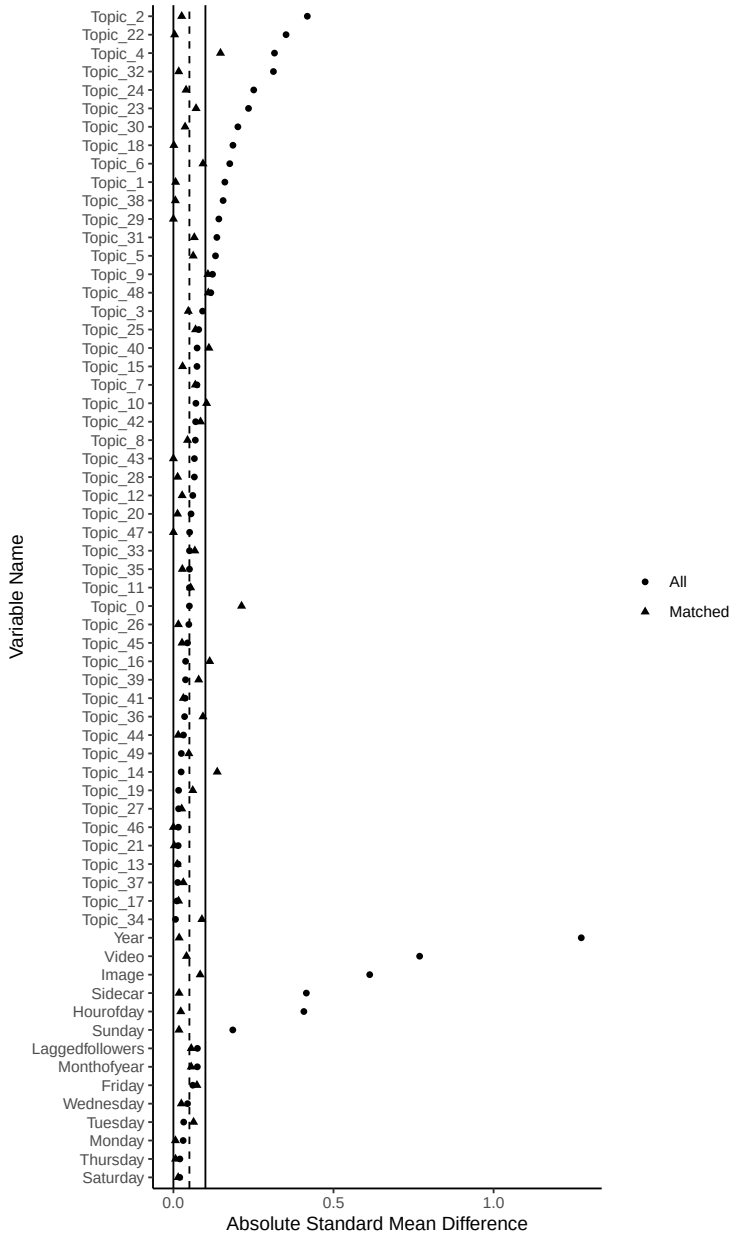


Figure A.3: Love Plot from propensity score matching

Note: There are two sets of matching variables: the 50 topics extracted from the post media described in Section 2.3, and the hand-selected features we describe in the same section. We sort variables within each group in order of decreasing absolute standard mean difference in the unmatched (All) data. The absolute standardized mean difference is computed as $\left|(\bar{X}_{treat} - \bar{X}_{control})/\hat{\sigma}_{treat}\right|$. *In the matched data, there are almost no instances where the matching variable has an absolute standard mean difference larger than 0.1, which is the threshold for successful matching.*

Table A.2: Balance table from propensity score matching

	Unmatched controls			Matched controls		
	Means Treated	Means Control	Std. Mean Diff.	Means Treated	Means Control	Std. Mean Diff.
Distance	0.005	0.0011	0.7594	0.005	0.005	0
HourOfDay	15.4256	13.8221	0.4074	15.4256	15.3349	0.023
Friday	0.1744	0.1514	0.0606	0.1744	0.2023	-0.0735
Monday	0.1628	0.1516	0.0304	0.1628	0.1605	0.0063
Saturday	0.1116	0.1179	-0.0198	0.1116	0.107	0.0148
Sunday	0.0767	0.1261	-0.1855	0.0767	0.0721	0.0175
Thursday	0.1419	0.1489	-0.0202	0.1419	0.1395	0.0067
Tuesday	0.1628	0.1509	0.0321	0.1628	0.1395	0.063
Wednesday	0.1698	0.1532	0.0441	0.1698	0.1791	-0.0248
MonthOfYear	6.4163	6.1661	0.0747	6.4163	6.6023	-0.0556
Year	2023.2	2022.2	1.2738	2023.2	2023.2	0.018
Lagged Followers	59968	56393	0.075	59968	62620	-0.0558
Image	0.1047	0.2923	-0.6131	0.1047	0.0791	0.0836
Sidecar	0.1884	0.3507	-0.4152	0.1884	0.1953	-0.0178
Video	0.707	0.357	0.769	0.707	0.7256	-0.0409
Topic 0	0.0758	0.0662	0.0499	0.0758	0.1171	-0.2129
Topic 1	0.0813	0.051	0.1606	0.0813	0.0826	-0.0068
Topic 2	0.0148	0.0476	-0.4182	0.0148	0.0128	0.0259
Topic 3	0.0687	0.0525	0.0911	0.0687	0.0604	0.0465
Topic 4	0.0258	0.0591	-0.3157	0.0258	0.0413	-0.1471
Topic 5	0.0683	0.045	0.1315	0.0683	0.0573	0.0618
Topic 6	0.0143	0.0313	-0.176	0.0143	0.0233	-0.0924
Topic 7	0.0476	0.0352	0.0735	0.0476	0.0361	0.0684
Topic 8	0.0321	0.0227	0.0684	0.0321	0.0382	-0.0441
Topic 9	0.0212	0.0321	-0.1222	0.0212	0.0117	0.1079
Topic 10	0.0467	0.0352	0.0703	0.0467	0.0297	0.1036
Topic 11	0.0226	0.028	-0.05	0.0226	0.0169	0.0538
Topic 12	0.0205	0.0261	-0.0605	0.0205	0.0179	0.0273
Topic 13	0.0349	0.0325	0.0146	0.0349	0.0368	-0.012
Topic 14	0.0341	0.031	0.0242	0.0341	0.0514	-0.1368
Topic 15	0.0224	0.0296	-0.0736	0.0224	0.0252	-0.0286
Topic 16	0.0307	0.0265	0.0381	0.0307	0.0184	0.1133
Topic 17	0.026	0.0245	0.0112	0.026	0.0239	0.0162
Topic 18	0.0469	0.0159	0.1859	0.0469	0.0467	0.0014
Topic 19	0.0205	0.0223	-0.0162	0.0205	0.0142	0.0605
Topic 20	0.0254	0.0189	0.0551	0.0254	0.0239	0.0129
Topic 21	0.0205	0.0222	-0.015	0.0205	0.0208	-0.0028
Topic 22	0.0058	0.0204	-0.3518	0.0058	0.0057	0.0037
Topic 23	0.0069	0.0172	-0.2346	0.0069	0.0037	0.0708
Topic 24	0.0047	0.0183	-0.2509	0.0047	0.0025	0.0395
Topic 25	0.0092	0.0151	-0.079	0.0092	0.0144	-0.0691
Topic 26	0.0072	0.0103	-0.0482	0.0072	0.0062	0.0155
Topic 27	0.0153	0.0135	0.0162	0.0153	0.0124	0.0259
Topic 28	0.0197	0.0122	0.0655	0.0197	0.0182	0.0132
Topic 29	0	0.0142	-0.1419	0	0	0
Topic 30	0.0058	0.0182	-0.2014	0.0058	0.0081	-0.0365
Topic 31	0.0263	0.0111	0.1357	0.0263	0.019	0.0653
Topic 32	0.0026	0.0098	-0.3124	0.0026	0.003	-0.0167
Topic 33	0.0045	0.0071	-0.0502	0.0045	0.001	0.0669
Topic 34	0.0074	0.0069	0.0066	0.0074	0.0138	-0.089
Topic 35	0.0047	0.0074	-0.0502	0.0047	0.0031	0.0279
Topic 36	0.0091	0.0067	0.0349	0.0091	0.0027	0.0923
Topic 37	0.0099	0.0088	0.0133	0.0099	0.0073	0.0309
Topic 38	0.0031	0.0077	-0.1552	0.0031	0.0033	-0.006
Topic 39	0.0093	0.006	0.0378	0.0093	0.0023	0.0789
Topic 40	0.0131	0.0061	0.0741	0.0131	0.0236	-0.111
Topic 41	0.0085	0.0058	0.0364	0.0085	0.0062	0.031
Topic 42	0.0085	0.0033	0.0694	0.0085	0.0149	-0.0848
Topic 43	0	0.0039	-0.0655	0	0	0
Topic 44	0.0035	0.0051	-0.0315	0.0035	0.0043	-0.0151
Topic 45	0.0058	0.003	0.0442	0.0058	0.0076	-0.027
Topic 46	0.0023	0.0016	0.0156	0.0023	0.0023	0
Topic 47	0	0.0024	-0.0507	0	0	0
Topic 48	0.0047	0.0012	0.1168	0.0047	0.008	-0.1091
Topic 49	0.0008	0.0012	-0.0247	0.0008	0	0.0482

Note: Std. Mean Diff. is the standardized mean difference, $(\bar{X}_{treat} - \bar{X}_{control})/\hat{\sigma}_{treat}$

B Identification and Estimation

B.1 Event study analysis

To formally explore whether there are pre-trends in the data, we specify the following event study model:

$$\text{FollowerGrowth}_{jt} = \sum_{s=-21}^{42} \beta_s \cdot \text{Treat}_{j,t-s} + \alpha_j + \alpha_t + \alpha_{t-t^*(j)} + u_{jt} \tag{B.1}$$

Instead of estimating a single cumulative effect β with decay given by $f(t-t^*(j), \theta)$, we estimate a β_s coefficient for the relative effect of being in the treatment group up to 21 days before to treatment, and 42 days after treatment. If there is no pre-trend, then β_s should not be estimated statistically significantly different from zero for $s < 0$. As in Equation (1), α_j is the creator fixed effect, α_t is the calendar date fixed effect, and $\alpha_{t-t^*(j)}$ is a time-to-post fixed effect that captures the baseline effect of posting.

Coefficients are plotted in Figure B.1. The 95% confidence intervals for each β_s with $s < 0$ clearly cover 0, and thus we fail to reject the hypothesis of no pre-trends using this framework.

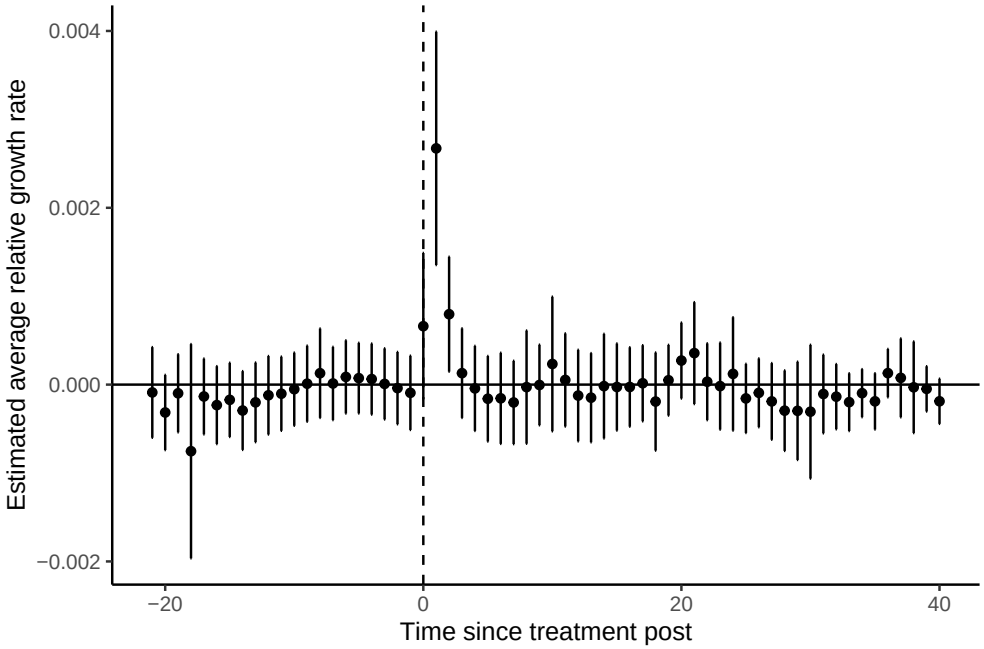


Figure B.1: Event study specification, follower growth rates

Note: This graph plots the estimated coefficients $\hat{\beta}_s$ from Equation (B.1) using follower growth as the dependent variable. All estimated coefficients for periods prior to treatment are near zero and insignificant, which provides evidence against pre-trends for treated creators. Error bars are 95% confidence intervals, clustered at the creator level.

B.2 Linear pre-trend analysis

Angrist and Krueger (1999) propose a test for pre-trends, which involves estimating a differential linear trend on the pre-treatment data for treated and control units. In the marketing literature, Bronnenberg et al. (2020) and Ching and Goetz (2024) implement this test to help validate parallel trends assumptions. We estimate:

$$Y_{jt} = \delta_1 \cdot t + \delta_2 \cdot \text{Treat}_j \cdot t + \alpha_j + \alpha_t + u_{jt} \qquad \text{for } t \in [-30, -1] \tag{B.2}$$

for each of the outcome variables, t is the time-to-post, and Treat_j is an indicator variable that equals one if creator j is a treated unit. δ_1 will capture the underlying joint linear pre-trend for both treatment and control creators in the 30 days prior to posting, while δ_2 estimated significantly different from 0 would capture a differential pre-trend for treated creators.

Estimates from Equation (B.2) are presented in Table B.1. We find no differential linear trend for treated creators prior to treatment. In additional specifications we also try looking at shorter time windows before posting, of 10 days and 5 days, and continue to find no evidence of a differential pre-trend for treated creators pre-treatment.

Table B.1: Testing for linear pre-trends

	Dependent variable:		
	Daily Follower Growth		
	(1)	(2)	(3)
t	-0.00002	0.00001	0.0001
	(0.00001)	(0.00001)	(0.00002)
	[p = 0.0044]	[p = 0.3648]	[p = 0.0215]
$t \times$ Treat	0.00001	-0.000005	-0.00005
	(0.00001)	(0.00001)	(0.00003)
	[p = 0.2225]	[p = 0.7434]	[p = 0.1293]
No. Days Prior to Treatment.	30	10	5
Creator FE	✓	✓	✓
Observations	24,688	8,393	4,232
R^2	0.3884	0.5838	0.7280

Note: This table reports estimates from Equation (B.2), using daily follower growth as the dependent variable. Column 1 includes 30 days for each treated and control creator prior to treatment, Column 2 includes 10 days, and Column 3 includes 5 days. Standard errors are clustered at the creator level and reported in (.) below the estimated coefficients. Exact p values are reported in [.] below the standard errors.

B.3 Timing test - post coordination with external shocks

Figure B.2 provides examples of local maxima in search volumes near collaborative post timings. We aggregate the panel of collaborative posting timing to the creator×week level, and run the following regression:

$$\text{CollabPost}_{jt} = \beta_0 + \sum_{s=-6}^{6} \beta_s \text{LocalMax}_{t-s} + \alpha_j + \epsilon_{jt}$$

where j indexes a creator, t indexes a week, CollabPost_{jt} is an indicator for whether j has their collaborative post published in week t, and LocalMax_{t-s} is an indicator for whether the local maximum in search volume occurred at time $t-s$, for $s = -6, \ldots, 6$. To understand the regression, consider the coefficient β_{t-3}. If this coefficient were to be positive and significant, it would indicate that there is a greater propensity for collaborative posts 3 weeks after the local maximum in search volumes for the creator.

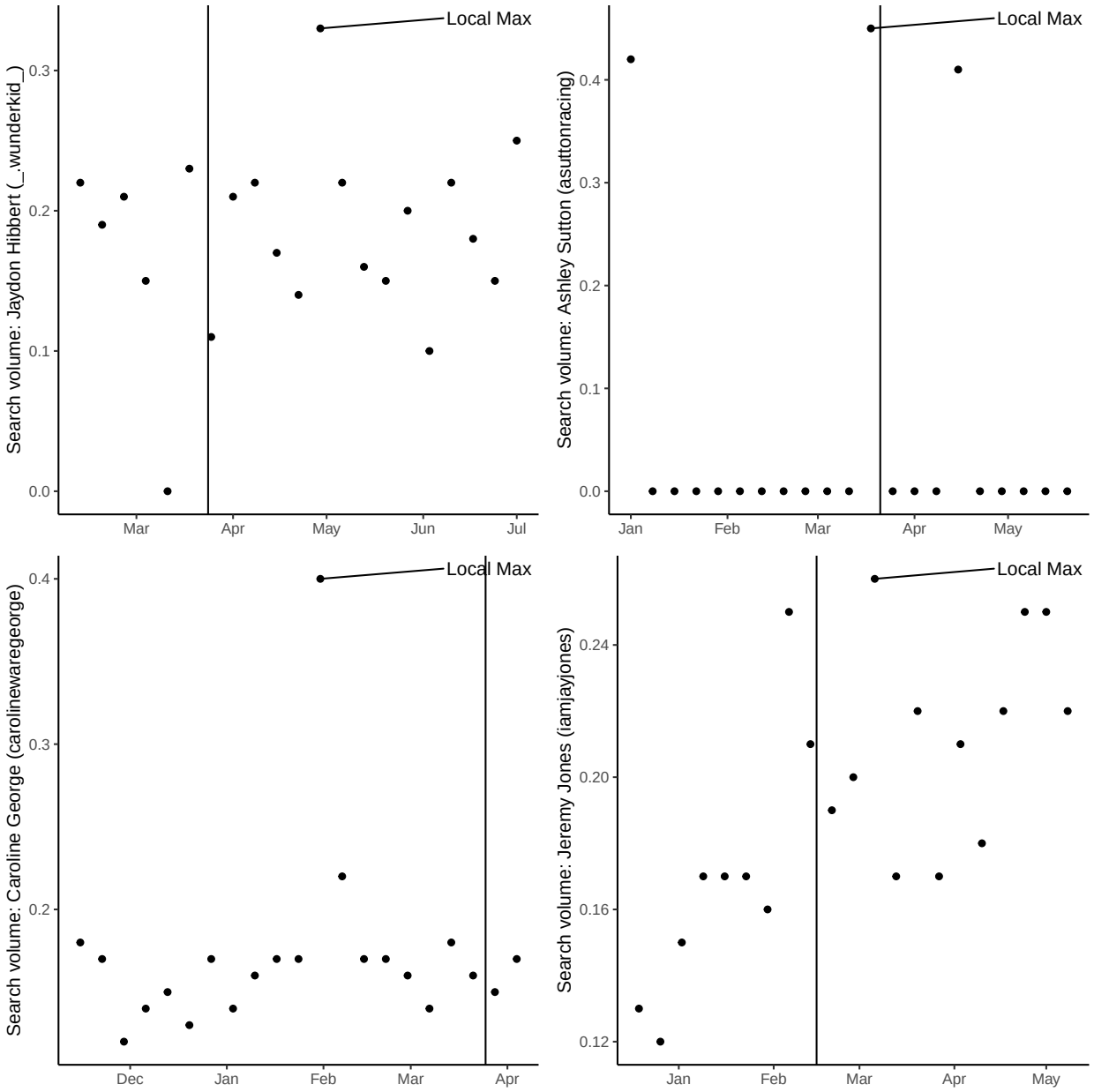


Figure B.2: Examples of local maxima in search volumes near collaborative posts

Note: Each vertical line is the date on which the collaborative post is published on Instagram for that creator. Search volumes are scraped at the weekly level from Google trends, and normalized by dividing by 100. Years are 2024, 2022, 2024, and 2023 moving from left to right starting at the upper left.

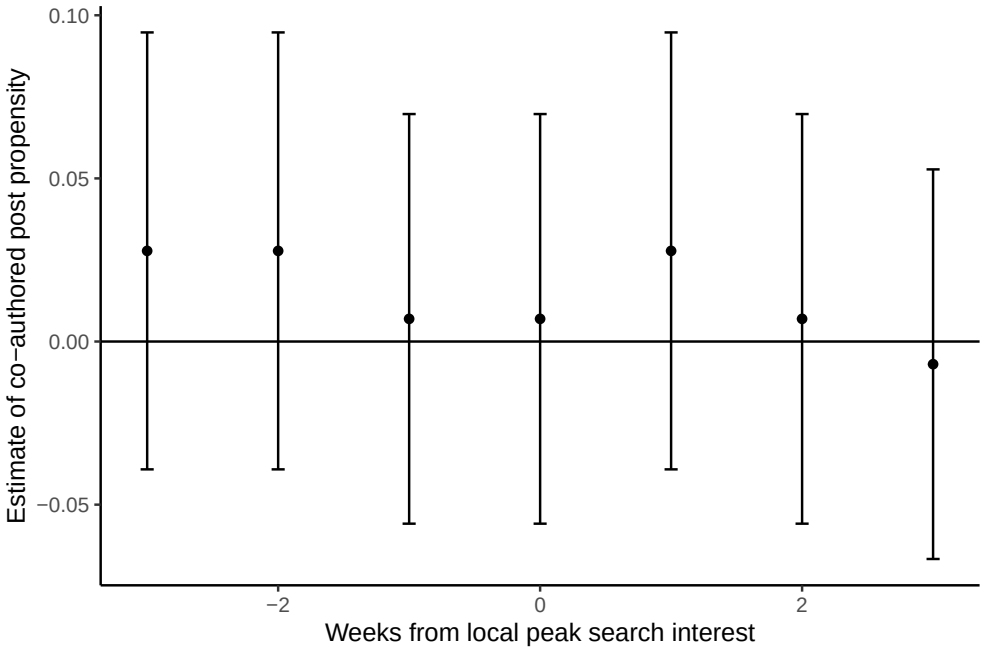


Figure B.3: Collaborative posting around times of peak search interest in creator, within a 3 week window

Note: This figure replicates the analysis in Figure 5, except defining the local search maximum within 3 weeks of the collaborative post. This figure is not equivalent to simply truncating Figure 5, since the timing of the local search maximum within 3 weeks of the collaborative post may be a different week compared to the timing of the local search maximum within 6 weeks of the collaborative post.

B.4 Estimation

Our assumption from Section 3.2 that treatment timing is exogenous conditional on our set of fixed effects and controls implies the following moment condition:

$$\mathbb{E}[u_{jt}|\text{Treat}_{jt}, \mathbf{X}_{jt}, \alpha_j, \alpha_t, \alpha_{t-t^*(j)}, \theta, \beta, \boldsymbol{\gamma}] = 0$$

We estimate Equation (1) using the generalized method of moments (GMM). The parameters to estimate are the non-linear Poisson parameter θ, the cumulative treatment effect β, the control parameters γ, and the fixed effects $\boldsymbol{\alpha}$. Since all parameters other than θ are linear, our estimation procedure first picks a value for θ, then estimates $(\beta, \boldsymbol{\gamma}, \boldsymbol{\alpha})$ using a fixed-effects regression. We recover $\hat{u}_{jt}(\theta, \beta, \boldsymbol{\gamma}, \alpha)$ and interact it with $\tilde{\text{Treat}}_{jt}$, which yields the empirical moment condition:

$$\hat{g}(\theta) = \frac{1}{N} \sum_j \frac{1}{T_j} \sum_t \hat{u}_{jt}(\theta, \beta(\theta), \boldsymbol{\gamma}(\theta), \boldsymbol{\alpha}(\theta)) \cdot \text{Treat}_{j,t} \tag{B.3}$$

We form the empirical GMM objective $\hat{g}(\theta)' \cdot \hat{g}(\theta)$. Because θ is a scalar, we search for the optimal θ using golden section search with parabolic interpolation.

Our estimation approach extends prior work in the influencer marketing literature on estimating non-linear parameters in a panel data setting with fixed effects, including Li et al. (2025) and Huang and Morozov (2025) who estimate a geometric decay of influencer content on purchases. Consistency, unbiasedness, and asymptotic normality of the GMM estimator are guaranteed as long as Equation (B.3) holds, and the standard technical assumptions are satisfied (Hayashi, 2011).

Since our setting features panel data with staggered treatment, GMM estimation may be biased if there are dynamic treatment effects and treated creators post-treatment are recruited as control units (Goodman-Bacon, 2021; Callaway and Sant'Anna, 2021; Roth et al., 2023). As we explained in Section 2.2, we only used treated creators' pre-treatment posts as controls which should help minimize this problem. Moreover, in Figure 4, the effect of treatment appears to decay rapidly to zero, so that it is very unlikely that treated creators' post-treatment periods featuring non-zero dynamic treatment effects will be used as controls.

B.5 Event study with alternate dependent variable: follower growth

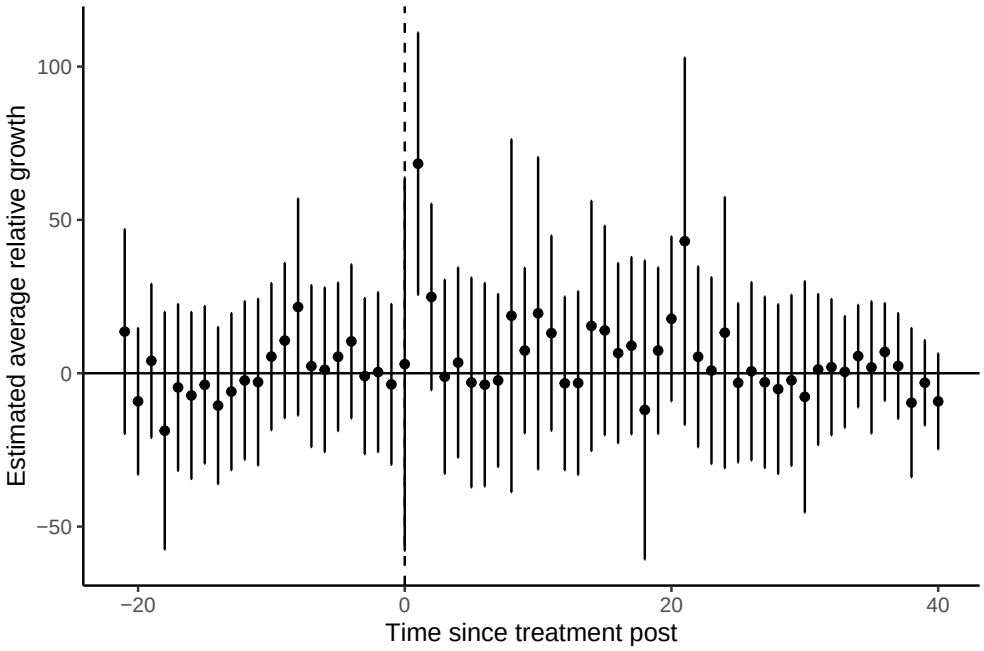


Figure B.4: Event study specification, follower growth

Note: This graph plots the estimated coefficients $\hat{\beta}_s$ from Equation (B.1), estimated using follower growth (instead of growth rate) as the dependent variable. All estimated coefficients for periods prior to treatment are near zero and insignificant, which provides evidence against pre-trends for treated creators. Error bars are 95% confidence intervals, clustered at the level of the creator.

C Robustness Checks

C.1 Robustness: subsample analysis

Table C.1: Robustness checks - subsample analysis

	Dependent variable:	
	Follower Growth Rate	
	(1)	(2)
β	0.0043	0.0055
	(0.0013)	(0.0012)
	[p = 0.0009]	[p = 0.0000]
θ	1.2878	1.1724
	(0.19)	(0.1408)
	[p = 0.0000]	[p = 0.0000]
Creator FE	✓	✓
Date FE	✓	✓
Time-To-Post FE	✓	✓
Cross-Posting Controls		✓
Avg. Value of Dep. Var.	0.0008	0.0008
Observations	41,176	38,638
R^2	0.1935	0.2811

Note: This table presents estimates of Equation (1) using different cuts of the data, and using daily follower growth rate as the dependent variable. Column 1 drops any creator-brand pairing where cross-posting by the brand to other social media platforms is detected, to verify that cross-posted collaborations are not driving our treatment effects. In Column 2, we only use posts by not-yet-treated creators at least 90 days prior to treatment as controls, to verify that spillovers from treated to control creators are not biasing our results. The estimated β coefficient, when multiplied by 100, is the cumulative percent increase in followers (relative to just before treatment) due to the collaborative post. θ is the Poisson parameter and governs the shape of the treatment effect decay over time. Cross-posting controls are a vector of dummy variables capturing if a brand posted about the creator to TikTok, Facebook, or Twitter/X, with 7 days of lagged dummies included for each platform. Standard errors are clustered at the creator level and reported in (.) below the estimated coefficients. Exact p values are reported in [.] below the standard errors.

C.2 Robustness: omitting creators with large followings on other social media platforms

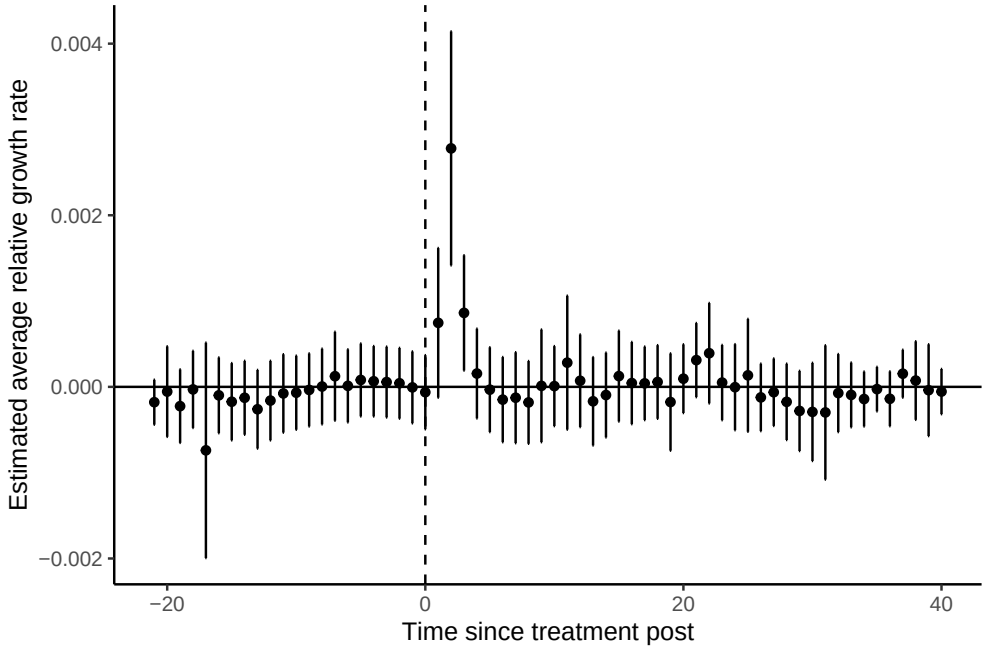


Figure C.1: Event study specification, follower growth rates, dropping creators active on other social media platforms

Note: This graph plots the estimated coefficients $\hat{\beta}_s$ from Equation (B.1), dropping any creators who mention another social media profile in their Instagram biography to screen out creators who are large on other platforms. All estimated coefficients for periods prior to treatment are near zero and insignificant. Error bars are 95% confidence intervals, clustered at the level of the creator.

C.3 Robustness: dropping creators with fast follow-on posts

Table C.2: Robustness checks - dropping creators who quickly follow up

	Dependent variable:	
	Follower Growth Rate	
	(3)	(4)
β	0.0039	0.0037
	(0.0015)	(0.0016)
	[p = 0.0093]	[p = 0.0208]
θ	1.5973	1.8030
	(0.2676)	(0.7725)
	[p = 0.0000]	[p = 0.0196]
Creator FE	✓	✓
Date FE	✓	✓
Time-To-Post FE	✓	✓
Cross-Posting Controls	✓	✓
Avg. Value of Dep. Var.	0.0008	0.0008
Observations	29,336	20,724
R^2	0.1496	0.1423

Note: This table presents estimates of Equation (1) using different cuts of the data, and using daily follower growth rates as the dependent variable. In Column 1 (Column 2) we drop any treated creator who creates a post within one day (two days) after the collaborative post, to verify that fast follow-on posts are not driving treatment effects. The estimated β coefficient, when multiplied by 100, is the cumulative percent increase in followers (relative to just before treatment) due to the collaborative post. θ is the Poisson parameter and governs the shape of the treatment effect decay over time. Cross-posting controls are a vector of dummy variables capturing if a brand posted about the creator to TikTok, Facebook, or Twitter/X, with 7 days of lagged dummies included for each platform. Standard errors are clustered at the creator level and reported in (.) below the estimated coefficients. Exact p values are reported in [.] below the standard errors.