Influencer Dynamics

Kazimier Smith

November 12, 2024

Before reading, please click here for the most recent version

Abstract

Consumers use social media for entertainment and to discover new products. To reach potential customers, brands pay influencers to feature products in their content. Payment depends on the size of the influencer's audience, and the effectiveness of the endorsement relies on trust. Excessive product recommendations may erode the relationship between an influencer and their followers. I develop a dynamic model in which an influencer produces sponsored posts (which include product recommendations) and organic posts (without product recommendations). Both affect the influencer's growth and require effort to produce. A sponsored post incurs additional costs: it requires searching for and negotiating with brands, and followers are less likely to engage with it. Using 2,780,011 Instagram posts and 136,453 TikTok posts from 1,369 influencers, I quantify the engagement penalty by comparing identical posts across platforms. Analyzing the influencers' career histories, I show that organic and sponsored posts have similar effects on follower growth. I leverage variation in the number and types of posts to estimate the unknown cost parameters in the model. An influencer with 100,000 followers optimally produces about 0.25 sponsored posts and two organic posts per week, and influencers with more followers produce more content of both types. Viral posts boost growth. Regulating sponsored content can theoretically backfire by decreasing incentives to produce organic posts, but counterfactual simulations assuage this concern. A cost increase that reduces sponsored content by 25% only causes a 2.7% drop in organic content.

1 Introduction

Social media influencers drive much of the modern digital economy. They produce the content to which the average American devotes two hours of their day (Statista (2024a)). They interact personally with fans and build loyal followings. They recommend products to these audiences, fueling the $220 billion social media marketing industry (Statista (2024b)). Large brands capitalize on the trust followers have in influencers to run more impactful campaigns. Niche brands can use influencers to target a specific audience most likely to enjoy their products. For any firm, a product recommendation that goes viral can generate overwhelming sales. This ecosystem attracts broad interest: one in four members of Gen Z wants to become a social media influencer (Novacic (2019)).

Influencer careers are highly variable. While viral posts generate a large upside for influencers and brands, most posts do not see stellar performance. Online influencer communities believe that posting regularly, rather than relying on a single hit, is the key to success, but even influencers who follow this advice often grow slowly. Both luck and effort seem to play a role, but the effect of content production on career outcomes is not fully understood.

In this paper, I combine theory with empirical analysis to better understand the economics behind an influencer's choices and career progression. I develop a dynamic model in which an influencer produces two types of content to grow their following. The first is *organic posts*, which are standard social media content that does not include a product recommendation. For example, a cooking-focused influencer films themselves making pizza dough and shares the recipe with their audience. The second type is *sponsored posts*. A sponsored post is like an organic post, but it includes a specific product recommendation. The cooking influencer films the pizza recipe and mentions that they prefer to use King Arthur Flour. King Arthur Flour pays the influencer for the post, and the payment depends on the size of the influencer's audience (their follower count).

The influencer solves a dynamic optimization problem. The single state variable is the influencer's follower count each week. Given the state, the influencer chooses the number of organic and sponsored posts to produce this week. The influencer is paid for each sponsored post based on their follower count, and influencers with more followers receive higher payments. Both types of content affect the growth of the influencer's audience. The change in follower count from the current period to the next period depends directly on the number of organic and sponsored posts the influencer makes. It also depends on the average performance of the influencer's content (how many likes it receives). A higher fraction of sponsored posts reduces average performance. I estimate these relationships from data. Follower growth includes a normally distributed random shock to account for factors the influencer cannot observe (like the platform's content distribution algorithm).

Making content is costly. Both organic and sponsored posts require time and energy (an "effort cost"). Sponsored content incurs an additional "match cost" because the influencer spends time searching for a sponsoring brand, negotiating with them, and conforming their content to the brand's requirements. Each type of post exhibits increasing marginal costs.

The performance penalty sponsored posts incur significantly affects the influencer's decision. Estimating it is challenging because I do not observe the amount of effort the influencer puts into each post. If influencers exert less effort on sponsored content, then sponsored posts perform worse than organic posts not because they are sponsored but because they are lower quality. Because I do not observe effort, I implicitly assume that the influencer's second sponsored post in a week always takes the same amount of effort. Since effort is not in the model, I need to estimate the effect of sponsorship on post performance while holding effort fixed.

I achieve this using a novel empirical strategy based on *cross-posts*, which are identical posts that an influencer uploads to multiple platforms. Effort is the same for both posts because the photo or video content is identical. When the post is sponsored on one platform and not on the other, I can isolate the impact of sponsorship on performance. To estimate this, I compile a list of 1,369 Instagram content creators in eight categories (beauty, cycling, fitness, food, lifestyle, mom, tech, travel), and I manually match them to their TikTok accounts when they exist. I collect each influencer's entire post history on both platforms, and I classify posts as sponsored by searching their text for keywords like #ad. I identify 9,697 cross-posted pairs by looking for matching text. Without controlling for effort, a sponsored Instagram post receives about 30% fewer likes than its organic counterpart. After adjusting for effort, the effect decreases to about a 19% penalty for sponsorship. Effort plays a major role in sponsored posts' reduced performance, and failing to control for it would affect my quantitative results.

With this key piece addressed, I estimate the remaining parts of my model. To determine the relationship between follower count and pay, I collect information on about 21,854 payments from brands to anonymous influencers. The data reveal a positive, linear relationship between follower count and pay for sponsored content. In line with anecdotal industry evidence, an influencer with 10,000 followers earns about \$145 per sponsored post. I cannot structurally estimate payment parameters because the payments data are anonymous, so I estimate this relationship offline and input it into my model.

Finally, I supplement the post data with a daily time series of each influencer's follower count. I focus my analysis on Instagram, where I am able to collect 4.9 years of follower count data for the median influencer. The post and follower count data reveal that the growth penalty for sponsored content is small. For an influencer with 100,000 followers who otherwise would have gained 1,000 followers, one extra organic post yields 39 additional followers while one extra sponsored post yields 18. Much of the growth penalty therefore comes from the fact that sponsored posts receive fewer likes, but even this effect is quite small. While a sponsored post gets 19% fewer likes, that translates into only about six fewer followers.

I input the pay-follower count relationship and the growth-content relationship into the model, and then I apply the method of simulated moments to obtain structural estimates of the parameters of the effort and match cost functions. The resulting model fits the data well. In both the data and the estimated optimal policy, influencers with more followers produce more organic and sponsored content. An influencer with 10,000 followers behaving optimally makes about 2.5 organic and 0.25 sponsored posts per week, while an influencer with 1,000,000 followers increases production to about 3.5 and 0.75 per week, respectively. Influencers with more followers make more sponsored content because brands pay them more and because the additional revenue outweighs the increasing marginal cost of another post. They make more organic posts because doing so generates a larger raw increase in follower count than it would for a smaller influencer. That increase translates to larger future payments, so organic content is more attractive for larger influencers. On net, influencers increase and then slightly decrease their share of sponsored content as they grow. Initially, the extra pay from sponsored content makes it more attractive, but at a certain size, the growth benefit from organic content outweighs pay, so the fraction of sponsored posts begins to decline.

To test the role of luck in an influencer's career, I simulate a viral post by introducing a large, positive follower count shock in one period. The influencer gains many followers, which changes the composition of their content according to the optimal policy. Soon after the viral post, though, the influencer's follower count trajectory returns to its previous trend. Growth is approximately percentage-based in my model, so influencers with more followers grow more quickly in terms of raw follower count. This means that the gap between the influencer's actual follower count and their counterfactual follower count without the viral post widens over time. Virality generates a persistent increase in audience growth. My follower count data lend weight to the model's prediction. The data show clear jumps in follower count after viral posts.

Existing theoretical models of influencer behavior (e.g. Nistor et al. (2024)) predict periods of "investing", when the influencer focuses on growth rather than revenue, and of "harvesting", when the influencer makes sponsored content to extract value from their audience. This result relies on a substantial follower growth penalty from sponsored content which does not appear in my data. Instead, my model predicts an increase in the fraction of sponsored content because influencers with more followers are paid more. Beyond about 100,000 followers, the extra growth from and lower cost of organic content become so attractive that influencers stop increasing their fraction of sponsored content. Observed content production choices are largely driven by short-term costs and benefits rather than dynamic incentives.

Counterfactual simulations analyze changes in incentives and in platform policies. The impact of these changes is not obvious because the return to sponsored content affects the return to organic content since the influencer's problem is dynamic. Increasing the follower growth penalty for a sponsored post slightly reduces organic content production and has almost no effect on sponsored content production. My calculated optimal policy is therefore robust to alternative estimates of the penalty.

Increasing the cost of a sponsored post by about 20% (e.g. through disclosure rules that make sponsored content more time-consuming) slightly reduces both organic and sponsored content production. When sponsored posts become more costly, both costs and benefits of organic posts decrease, and the net effect is small. Regulations that affect the return to sponsored content will likely have little impact on organic content, alleviating concerns that they might cause influencers to substantially reduce their content production.

2 Literature

I contribute to several parts of the literature on influencer marketing. The payment data I collect allow me to test the predicted relationships between pay and follower count and between pay *per follower* and follower count. Wies et al. (2023) hypothesize and empirically confirm an inverted U relationship between follower count and engagement with sponsored content. This confirms industry wisdom that the most valuable influencers on a per-follower basis are those in the middle of the follower count distribution. Brands running influencer marketing campaigns care about reaching many people but also about converting them to purchases, so they should pay more per follower to mid-sized influencers than to very large influencers. Instead, I find a consistent negative relationship between pay per follower and follower count (Figure 27). Precisely explaining this pattern requires further study, but it could reflect small influencers' unwillingness to work for low pay. If producing a sponsored post has some fixed cost independent of follower count, then brands must pay more per follower to small influencers to make their offers worthwhile.

Tian et al. (2024) predict an S-shaped relationship between impressions on a sponsored post and follower count, although the shape varies depending on the specific marketing campaign. If firms set pay by equating it to the marginal benefit (in terms of impressions) of another follower, then the relationship between pay and follower count should be similar to the relationship between impressions and follower count. I find a positive and approximately linear relationship (Figure 26), so brands either are not aware of or do not internalize the declining marginal benefit of an additional follower. While smaller influencers may see higher engagement on their content, brands seem to care mostly about reach (i.e. follower count). If they took engagement into account, the pay vs. follower count curve would likely exhibit nonlinearities.

A key assumption in existing theoretical models of influencers is the dynamic cost of sponsored content, which generates revenue but either reduces follower count or slows its growth. Nistor et al. (2024) assume that when an influencer starts endorsing products that are a poor fit for their audience, a certain fraction unfollows them. Mitchell (2021) models the length of the relationship between an influencer and a follower, but the same tradeoff arises: sponsored content reduces the length of the relationship. I use my post and follower count data to test the empirical validity of this assumption. Contrary to expectations, both organic and sponsored content have a positive effect on follower growth, and the magnitudes of the effects are almost identical. This result could be unique to Instagram, since Cheng and Zhang (2022) find a negative effect of sponsored Youtube videos on subscriber count. Youtube is a very different platform from Instagram, and Hughes et al. (2019) find that consumer reactions to sponsored content differ across platforms. If consumers view Instagram as a source of product recommendations and Youtube as an entertainment platform, they may be more amenable to sponsored content on Instagram. Advertising can also be useful for consumers, for example, through the signaling effect for experience goods theorized in Nelson (1974). Sahni and Nair (2020) find this effect to be large and positive for restaurants. Influencers in my sample often advertise experience goods: beauty influencers use specific brands of makeup, tech influencers recommend particular

PC components, and food influencers favor one olive oil over another. If brands signal their quality by paying influencers to advertise, sponsored content could have a positive effect on consumers.

Behavioral literature empirically demonstrates several variables that mediate negative effects of sponsored content, such as explicit disclosure (Giuffredi-Kähr et al. (2022)), the number of other people the influencer follows (Valsesia et al. (2020)), and use of high-arousal language (Cascio Rizzo et al. (2024)). Many factors simultaneously influence post performance, so pinning down the true effect of sponsorship is empirically challenging. Bairathi and Lambrecht (2023) use the Federal Trade Commission's warning to influencers about disclosure as an instrument to identify the effect of sponsorship and find a negative effect. I offer another empirical strategy leveraging *cross-posts* uploaded to both Instagram and TikTok. I show that sponsorship has a negative impact on follower engagement, but controlling for post quality using a counterfactual non-sponsored post reduces the magnitude of the effect. The studies above show varied impacts of sponsorship depending on the nature of the content, and my result indicates that influencers can mitigate sellout effects by making higher quality sponsored posts.

I use the pay and follower growth facts above to inform a dynamic model of influencer content production similar to the model of authenticity choice in Nistor et al. (2024). The influencer's key choice in their model is whether to accept sponsorship offers that are a poor fit for their audience. Accepting poor-fit offers generates more revenue but causes some followers to abandon the influencer. I focus on the distinction between sponsored and organic posts rather than the fit of sponsored content. This allows me to test whether the follower growth penalty or other costs primarily explain observed content production. Modeling organic and sponsored posts also adds a degree of flexibility since influencers can substitute between the two. Authenticity arises as either the total number or fraction of sponsored posts, and I can describe the choices from which a given level of authenticity derives.

I implicitly assume, as do existing models, that making more sponsored posts requires accepting poor-fit offers that alienate followers. Since making more sponsored posts decreases their average fit in my model, the influencer covers a broader range of topics as they grow. Gong (2021) establishes the same fact as a solution to the cold-start problem: influencers initially make niche content to attract users with specific tastes. As the influencer develops their reputation, they broaden their content to attract more users. They also produce more sponsored content because, as in my paper and other models, their larger follower base makes it more lucrative. I do not explicitly model followers or the fit between content and audience. Instead, they appear in a reduced form way via the cost function for sponsored content. The cost of a sponsored post increases as the influencer makes more of them because the influencer starts with the highest fit sponsorship offers and then accepts worse ones. Notably, Leung et al. (2022) find an inverted U relationship between follower-brand fit and the effectiveness of influencer marketing campaigns. Future work could develop a dynamic model that captures this fact.

In existing literature, the dynamic cost of sponsored content generates two distinct behaviors. First, an influencer spends time growing their audience by prioritizing organic content ("investing"). Once their audience is sufficiently valuable, they "harvest" by making sponsored content for brands in exchange for money. Nistor et al. (2024) and Mitchell (2021) both show this pattern in different ways. The influencer is willing to spend time investing because it generates higher future advertising revenues. Forgoing payment today can maximize total lifetime utility. My data show that this tradeoff is empirically small, so while it appears in my model, it is not the main force disincentivizing sponsored posts. Instead, the marginal benefit of another sponsored post increases with follower count more quickly than the marginal cost, so influencers make more sponsored content as their audiences grow.

Viral posts appear in my model because follower growth is subject to a random shock each period, and these shocks are occasionally large and positive. While literature beginning with Berger and Milkman (2012) identifies factors that influence virality, influencers' post histories indicate that "going viral" is rare and unpredictable. Influencers instead try to post high-quality content regularly and hope that it will occasionally perform extremely well. When influencers do go viral in my model, they grow quickly for a short period of time before returning to their previous trend.

Only a few existing papers structurally estimate the parameters of an influencer's problem. Tang et al. (2012) model and estimate a Youtuber's content production decision. They include revenue sharing but do not explicitly model sponsored vs. organic content. Li (2023) studies the disclosure decisions of Twitch streamers and, in counterfactual analyses, examines the impacts of regulating disclosure. Instagram is a very different platform from YouTube and Twitch, so studying Instagram influencers is valuable in its own right. I also collect novel data on payments from brands to influencers so I can more accurately estimate how pay impacts influencer utility. Adding this relationship to the model means I can identify unknown cost function parameters. Modeling explicit costs of both types of content allows me to analyze different counterfactual scenarios, such as the introduction of a platform tool matching influencers to potential sponsors.

3 Data and reduced form evidence

There are three key pieces to understanding influencers' content choices:

1. How pay for sponsored content depends on follower count
2. How follower count evolves over time given content production
3. How costly it is to produce organic and sponsored content

I address the first two points with reduced form analysis, while I leave the third for structural estimation.

3.1 Pay for sponsored content

To estimate the dependence of pay for sponsored posts on follower count, I collect a novel dataset from FYPM.vip. On the site, influencers submit reviews of their collaborations with brands. Site staff verify both the influencers and the collaborations. The site's primary goal is to make it easier for influencers "to figure out what price to charge for their services" (FYPM (2024)); they can use the site to see how much similar influencers are paid for brand collaborations. Figure 1 shows a sample review. The influencer who submitted the review collaborated with a brand called "Cat Person" in late 2023. The influencer has 100,000 TikTok followers and 20,000 Instagram followers, and the brand asked them to post a short-form video on both platforms. The brand paid the influencer \$500 cash and gave them cat chews worth \$60. The "review" section at the bottom is the influencer's free-form description of their experience working with the brand.

From these reviews, I estimate the relationship between follower count and pay. I average all the listed follower counts in the review, so in the example above the influencer's aggregated follower count is $(100,000+20,000)/2 = 60,000$. I calculate total pay by summing cash pay and the value of free products, so total compensation is \$560. Finally, I adjust for the fact that collaborations requiring multiple pieces of content tend to pay more (Figure 28). The platform reports an aggregate number of "deliverables" for each review; it is three in the example above. My outcome variable is pay per deliverable, which is $\$560/3 = \$186.66$. I collect 21,854 reviews and exclude those for which follower count, number of deliverables, or pay is zero.

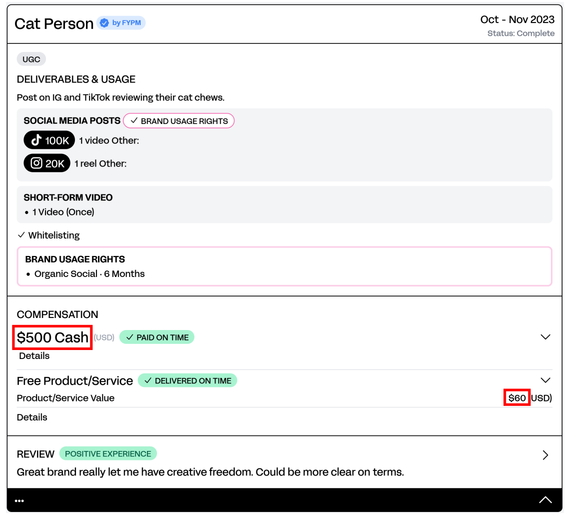


Figure 1: Review of accepted collaboration

The site also includes reviews of *declined* collaborations; influencers typically describe them as paying too little to be worthwhile. I exclude these since they do not describe the true relationship between follower count and pay for a sponsored post. The resulting sample has 15,047 reviews.

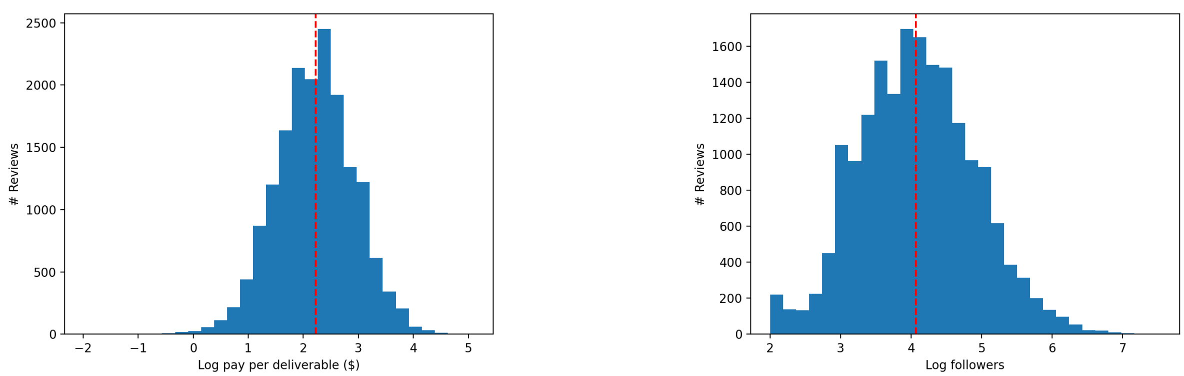


Figure 2: Main review data variables

Figure 2 shows the distributions of the two main variables from the review data. The median accepted review is from an influencer with about 10,000 followers who is paid about $125 per deliverable.

Table 1 estimates the relationship between pay and follower count. Since date fixed effects do not substantially increase explanatory power and since it is not clear how to include them in a dynamic model of a single influencer, I input the left column into my model. An influencer with 10,000 followers receives $145 per deliverable, while pay increases to $448 per deliverable for an influencer with 100,000 followers. Table 22 estimates the same regression using only Instagram followers; the coefficient on log followers increases slightly to 0.509, which is unlikely to make a large difference in subsequent estimation.

	Log pay per deliverable	Log pay per deliverable
Log followers	0.489***	0.482***
	(0.006)	(0.006)
Intercept	0.206***	1.878**
	(0.024)	(0.582)
Date FE	No	Yes
N	15,047	15,047
R2	0.327	0.337

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 1: Dependence of pay on follower count

3.1.1 Sample selection

Influencers voluntarily write reviews on FYPM.vip, so my payment data could suffer from selection bias if influencers only review particulary positive experiences or if only larger influencers write reviews. The presence of 809 declined reviews suggests that at least some influencers view the site as a place to record all offers. If influencers are willing to post unacceptable offers, then they are likely also willing to post low-paying offers or low-quality offers. My pay calculations are also broadly in line with industry estimates of about $100 per 10,000 followers (Geyser (2022)). If anything, the relationship I estimate from the data shows pay increasing more slowly with follower count than in the industry numbers, so upward bias in the pay data is unlikely. Downward bias is also unlikely since influencers rate the majority of accepted reviews positively. Figure 3 shows details of the rating section of the review. The influencer rates the collaboration

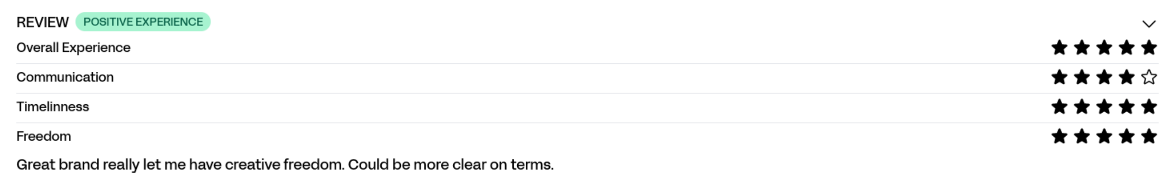


Figure 3: Review of accepted collaboration

in each of four categories, and I assign the review an "average score" by taking the mean of the star ratings in each category. Figure 4 shows the distribution of these average scores separately for accepted and declined reviews. The majority of accepted reviews receive close to five stars in all categories, so it seems unlikely that influencers only review collaborations with which they are unsatisified.

3.1.2 Declined reviews

In my sample, 809 reviews are marked "Declined Offer", which means a brand proposed a sponsorship to an influencer and the influencer said no. Figure 5 shows an example in which the influencer declined because the compensation (free water) was insufficient. Reading through the reviews, I also found several in which the influencer declined because the brand was a poor fit for their audience. Table 2 predicts the probability of declining given several review characteristics. Influencers with more followers are more likely to decline offers, probably because they receive more offers and can be pickier. Pay and the influencer's "overall experience" with the brand have a significant negative impact on the probability of accepting the offer, as expected. Freedom, the level of creative control the influencer has over the sponsored post, increases the probability

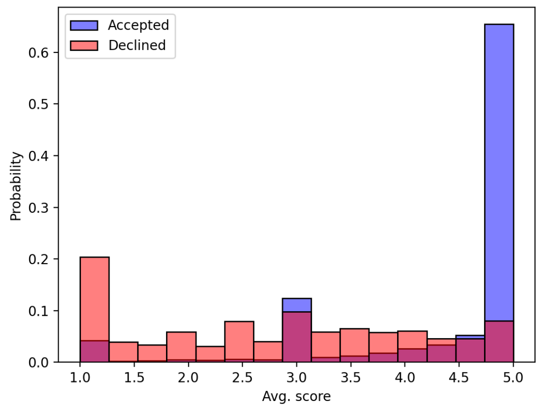


Figure 4: Review scores

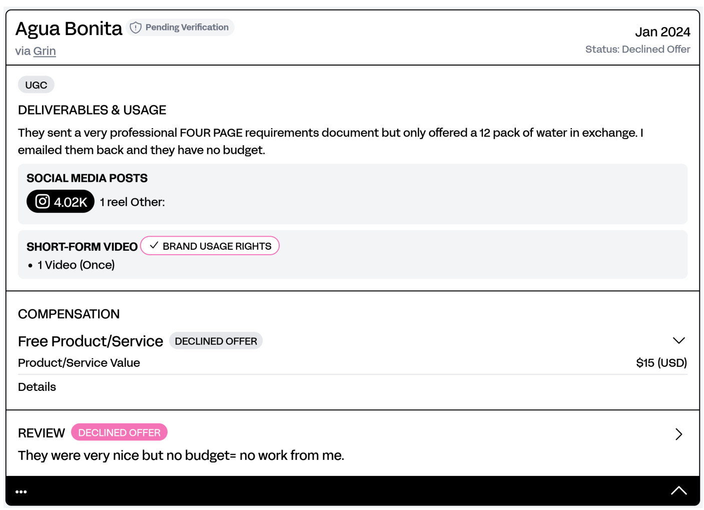


Figure 5: Review scores

of accepting an offer. In Pei and Mayzlin (2022), a firm that affiliates too closely with an influencer can reduce the persuasivness of that influencer's product review. The result in Table 2 suggests that influencers understand this cost.

3.2 Follower count transition

To determine how follower count evolves over time, I collect post and follower count data for 1,369 influencers. I compile the list from feedspot.com, an influencer search website that maintains pages such as "Top 100 Lifestyle Influencers". I collect these pages for the categories beauty, cycling, fitness, food, lifestyle, mom, tech, and travel. Each page contains influencers of varying sizes; the smallest influencer in my sample had 883 Instagram followers on January 1st, 2023, while the largest had nearly 52 million.

The lists focus on Instagram, but many of the influencers also use TikTok. I manually match each Instagram account to the corresponding TikTok account by searching Google, checking the influencer's Instagram profile, and checking their website. I find a TikTok account for about 80% of the influencers. The lists from feedspot.com focus on Instagram influencers, so TikTok is often not their primary platform. Many

	Declined
Log followers	0.533***
Log pay	-1.040***
Free product	0.218
Log deliverables	0.105
Usage rights	0.204+
Verified	-0.141
Overall experience (1-5)	-1.637***
Communication (1-5)	0.385***
Timeliness (1-5)	0.322***
Freedom (1-5)	-0.135**
Via agency	-0.135
# Words	-0.004***
Intercept	0.299
N	15,559
N (declined)	512

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Table 2: Predicting declined offers

of the TikTok accounts have just one or two videos from a few years ago.

Next, I use a data collection platform [1] to collect all Instagram and TikTok posts for the influencers in my sample. After dropping duplicate posts (which likely arise from scraping issues), posts missing a username, and posts missing a date, I end up with 2,780,011 Instagram posts and 136,453 TikTok posts. For each of these I collect all text-based information about the post, including the post caption, hashtags, number of likes, post date, comments, any paid partnership labels, and other features. Figure 6 shows a sample post from the recipe creator @dadaeats, and Table 3 compares the means of several post characteristics on the two platforms. Average follower count and average likes are higher on TikTok because of selection: the smaller influencers in my sample tend to post less on TikTok or do not use it at all. Instagram posts typically have more text. For example, food influencers in my data typically post full recipes in the post caption on Instagram but omit them on TikTok.

	Instagram	TikTok	p-value
Followers	849,979.68	1,080,461.57	0.00
Sponsored	0.09	0.09	0.01
Likes	10,962.13	28,477.63	0.00
# Comments	160.45	133.02	0.00
# Words in post	55.11	25.39	0.00
Viral	0.07	0.44	0.00
N	2,780,011.00	136,453.00	

Table 3: Comparing Instagram and TikTok posts

[1] Brightdata throught the Bright Initiative

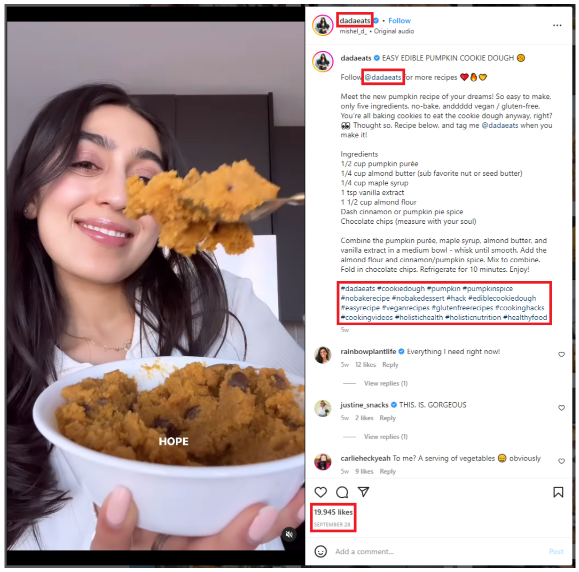


Figure 6: Post information

3.2.1 Classifying sponsored posts

The influencer's ability to produce different numbers of organic and sponsored posts is a key element of my model. I analyze post text to classify sponsored posts in my data. I first create a list of 8,094 brands sponsoring collaborations that appear in the review data from FYPM.vip. I classify a post as sponsored if it meets at least one of the following conditions:

1. Mentions one of the 8,094 brands (a mention begins with the @ symbol)

2. Uses the official platform disclosure tool (Paid partnership with...)

3. Contains explicit disclosure like #ad, #sponsored, or #lululemon_partner

244,202 Instagram posts (8.78%) and 12,005 TikTok posts (8.80%) are sponsored. Figure 7 shows a post sponsored by Pura, a home fragrance producer.

Table 4 compares organic and sponsored Instagram posts. Notably, the difference in average likes for both post types is statistically insignificant. I will return to this fact when I estimate the influencer's follower count transition function. The average follower count for organic posts is probably higher because larger influencers make more organic posts (Figure 13).

	Organic	Sponsored	p-value
Followers	853,998.02	822,795.17	0.00
Likes	10,949.56	11,092.87	0.13
# Comments	155.49	211.29	0.00
# Words in post	52.13	85.66	0.00
Viral	0.06	0.10	0.00
N	2,532,763.00	247,248.00	

Table 4: Organic vs sponsored Instagram posts

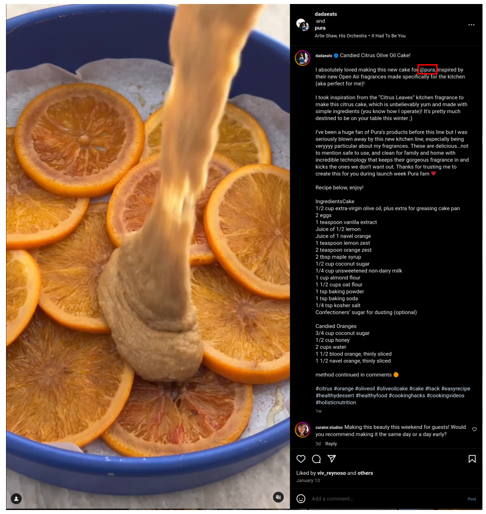


Figure 7: Brand mention

3.2.2 Alternative classifications of sponsored posts

There are many ways to separate sponsored from organic posts. Rather than claim a perfect classification system, I implement some other methods to provide upper and lower bounds on the number of sponsored posts in my data. I start with the least restrictive classification: assuming any post that mentions another Instagram account is sponsored. This creates false positives because influencers sometimes mention their friends or other influencers, and these mentions do not indicate sponsorship. On the other hand, Figure 7 demonstrates that a mention is sometimes the only evidence of sponsorship in the post text. My primary classification method only looks for mentions of brands in the FYPM.vip data; treating all mentions as evidence of sponsorship ensures no brand is excluded. It therefore provides an upper bound on the number of sponsored posts in the data. Table 5 compares organic and sponsored Instagram posts, where any post

	Organic	Sponsored	p-value
Followers	802,182.95	904,296.23	0.00
Likes	10,911.56	11,034.12	0.05
# Comments	142.79	185.48	0.00
# Words in post	46.20	67.74	0.00
Viral	0.06	0.07	0.00
N	1,629,882.00	1,150,129.00	

Table 5: Organic vs sponsored (any mention) Instagram posts

with a mention other than the influencer's own username is classified as sponsored. Unsurprisingly, the number of sponsored posts increases significantly.

Some, but not all, posts mention brands from the FYPM.vip data. To gauge how well FYPM covers the universe of possible sponsors, I try classifying any post mentioning a brand from FYPM as sponsored. Table 6 describes sponsored posts classified this way. Many of the mentions in Table 5 are either false positive or mention brands not in the FYPM data.

Checking for explicit sponsorship disclosure should almost completely avoid false positives. It seems unlikely that "#sponsored" would appear in an organic post, so 141,747 is probably a lower bound on the number of sponsored posts in my data. Table 7 classifies any post using the paid partnership label or hashtags

	Organic	Sponsored	p-value
Followers	838,348.30	1,018,910.05	0.00
Likes	10,818.77	13,842.30	0.00
# Comments	155.41	260.71	0.00
# Words in post	53.88	79.65	0.00
Viral	0.06	0.10	0.00
N	2,646,889.00	133,122.00	

Table 6: Organic vs sponsored (mentions FYPM brand) Instagram posts

like "#ad" as sponsored. Table 8 only looks for the paid partnership label. There is a remote possibility that some influencers use disclosure like "#ad" on posts that are not sponsored so they appear to receive more sponsorship offers than they actually do. The paid partnership label cannot be abused this way because the partnering brand also participates in the disclosure. Note that my primary classification method in Table 4 is the union of those in Tables 6 and 7.

	Organic	Sponsored	p-value
Followers	870,741.05	618,184.47	0.00
Likes	11,086.14	8,607.34	0.00
# Comments	160.11	166.85	0.15
# Words in post	53.08	92.99	0.00
Viral	0.06	0.11	0.00
N	2,638,264.00	141,747.00	

Table 7: Organic vs sponsored (disclosed) Instagram posts

	Organic	Sponsored	p-value
Followers	859,806.66	532,542.37	0.00
Likes	10,983.67	9,321.75	0.00
# Comments	159.92	199.00	0.01
# Words in post	54.35	109.81	0.00
Viral	0.07	0.15	0.00
N	2,741,827.00	38,184.00	

Table 8: Organic vs sponsored (disclosed with platform tool) Instagram posts

Finally, I apply sponsorship classification methods from Ershov and Mitchell (2020). Ideally I would use their machine learning algorithm to detect sponsored content, but it relies on a reliable ground truth. The authors use Instagram posts after a regulation change in Germany which required disclosure on sponsored posts. Fines and enforcement followed the change, so Ershov and Mitchell (2020) assume all sponsored posts are disclosed after the new regulations. Their German Instagram data therefore serves as training data for machine learning models that differentiate organic and sponsored posts. Since I used United States data, I have no equivalent training dataset. In fact, Ershov et al. (2023) suggests that most sponsored content is undisclosed.

Instead, I use the manual classification from Ershov and Mitchell (2020). It consists of two lists of words. The first (Table 23) contains explicit disclosure indicators and the second (Table 24) consists of words suggesting sponsorship. The second list in particular is broad; it contains words like "until" that could be part of a limited time offer ("coupon valid until October 1st") but could also be present in organic posts. Table 9 shows that searching text for these words classifies about two-thirds of the Instagram posts in my data as sponsored. Ershov and Mitchell (2020) agree that "our manual definition likely overstates the amount of sponsored content". Searching post text for Ershov and Mitchell (2020)'s explicit sponsorship disclosures

	Organic	Sponsored	p-value
Followers	1,088,051.74	781,707.98	0.00
Likes	12,606.84	10,147.79	0.00
# Comments	137.01	171.99	0.00
# Words in post	19.10	72.83	0.00
Viral	0.04	0.08	0.00
N	916,810.00	1,863,201.00	

Table 9: Organic vs undisclosed sponsored (Ershov and Mitchell (2020)) Instagram posts

yields about 200,000 sponsored posts (Table 10). This is the closest number to my original classification, which had about 244,000 sponsored posts. Finally, Table 11 classifies posts as sponsored if they contain a

	Organic	Sponsored	p-value
Followers	886,098.76	555,518.97	0.00
Likes	11,102.01	9,138.96	0.00
# Comments	152.19	266.66	0.00
# Words in post	51.50	101.50	0.00
Viral	0.07	0.08	0.00
N	2,579,343.00	200,668.00	

Table 10: Organic vs disclosed sponsored (Ershov and Mitchell (2020)) Instagram posts

word either from Table 23 or from Table 24. Combining the two shows that most posts classified as sponsored because they contain explicit disclosure words also contain other indicators of sponsorship.

	Organic	Sponsored	p-value
Followers	1,099,971.93	780,148.33	0.00
Likes	12,683.96	10,129.19	0.00
# Comments	136.59	171.93	0.00
# Words in post	18.76	72.59	0.00
Viral	0.04	0.08	0.00
N	902,620.00	1,877,391.00	

Table 11: Organic vs sponsored (Ershov and Mitchell (2020)) Instagram posts

3.2.3 Match value

The alignment between a sponsored post and the author's typical content affects the audience's reaction to it (Leung et al. (2022)). I calculate a measure of follower-brand fit based on text similarity as follows:

1. Generate vector representations of each post's text using Doc2Vec with 100 features (Řehůřek and Sojka (2010)).
2. Calculate each influencer's "average" organic post on each platform by taking the componentwise mean of all their organic posts.
3. Calculate the match value of a sponsored post as the cosine similarity between the post's vector representation and the influencer's average organic post.

Figure 8 shows the distribution of match values for sponsored Instagram posts. The median Instagram post has a similarity score of 0.56. Leung et al. (2022) empirically show an inverted U-shaped relationship between

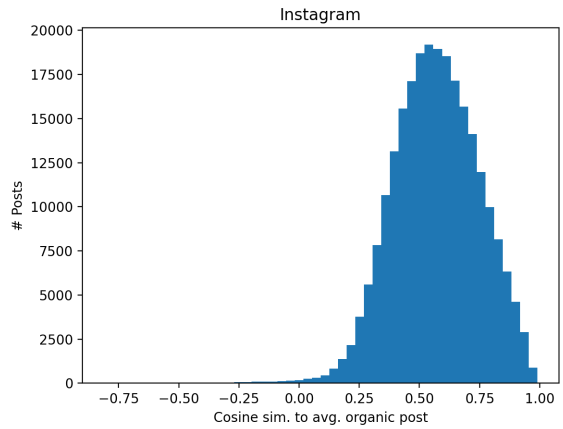


Figure 8: Similarity to average organic post

follower-brand fit and the performance of an influencer marketing campaign, although they calculate fit in a different way. I test this in my data by plotting engagement (likes plus comments divided by followers) and match value for sponsored Instagram posts (Figure 9). The relationship does seem to have an inverted U

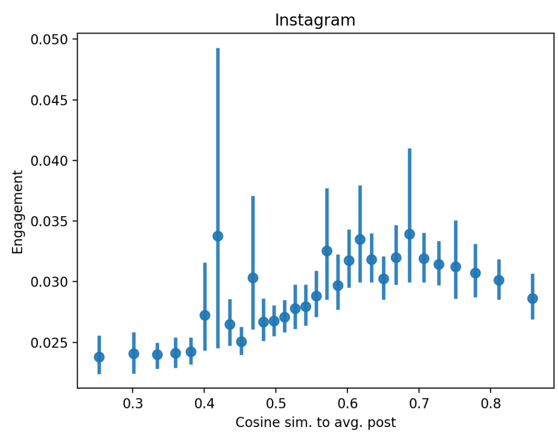


Figure 9: Engagement vs similarity to average organic post

shape. The sponsored posts that see the highest engagement are not the ones most similar to the influencer's typical content. My measure of fit is different from Leung et al. (2022). They measure the overlap between the brand's category (e.g. beauty) and the categories in which the audience is interested.

With this measure, crowding out is one reason for an inverted U relationship. If the audience is interested in beauty and travel, beauty/travel brands will face fierce competition since many similar brands will offer to sponsor the influencer. A beauty/tech brand can still appeal to the audience while avoiding competition.

My text similarity-based measure exhibits a similar pattern for different reasons. Sponsored posts very similar to the influencer's average organic post might seem to followers like an attempt to conceal the sponsorship. Alternatively, they might not stand out from the influencer's typical content, while a sponsored post that is very different attracts more attention from followers. The pattern could also reflect different behavior by follower count. Some of the largest influencers have a set "formula" for their content, so their sponsored posts will tend to have very high similarity scores. Larger influencers also tend to see lower engagement than mid-sized influencers.

3.2.4 Base specification for follower count transition

To assess how follower count changes over time based on content production decisions, I collect daily follower count histories on both platforms for the creators in my sample from a social media analysis platform [2]. While the platform has limited data on TikTok follower counts, the Instagram data is extensive. Figure 10

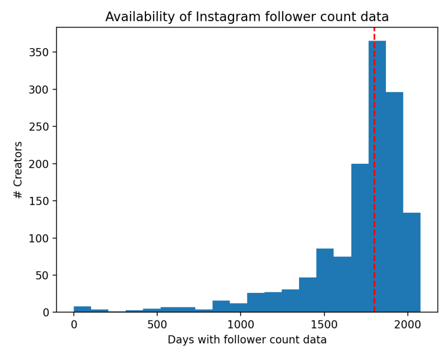


Figure 10: Instagram follower count data availability

shows the distribution of the number of days of follower count data I have; the median creator has about 4.9 years of data.

I assume that an influencer's change in followers from one week to the next depends on the amount of organic and sponsored content they make and the performance (measured by likes) of that content. I aggregate the data to form a weekly panel of influencers. My main specification is the following:

$$\log f_{it+1} - \log f_{it} = \tau_o o_{it} + \tau_s s_{it} + \tau_\ell \ell_{it} + \epsilon_{it} \tag{1}$$

The variable definitions are:

- f_{it+1} Influencer i's follower count in week $t+1$
- f_{it} Influencer i's follower count in week t
- o_{it} Total organic posts by influencer i in week t
- s_{it} Total sponsored posts by influencer i in week t
- ℓ_{it} Average likes across all posts by influencer i in week t, or zero if the influencer did not post
- ϵ_{it} Normally distributed follower count "shock" to capture randomness in growth (e.g. viral posts). The mean is zero and the standard deviation is σ_f, which I estimate from data.

[2]Social Blade

I focus on the change in follower count because of the unit root problem: follower count in week $t+1$ is typically very similar to follower count in week t, so regressing $\log f_{it+1}$ on $\log f_{it}$ would yield a coefficient close to one. Such a regression does not produce reliable estimates of the other coefficients. Most formal unit root tests require a balanced panel, so I create a balanced version of my panel running from June 30th, 2020 to June 30th, 2023. I exclude influencers who have any weeks with missing follower count data in this period; the resulting panel has 1,130 influencers. I apply a test that combines hypothesis tests for individual influencers. I can reject the null hypothesis of having a unit root for only 111 of the influencers, so it is possible that many of the follower count time series do have unit roots. This justifies my choice to use the *change* in follower count as the dependent variable.

Follower count typically changes only a small amount from one week to the next, so log differences are almost identical to percent change. Of the influencer-week observations for which I can calculate change in followers, 98.6% have less than a 5% change. Figure 11 shows the distribution of the observations with less than a 5% change. Table 12 estimates equation 1. Organic posts, sponsored posts, and likes all have small

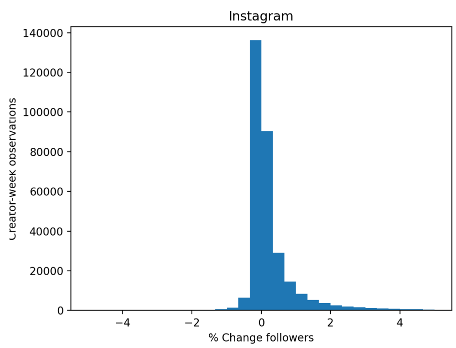


Figure 11: Distribution of percent change in followers

	Change log followers	Change log followers	Change log followers
Posted	0.00077***		
	(4.664e-05)		
# Organic posts		0.00025***	0.00019***
		(1e-05)	(1e-05)
# Sponsored posts		0.00016***	0.00010***
		(2e-05)	(2e-05)
Posted * Log likes			0.00028***
			(2e-05)
N	301,392	301,392	301,392
Creator FE	Yes	Yes	Yes
Date FE	Yes	Yes	Yes
R2	0.000917	0.0036	0.00451

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 12: Empirical transition function

but positive and statistically significant effects on follower growth. The mean of the regression residuals is $1.06 \cdot 10^{-18}$, validating my assumption about the error term. The standard deviation is about 0.018; I input this value into my model. The most surprising result is the similarity between the estimates of τ_o

and τ_s. Organic and sponsored content appear to have similar effects on future follower count. Although the coefficient on sponsored posts is smaller, the difference between the two coefficients implies very little difference in follower count growth. Taken at face value, this result challenges the common perception that sponsored content leads to slower or even negative follower count growth. To investigate this further, I include other variables in the transition regression.

3.2.5 Fraction of sponsored posts

Followers might care not about the total quantity of advertising, but about its pervasiveness throughout the influencer's content. Influencers who are primarily focused on commercializing their content might see slower growth. Anecdotally, an influencer marketing firm told me they recommend influencers keep their fraction of sponsored content below a threshold. In this case the influencer's follower growth depends on the total number of posts they produce and on the fraction of those that are sponsored. To determine this effect I estimate the following equation:

$$\log f_{it+1} - \log f_{it} = \tau_p(o_{it} + s_{it}) + \tau_r \frac{s_{it}}{o_{it} + s_{it}} + \tau_\ell \ell_{it} + \epsilon_{it} \tag{2}$$

Table 13 shows the result. Increasing the fraction of sponsored posts reduces follower growth, but the effect is small. An influencer who grows from 100,000 followers to 101,000 followers with zero sponsored content would grow to about 100,894 followers if 100% of their content were sponsored that period, a loss of 106 followers.

	Change log followers	Change log followers	Change log followers
# Posts	0.00024***	0.00024***	0.00018***
	(7.3251e-06)	(1e-05)	(1e-05)
Posted * Frac. sponsored		-0.00030***	-0.00046***
		(8e-05)	(8e-05)
Posted * Log likes			0.00029***
			(2e-05)
N	301,392	301,392	301,392
Creator FE	Yes	Yes	Yes
Date FE	Yes	Yes	Yes
R2	0.00354	0.00359	0.00457

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Table 13: Empirical transition function including authenticity

3.2.6 Cumulative effects in the transition function

The negative effects of sponsored content might build up over time. Followers might be willing to tolerate the occasional week of mostly sponsored posts, but if the influencer produces heavily commercialized content week after week, their growth could slow. Other cumulative effects might impact the probability that a user discovers the influencer. Influencers with more total posts or higher quality past posts could appear more often in searches and on algorithmically curated discovery pages, in which case they would see faster follower growth. I calculate an influencer's total "stock" of posts P_T at time T as

$$P_T = \sum_{t=0}^{T-1} \beta^{T-t}(o_t + s_t) \tag{3}$$

where $\beta = 0.99$ is the discount factor. I calculate a time-discounted stock because Instagram's discovery algorithm seems to put some weight on recency. The details of the algorithm are not public, but my own Explore page shows exclusively posts made this year. Viral trends also arise and die out quickly on social media, so older content quickly loses relevancy. Discounting old posts when calculating an influencer's post stock accounts for these effects. To measure the past performance of an influencer's content, I calculate the average number of likes on all posts prior to the current period. Table 14 adds these cumulative effects to

	Change log followers	Change log followers	Change log followers	Change log followers
# Organic posts	0.00019***	0.00029***	0.00015***	0.00023***
	(1e-05)	(1e-05)	(0.00001)	(0.00001)
# Sponsored posts	0.00010***	0.00020***	0.00011***	0.00017***
	(2e-05)	(2e-05)	(0.00002)	(0.00002)
Posted * Log likes	0.00028***	0.00033***	0.00063***	0.00063***
	(2e-05)	(2e-05)	(0.00002)	(0.00002)
Post stock		-0.00001***		-0.00001***
		(0e+00)		(0.00000)
Log avg. likes			-0.01003***	-0.00905***
			(0.00012)	(0.00013)
N	301,392	301,392	295,174	295,174
Creator FE	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes
R2	0.00451	0.0146	0.0268	0.0319

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 14: Empirical transition function with cumulative effects

the transition function regression. The cumulative variables both negatively affect follower growth, probably because they proxy for influencer size (follower count) and time on the platform. Influencers who started making content earlier will likely have a larger stock of posts, and influencers with more followers tend to get more likes. If a particular content category has a "carrying capacity", or a maximum number of users interested in the subject, then influencers in that category will see slower growth as they approach the upper bound. This could explain the negative effects of post stock and cumulative average likes.

3.2.7 Persistence

Reverse causality might cause problems in the transition function regression. If influencers who are growing more quickly receive more offers to make sponsored content, then growth causes an increase in the number of sponsored posts rather than vice versa. For a brand, fast-growing influencers are particularly valuable. The price of a sponsored post in the current period is relatively small, but the influencer will command a much larger following in the future, and the brand's advertisement will be distributed to that larger audience. The dependent variable in my transition function regressions is the change in log followers from week t to week $t + 1$, but a brand does not know the influencer's future follower count. Instead, it might forecast based on the influencer's growth in the previous period. To account for this possibility, Table 15 introduces lagged change in follower count to explain follower growth. The first column is my baseline specification. The second column includes the change in log followers from week $t-1$ to week t, and the third column adds the change in followers from week $t-2$ to week $t-1$. Including past follower growth increases the R^2 substantially, and including an additional lag yields a slightly larger increase. Given the changes in the R^2, a single lag captures most of the effect of past follower growth on current follower growth.

I think of past follower growth as a control for the frequency with which the influencer receives offers to make sponsored content. Influencers who grow more quickly receive more offers. When I include this control, the coefficients on organic and sponsored posts shrink but remain positive and are of the same order of magnitude as in my baseline specification. Given the already-small coefficients, these changes will not result in a significant change in follower growth and should not substantially affect my results.

	Change log followers	Change log followers	Change log followers
# Organic posts	0.00019***	0.00012***	0.00010***
	(1e-05)	(0.00001)	(0.00001)
# Sponsored posts	0.00010***	0.00005**	0.00004*
	(2e-05)	(0.00002)	(0.00002)
Posted * Log likes	0.00028***	0.00020***	0.00019***
	(2e-05)	(0.00002)	(0.00002)
Change log followers (1 lag)		0.33673***	0.30688***
		(0.00166)	(0.00180)
Change log followers (2 lags)			0.11501***
			(0.00175)
N	301,392	299,221	297,080
Creator FE	Yes	Yes	Yes
Date FE	Yes	Yes	Yes
R2	0.00451	0.125	0.141

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 15: Empirical transition function with lagged change in followers

3.2.8 Match value in the transition function

The alignment between the brands that sponsor an influencer and the influencer's audience affects engagement (Figure 9), so it might also affect follower growth. When sponsored posts are not too far from the influencer's typical content, they might not have a large impact on growth. Table 16 adds my measure of follower-brand fit to the transition function regression (I describe the measure in Section 3.2.3). The negative coefficient on match value might reflect the inverted U relationship between engagement and follower-brand fit: sponsored posts that are too similar to the influencer's typical content perform worse and therefore generate less follower growth.

3.2.9 Ershov and Mitchell (2020) classification of sponsored posts

Table 17 summarizes my estimates of the empirical transition function using different methods to classify sponsored posts. The coefficient on the number of organic posts is qualitatively very similar across classifiation methods. It shrinks in the rightmost column most likely becasue Ershov and Mitchell (2020) use a broad definition of sponsorship that probably includes some organic posts (because of classification words like "have"). Classifying sponsored post using only the platform's official "paid partnership" label is the

	Change log followers	Change log followers
# Organic posts	0.00019***	0.00019***
	(1e-05)	(1e-05)
# Sponsored posts	0.00010***	0.00017***
	(2e-05)	(3e-05)
Posted * Log likes	0.00028***	0.00029***
	(2e-05)	(2e-05)
Posted * Match value		-0.00036***
		(1e-04)
N	301,392	301,392
Creator FE	Yes	Yes
Date FE	Yes	Yes
R2	0.00451	0.00456

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Table 16: Empirical transition function with match value

only method that generates a negative coefficient on the number of sponsored posts. These are the most obviously sponsored, so consumers react the most negatively to them. The coefficients on the number of sponsored posts are qualitatively quite similar and typically smaller than the coefficient on the number of organic posts. Although my original classification is neither perfect nor the only option, the magnitudes of the coefficients and the difference between them seem broadly correct, so using a different method would likely not change the results of my structural estimation.

	Change fol.	Change fol.	Change fol.	Change fol.	Change fol.	Change fol.	Change fol.	Change fol.
# Organic posts	0.00019***	0.00018***	0.00019***	0.00015***	0.00019***	0.00018***	0.00018***	0.00009***
	(1e-05)	(1e-05)	(1e-05)	(1e-05)	(1e-05)	(1e-05)	(1e-05)	(1e-05)
# Sponsored posts	0.00010***							
	(2e-05)							
Posted * Log likes	0.00028***	0.00030***	0.00028***	0.00029***	0.00029***	0.00030***	0.00029***	0.00028***
	(2e-05)	(2e-05)	(2e-05)	(2e-05)	(2e-05)	(2e-05)	(2e-05)	(2e-05)
# Spon. discl. posts		0.00002						
		(2e-05)						
# Spon. undiscl. posts			0.00023***					
			(3e-05)					
# Posts w/ mention				0.00007***				
				(1e-05)				
# Posts w/ FYPM mention					0.00015***			
					(3e-05)			
# Paid partnership posts						-0.00013**		
						(4e-05)		
# Spon. discl. posts (EM)							0.00008***	
							(2e-05)	
# Spon. posts (EM)								0.00013***
								(1e-05)
Creator FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Date FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	301,392	301,392	301,392	301,392	301,392	301,392	301,392	301,392
R2	0.00451	0.00442	0.00461	0.00453	0.00451	0.00445	0.00447	0.00478

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Table 17: Empirical transition function with sponsorship classifications from Ershov and Mitchell (2020)

3.3 Effect of sponsorship on content performance

To incorporate the effect of sponsored content on post performance, I assume the number of likes in a given period is the average of the likes on each post in that period. Assuming an influencer makes N posts in a period and suppressing the i, t subscripts, the number of likes ℓ is

$$\ell = \frac{\ell_1 + \cdots + \ell_N}{N}$$

I assume organic and sponsored posts get ℓ_o and ℓ_s likes, respectively and that these values depend on follower count. There is a penalty for being sponsored, so $\ell_s = \ell_o - p$ for some p. If o of the N posts are organic and s are sponsored, then

$$\ell = \frac{o(\ell_o) + s(\ell_o - p)}{o + s} = \ell_o - p\frac{s}{o + s} \tag{4}$$

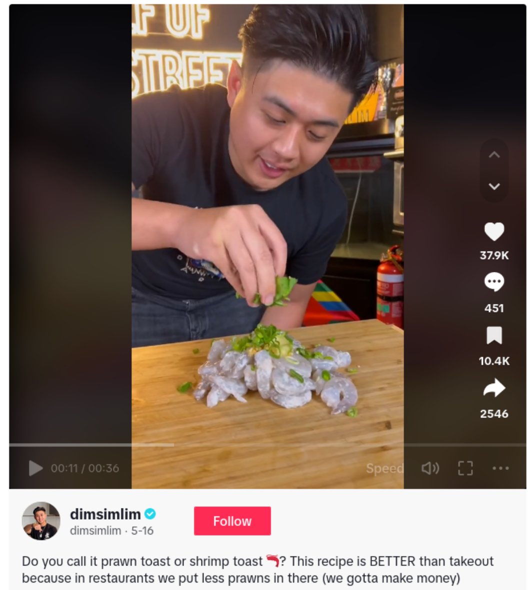
I will use this equation to calculate likes when I simulate my model. To operationalize this model of likes I need estimates of ℓ_o and of p. I calculate the former with a simple prediction of the effect of follower count on likes for organic posts, shown in Table 18.

	Log Instagram likes
Log followers	0.664***
	(0.000)
N	914,809
R2	0.979
+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001	

Table 18: Effect of follower count on likes, organic posts only

Estimating p is trickier because sponsored posts might get fewer likes for two reasons. First, followers dislike advertising. Second, influencers might put less effort into sponsored posts, so the posts perform poorly because they are low quality. I could adjust for quality by allowing organic and sponsored posts to require different levels of effort, but I then would not be able to identify whether effort or other costs like brand negotiations rationalize observed production of sponsored posts. That is, the model would have two effort parameters, θ_e^o and θ_e^s for organic and sponsored posts, respectively. I would have no way to separately identify θ_e^s and θ_m. Instead, I leverage the TikTok data I collected to control for post quality. Influencers sometimes *cross-post* content: they submit an identical video to both platforms. Figure 12 gives an example. The post on the left (Instagram) is clearly sponsored since it includes a “paid partnership” label. The post on the right (TikTok) includes no clear sponsorship disclosure. I assume that followers do not realize the post on the right is sponsored, so they treat it as organic and it receives likes accordingly, while followers treat the left post as sponsored. I use the organic version of the post as a control for the sponsored version, that is, I assume the organic version provides a good estimate of how the sponsored version *would have* performed had it been organic. The organic version essentially measures the post’s quality, and since the videos are identical, I can use the quality measure to isolate the effect of sponsorship. I find all pairs of cross-posts by looking for seven or more consecutive matching words. I estimate the effect of sponsorship on the subset where I have an accurate measure of quality, that is, the cross-posts where the Instagram version is sponsored and the TikTok version is organic. Table 19 shows the results using the TikTok post’s performance to control for the quality of the Instagram post. When I introduce the quality measure, the negative effect of sponsorship shrinks by almost half, indicating that about half of the penalty for sponsored content comes from lower post quality. I use the coefficient on sponsorship in the righthand column (-0.091) as my estimate of p. With these pieces in place I proceed to describe my dynamic model of influencer content production.

Sponsored



"Not sponsored"

Figure 12: Matched posts

4 Model

I model an influencer's content choice as a dynamic optimization problem in discrete time. Each period, the influencer chooses the number of organic and sponsored posts to produce given their follower count. The choices are continuous to simplify estimation, so an influencer can make 3.6 organic posts. My goal is not to match observed influencer behavior exactly but to describe content production patterns over an influencer's career and to examine how these patterns change when model parameters vary. Moreover, in my data I often observe influencers playing a "mixed strategy" like making a sponsored post every other week, which corresponds to making 0.5 sponsored posts per week in my model.

The influencer derives utility from monetary payments for sponsored content, but producing content is costly. If the influencer has f_t followers at the beginning of period t and makes o_t organic posts and s_t sponsored posts, their utility is

$$\alpha(f_t)s_t - c_e(o_t, s_t) - c_m(s_t)$$

The first term, α, is the payment the influencer receives for making one sponsored post. I assume

$$\log_{10} \alpha(f_t) = [\pi_0 + \pi_1 \log_{10}(f_t)]$$

for some coefficients π_0 and π_1. This functional form fits my payments data very well and reflects the standard industry fact that influencers with more followers are paid more.

The second term c_e is the *effort cost* of content; it captures the time and energy the influencer puts into making organic and sponsored posts. I assume that for some parameters θ_e and η,

$$c_e(o_t, s_t) = \theta_e(o_t + s_t)^\eta$$

In terms of pure effort, organic and sponsored content have the same cost: I assume setup, planning, filming,

	Log Instagram likes	Log Instagram likes
Log Instagram followers	0.657***	0.489***
	(0.001)	(0.003)
Sponsored (Instagram)	-0.155***	-0.090***
	(0.027)	(0.022)
Log TikTok likes		0.321***
		(0.005)
Creator FE	Yes	Yes
Date FE	Yes	Yes
N	9,697	9,697
R2	0.970	0.980

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Table 19: Effect of sponsorship on likes, controlling for quality

etc. takes the same amount of time for both types of post. The parameter η makes the cost function convex, so the marginal cost of one or two posts is low, but additional posts quickly become more costly because ideas are scarce. The influencer has a couple of good post ideas on which they can start working immediately. Making more posts requires thinking of new ideas which are increasingly difficult to find.

In my data, the largest influencers make 4-5 posts per week on average, but fewer than one of those is sponsored. If utility depended only on the two components described above, the only way to rationalize observed behavior is through a large negative effect of sponsored content on future follower count. I do not observe such an effect in regression analysis. Sponsored content must therefore have unique costs that do not apply to organic content. To capture these I introduce a *match cost* c_m. The total match cost of s_t sponsored posts follows an exponential function according to some parameter θ_m:

$$c_m(s_t) = e^{\theta_m s_t} - 1$$

c_m could represent many things. For example, to make more sponsored posts, an influencer must accept sponsorship offers from brands that are increasingly far from their typical content. A vegan recipe influencer may receive a few offers from vegan ingredient brands, but producing more sponsored posts might require accepting a sponsorship offer from a meditation app company. The latter sponsored post is more costly for the influencer because (1) they must find a way to incorporate meditation into their usual content (vegan recipes) and (2) the influencer feels bad about "selling out" to a brand in which their audience has little interest. Alternatively, making more sponsored content might require the influencer to search for brands with which to partner, creating a time cost unique to sponsored content. Finally, regulations could impose additional costs on sponsored content. If sponsored posts have to follow a specific format different from the influencer's usual post, then the influencer will have to spend time conforming to the requirements. Ultimately c_m is the reduced form of a more complex model of matching among influencers and brands. My goal is to describe influencer behavior and its changes in response to changes in the costs of making content, so while modeling the matching process explicitly is an interesting avenue for future work, it is not necessary here.

The flow utility specification is novel because it explicitly models production of content rather than an aggregate authenticity measure (e.g. the fraction of sponsored content or the probability of accepting a sponsorship offer). If the influencer can costlessly adjust their authenticity, equilibrium content choices are driven by other forces like follower dynamics. Since I do not observe a dynamic cost to sponsored content, it must have an increasing marginal cost, otherwise influencers would advertise infinitely. It seems natural to think of social media posts as economic objects that require resources to produce; they have increasing marginal costs because time and ideas become scarce. Modeling content production explicitly means I can examine substitution between the two types of content. Authenticity appears in my model as the fraction of the influencer's content that is sponsored, but a given fraction is achievable through multiple policies. If a decrease in the fraction of sponsored content comes with a significant reduction in total content, the apparent increase in authenticity could harm consumers. On the other hand, the ability to expend effort to produce more content alleviates the tradeoff between pay and authenticity: an influencer who increases both sponsored and organic content will see current period revenue go up with no accompanying decrease in authenticity (since the fraction of sponsored content stays the same).

Given the utility specification, the influencer chooses a sequence $\{o_t, s_t\}_{t=0}^{\infty}$ to maximize discounted lifetime utility

$$\sum_{t=0}^{\infty} \beta^t \left[\alpha(f_t)s_t - c_e(o_t, s_t) - c_m(s_t)\right]$$

where β is the discount factor. I can summarize the problem in a Bellman equation in which the state is the influencer's follower count f_t:

$$V(f_t) = \max_{o_t, s_t} \alpha(f_t)s_t - c_e(o_t, s_t) - c_m(s_t) + \beta \mathbb{E} V(f_{t+1})$$

I assume $\beta = 0.99$. The influencer's problem is dynamic because content choice (o_t, s_t) affects the transition from f_t to f_{t+1}. The data show that organic and sponsored posts both generate positive follower growth. To model the transition I assume

$$\log f_{t+1} - \log f_t = \tau_o o_t + \tau_s s_t + \tau_\ell \ell_t + \epsilon_t \tag{5}$$

where $\epsilon_t \sim \mathcal{N}(0, \sigma_f^2)$. I estimate $\tau_o, \tau_s, \tau_\ell, \sigma_f$ from panel data. Define ϕ to be the density of $\log f_{t+1}$ conditional on f_t, o_t, s_t, σ_f, that is, ϕ is the density of a random variable with distribution $\mathcal{N}(\log f_t + \tau_o o_t + \tau_s s_t + \tau_\ell \ell_t, \sigma_f^2)$.

The fact that τ_o and τ_s are positive and almost identical in magnitude is a surprising fact that partly informs the other components of my model. If τ_s were large and negative, influencers would have an incentive to produce organic content to grow their audience. After enough growth, they might begin producing more sponsored content to take advantage of higher payments for their now larger audience. My data do not support this behavior. Instead, influencers produce more organic than sponsored content because an organic post is less costly than a sponsored post and because content composition has little impact on follower growth. The absence of a large dynamic cost to sponsored content could be platform dependent. Consumers use Instagram in part to discover new products, so they may see sponsored content as no less valuable than organic content (which entertains them). In contrast, if consumers see YouTube purely as a source of entertainment, they might punish content creators for advertising too much (as in Cheng and Zhang (2022)). Models in which followers dislike sponsored content are certainly useful to explain some platforms, but their applicability to Instagram seems limited. The lack of dynamic cost also reinforces the need for the additional cost c_m of sponsored content: without it I cannot explain why influencers make so few sponsored posts.

5 Simulation

Before structurally estimating θ_e and θ_m, I assume their values and calculate the resulting value function and optimal policy. I approximate the value function with cubic Hermite splines. I prefer a continuous approximation because follower counts in my data change very little each period. A discretized value function would require a large and computationally burdensome number of grid points to properly approximate the follower count transition process. I choose cubic Hermite splines to maintain monotonicity: the value function should increase with follower count because more followers imply higher pay. I avoid linearly interpolating the value function because I calculate the expected value function with numerical integration. A non-differentiable value function slows down this calculation.

I simulate the model on the state space $(2, 9)$; this interval contains all the observed follower counts (in base 10 logs) I observe in my data and extends beyond the maximum observed follower count. Limiting the grid to span only observed follower counts would force the value function to decline artificially as it approached the right endpoint of the grid since at that point the influencer would have no room to grow. I interpolate the value function with $K = 40$ grid points $g_1, \ldots, g_K$ evenly spaced on the state space.

The algorithm to calculate the value function is a version of modified policy iteration (Judd (1998)). Let C be the polynomial approximation to the value function. I initialize it to be zero everywhere. I initialize the policy function for organic posts as the linear interpolation between one and five. I initialize the policy function for sponsored posts as the linear interpolation between zero and one. Let $o^*(f)$ be the optimal number of organic posts given the current optimal policy and f followers, and let $s^*(f)$ be the optimal number of sponsored posts. The algorithm proceeds as follows:

1. Let $\mathbf{V} = (C(g_1), \ldots, C(g_k))$ (the current value function evaluated at each grid point).

2. For $k = 1, \ldots, K$, calculate W_k as

$$W_k = \alpha(g_k)s^*(g_k) - c_e(o^*(g_k), s^*(g_k)) - c_m(s^*(g_k)) + \beta \int_2^9 C(x)\phi(x; g_k, o^*(g_k), s^*(g_k), \sigma_f)\; dx$$

   I calculate the integral with numerical quadrature.

3. Update C to be the piecewise cubic Hermite interpolating polynomial for $g_1, \ldots, g_K$ and $W_1, \ldots, W_K$.

4. Repeat steps 2 and 3 M times. Afterwards, each W_k represents the value of following the optimal policy (o^*, s^*) for M periods given g_k followers in the first period.

5. For $k = 1 \ldots, K$, let

$$(o_k^{**}, s_k^{**}) = \text{argmax}_{o,s}\; \alpha(f)s - c_e(o, s) - c_m(s) + \beta \int_2^9 C(x)\phi(x; f, o, s, \sigma_f)\; dx$$

6. Update the optimal policy: update $o^*(f)$ to be the piecewise cubic Hermite interpolating polynomial for $g_1, \ldots, g_K$ and $o_1^{**}, \ldots, o_K^{**}$, and update $s^*(f)$ to be the piecewise cubic Hermite interpolating polynomial for $g_1, \ldots, g_K$ and $s_1^{**}, \ldots, s_K^{**}$.

7. Let $\mathbf{V}' = (C(g_1), \ldots, C(g_k))$ (the new value function evaluated at each grid point). Let $\overline{\mathbf{V}}'$ be the maximum of the elements of $\mathbf{V}'$. Define

$$\iota = \frac{\|\mathbf{V} - \mathbf{V}'\|}{\overline{\mathbf{V}}'}$$

8. If $\iota \leq 10^{-5}$, stop and use C as the final value function and (o^*, f^*) as the final policy function. Otherwise, return to step 1.

The following table summarizes the parameter values I use for the simulation:

Parameter	Value	Source
π_0	0.489	Regression
π_1	0.206	Regression
τ_o	0.00019	Regression
τ_s	0.00010	Regression
τ_ℓ	0.00028	Regression
σ_f	0.0079202	Regression
β	0.99	Assumed
η	6	Assumed
θ_e	0.0025	Assumed (will be estimated)
θ_m	9.5	Assumed (will be estimated)

Table 20: Initial parameter values

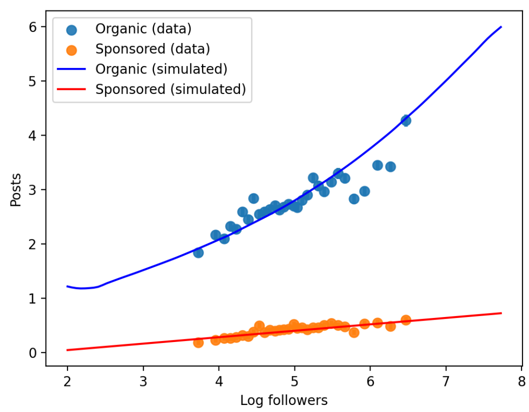


Figure 13: Optimal policy vs data

Figure 13 shows the calculated optimal content choice overlaid on a binscatter of observed choices in the data. My parameter choices fit the data reasonably well, and I will use them as a starting point for structural estimation.

In my model, influencers with more followers are paid more for sponsored content, so calculated optimal production of sponsored posts is not surprising. Why, though, do influencers with more followers make more organic posts? Given the follower transition equation (5), two influencers with different follower counts receive the same *percentage* increase in followers from increasing the number of organic posts they produce. The larger influencer thus gains more followers and will see a larger increase in their payments for sponsored content. As influencers grow, their incentive to grow even more increases.

Figure 14 shows the actual content choices of three influencers with varying follower counts. Posting behavior deviates substantially from the optimal policy my model implies. The largest of the three influencers posts almost three times as much per week as the optimal policy suggests, while in many weeks the mid-sized influencer posts only half as much. A natural extension of my model to rationalize these deviations is

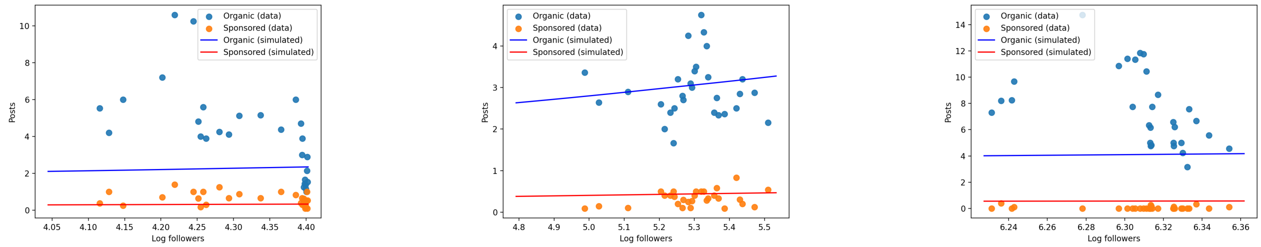


Figure 14: Example policy choices

to allow the effort cost parameter θ_e to vary with follower count. The marginal cost of a post could vary with audience size since, for example, large influencers might hire production teams to write and edit their content.

6 The influencer's career

How does an influencer's career evolve over time? Influencers who grew from unknown to celebrity overnight exist (e.g. Tube Girl), but according to my data and model, they are not the norm. Instead, the influencer's job is to create a steady stream of content that generates slow but positive growth. Viral posts provide a useful boost because two influencers who make the same choices see the same *percentage* growth. After a viral post, an influencer has a larger audience, so constant percentage growth translates to a larger increase in raw follower count. For example, Figure 15 shows follower count history for @ryanpeterspgh, a recipe

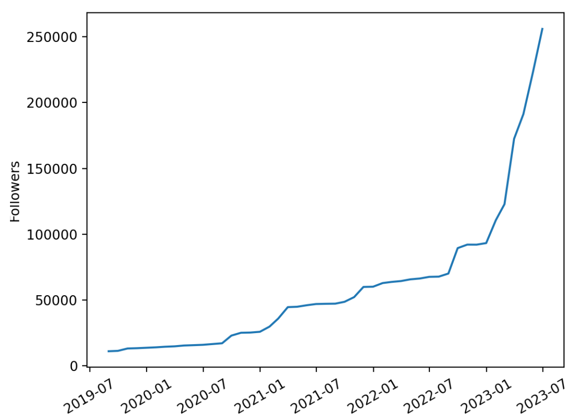


Figure 15: Followers over time for @ryanpeterspgh

creator with 731,000 Instagram followers today. There are periods of rapid growth, likely generated by viral posts. After those periods, growth seems to return to its previous trend. Although viral posts do not change the shape of the influencer's growth curve, the influencer sees larger raw increases in follower count after the viral post because their percentage growth is the same and their audience is larger.

My model generates a similar pattern. Figure 16 simulates the model for 200 weeks for two influencers who both start with 100,000 followers. They receive (different) follower count shocks in each period drawn from the distribution described above, and I artificially introduce a "viral post" by giving one of them (the

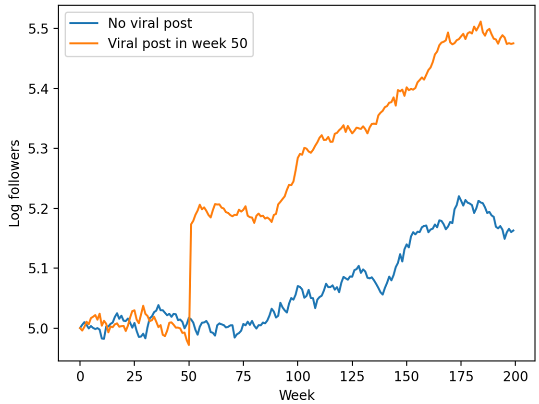


Figure 16: Follower growth over time

orange line) a large, positive shock in week 50. The influencer with the shock sees about a 25% increase in followers as a result, but the shock is transient in the sense that it does not change the shape of the curve. It shifts it up, at which point the curve returns to its pre-shock trend. Although the shape of the curve does not change, the gap in raw follower count grows after the viral post. In week 75, the influencer with the viral post has about 58,000 more followers than the influencer without a viral post. In week 175, the gap increases to about 92,000 followers.

How much effort is required to produce content, and what are the rewards? In 2021, HypeAuditor, a social media marketing firm, surveyed influencers about their work hours and income (Baklanov (2021)). The survey estimated that influencers with 1,000 to 10,000 followers earn about $1,420 per month, while influencers with 500,000 to 1,000,000 followers earn $5,847 per month. In my model, an influencer with 10,000 followers behaving optimally should produce about two sponsored posts per month; this influencer receives $147 per sponsored post according to my payment data, so their monthly income is about $300. An influencer with 1,000,000 followers should produce three sponsored posts per month and receives $1,439 per post for a monthly income of about $4,500. My model might underestimate true pay because I only account for revenue from sponsored posts. Influencers have other sources of income, like Patreon, that I do not observe but that could increase their total earnings. On the other hand, influencers might overreport their income in surveys.

According to the survey, influencers with under 1,000,000 followers spend about 30 hours per week on their content (including audience interactions, negotiating with sponsors, etc.), while the largest influencers spend 40-50 hours per week. Pay increases with follower count far faster than hours worked: taking revenue numbers from the survey, an influencer with 1,000,000 followers earns $30 per hour, while an influencer with 10,000 followers earns about $13 per hour. Relying only on income from sponsored Instagram content requires nearly a seven-figure follower count to earn a living. This partly explains why influencers use multiple platforms, start websites, and try to develop their own products. Much of the value of follower growth on Instagram comes from leveraging the fame elsewhere.

There is little variation in the time it takes to produce a post as an influencer grows. Assuming optimal behavior from my model, an influencer with 1,000,000 followers makes about three posts per week. They spend 32 hours doing so according to the HypeAuditor survey, yielding about 10 hours per post. An influencer with 10,000 followers spends 27 hours making 2.5 posts per week on average, yielding nearly identical hours per post. The largest content creators, especially on YouTube, have dedicated content creation teams (editors, writers, etc). If larger influencers in my model made far more posts than small influencers or if hours per post became unrealistically small as follower count increased, I could conclude that influencers in my data outsource some of their work. Instead, I see no variation in hours per post. The influencers in my data (mostly under 1,000,000 followers) probably make content on their own.

The above calculations suggest that the marginal cost of a post, measured as hours per post, does not vary with follower count. This is consistent with my model: neither c_e nor c_m depends on followers. Instead, posting is costly because it takes time and because ideas get harder to find as they are exhausted. An influencer with more followers is willing to spend more time searching for post ideas because they are rewarded more highly by brands for their sponsored posts.

6.1 Invest and harvest?

Theoretical literature modeling influencers (Mitchell (2021), Nistor et al. (2024)) predicts distinct periods of "investing" and "harvesting". When investing, an influencer avoids making sponsored content in order to grow their audience. After enough growth, the influencer harvests by making sponsored content in exchange for pay from brands.

Because I model production of organic and sponsored content separately, harvesting can mean two things. Making more sponsored content is a form of harvesting, and it increases linearly as an influencer grows. The *fraction* of sponsored content is another measure of harvesting. If followers are happy to see sponsored content so long as it is accompanied by sufficient organic content, then increased sponsored posting does not necessarily harvest followers' goodwill. The influencer can simultaneously increase organic content to keep their audience happy (although this requires additional effort). Harvesting occurs when the fraction of sponsored content increases. Figure 17 plots the influencer's optimal policy as the fraction of sponsored content.

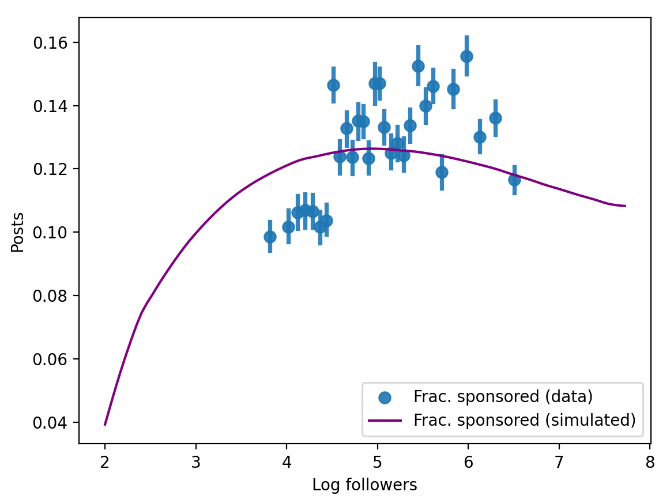


Figure 17: Optimal policy (frac. sponsored) vs. data

Both my model and the data show increased harvesting as the influencer gains followers. There is not, as in the primary model of Nistor et al. (2024), a distinct point at which the influencer begins harvesting.

The calculated optimal policy is more consistent with their extension in which influencers can continuously choose a fraction of poor fit sponsorship offers to accept.

When measured as the fraction of sponsored content, my model implies that larger influencers harvest more up to about 100,000 followers, after which harvesting declines slightly. The incentives are different from existing models of influencers. There is no follower count penalty to sponsored content that would encourage influencers to build an audience before beginning to harvest. Instead, as an influencer grows, sponsored content becomes more attractive because more followers generate larger payments, which outweigh the additional cost of producing more sponsored posts. Beyond 100,000 followers, the reward in terms of future followers for organic posts outweighs the additional pay for sponsored posts, and the fraction of sponsored content begins to decline.

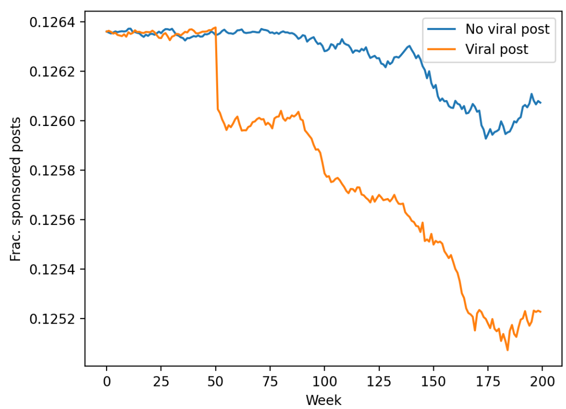


Figure 18: Optimal policy (frac. sponsored posts) over time

Figure 18 plots the fraction of sponsored posts an influencer following the optimal policy produces over four years (starting with 100,000 followers). Again, a viral post causes a discrete jump in the influencer's follower count, which generates an abrupt change in the optimal policy, but the viral post does not fundamentally change the shape of the curve. There is a slighty decrease in harvesting over time: the influencer produces more sponsored posts and more organic posts, but the latter outweighs the former. The magnitude of the decrease is very small: for the influencer with the viral post, the fraction of sponsored posts decreases from 12.6% to 12.5% from start to end.

Overall, simulations reveal a key fact supported by data: rather than alternate between investing and harvesting or switching to harvesting at a distinct point, influencers smoothly increase both the raw quantity and fraction of sponsored content they make as they gain followers. Testing how this pattern changes under counterfactual scenarios requires structural estimation.

7 Estimation

I use the method of simulated moments to estimate the cost parameters θ_e and θ_m. I simulate the careers of many influencers with different starting follower counts. I assign the simulated follower counts to bins $b = 1, \ldots, B$ and calculate the average number of organic and sponsored posts in each bin across all simulated observations. Denote these averages $\overline{o}_b^*$ and $\overline{s}_b^*$ respectively. The moment conditions are based on these simulated averages. Let o_{it} be the number of organic posts individual i makes in period t and let s_{it} be the number of sponsored posts. The moment conditions for each bin b are

$$\mathbb{E}\left[o_{it} - \overline{o}_b^* \mid \text{individual } i \text{ in bin } b \text{ at time } t\right] = 0$$
$$\mathbb{E}\left[s_{it} - \overline{s}_b^* \mid \text{individual } i \text{ in bin } b \text{ at time } t\right] = 0$$

yielding $2B$ conditions in total. Define

$$g^b(o_{it}, s_{it}, \theta_e, \theta_m) = \begin{bmatrix} (o_{it} - \overline{o}_b^*)\mathbb{1}(\text{individual } i \text{ in bin } b \text{ at time } t) \\ (s_{it} - \overline{s}_b^*)\mathbb{1}(\text{individual } i \text{ in bin } b \text{ at time } t) \end{bmatrix}$$

Let $g(o_{it}, s_{it}, \theta_e, \theta_m)$ be the $2B \times 1$ column vector formed by stacking $g^1, \ldots, g^B$. The unconditional moment conditions are then given by

$$\mathbb{E}\left[g(o_{it}, s_{it}, \theta_e, \theta_m)\right] = 0$$

I calculate sample analogs by summing over individuals $i = 1, \ldots, I$ and time periods $t = 1, \ldots, T$. For example, for bin b and parameters (θ_e, θ_m), the sample analog of $g^b(o_{it}, s_{it}, \theta_e, \theta_m)$ is

$$g_{IT}^b(\theta_e, \theta_m) = \begin{bmatrix} \frac{1}{IT}\sum_{t=1}^T \sum_{i=1}^I o_{it}\mathbb{1}(i \text{ in } b \text{ at } t) - \frac{\overline{o}_b^*}{IT}\sum_{t=1}^T \sum_{i=1}^I \mathbb{1}(i \text{ in } b \text{ at } t) \\ \frac{1}{IT}\sum_{t=1}^T \sum_{i=1}^I s_{it}\mathbb{1}(i \text{ in } b \text{ at } t) - \frac{\overline{s}_b^*}{IT}\sum_{t=1}^T \sum_{i=1}^I \mathbb{1}(i \text{ in } b \text{ at } t) \end{bmatrix}$$

Let $g_{IT}(\theta_e, \theta_m)$ be the $2B \times 1$ column vector formed by stacking $g_{IT}^1, \ldots, g_{IT}^B$.

The estimation proceeds as follows:

1. Set $S = 1000$ to be the number of simulations and $T = 200$ the number of periods per simulation. Draw starting follower counts f_s^0 for $s = 1, \ldots, S$ from the empirical distribution of follower counts on June 30, 2019. For $s = 1, \ldots, S$ and $t = 1, \ldots, T$, draw follower transition shocks $\epsilon_{st} \sim \mathcal{N}(0, \sigma_f^2)$.

2. Fix a guess of (θ_e, θ_m). I start with $(0.0025, 9.5)$ from Table 20 above.

3. Use the algorithm described above to calculate the optimal policy. Let $o^*(f; \theta_e, \theta_m)$, $s^*(f; \theta_e, \theta_m)$, and $\ell^*(f; \theta_e, \theta_m)$, denote the optimal numbers of organic posts, sponsored posts, and likes, respectively, for an influencer with f followers, given cost parameters θ_e and θ_m.

4. Simulate the careers and content choices of each influencer s for T weeks as follows:

   (a) Set $f_{st} = f_s^0$

   (b) Use the optimal policy to find $o^*(f_{st}; \theta_e, \theta_m)$ and $s^*(f_{st}; \theta_e, \theta_m)$.

   (c) Set

   $$\log f_{st+1} = \log f_{st} + \tau_o o^*(f_{st}; \theta_e, \theta_m) + \tau_s s^*(f_{st}; \theta_e, \theta_m) + \tau_\ell \ell^*(f_{st}; \theta_e, \theta_m) + \epsilon_{st}$$

   (d) Repeat steps (b) and (c) to generate $f_{s1}, \ldots, f_{sT}$.

5. Generate $B = 15$ equally sized follower count bins. For each bin b, calculate $\overline{o}_b^*$ and $\overline{s}_b^*$.

6. Let

   $$g_{IT}(\theta_e, \theta_m) = \left[g_{IT}^1(\theta_e, \theta_m), \ldots, g_{IT}^B(\theta_e, \theta_m)\right]^\top$$

7. Repeat steps (2)-(6) to find $(\widehat{\theta}_e, \widehat{\theta}_m)$ that minimize

   $$g_{IT}(\theta_e, \theta_m)^\top g_{IT}(\theta_e, \theta_m)$$

8. Calculate the weighting matrix

$$\widehat{W} = \left(\frac{1}{IT} \sum_{t=1}^{T} \sum_{i=1}^{I} \left(g(o_{it}, s_{it}, \widehat{\theta}_e, \widehat{\theta}_m) - g_{IT}(\widehat{\theta}_e, \widehat{\theta}_m) \right) \left(g(o_{it}, s_{it}, \widehat{\theta}_e, \widehat{\theta}_m) - g_{IT}(\widehat{\theta}_e, \widehat{\theta}_m) \right)^{\top} \right)^{-1}$$

9. Repeat steps (2)-(6) to find $(\widehat{\theta}_e, \widehat{\theta}_m)$ that minimize

$$g_{IT}(\theta_e, \theta_m)^{\top} \widehat{W} g_{IT}(\theta_e, \theta_m)$$

7.1 Results

Parameter	Estimate
θ_e	0.0034050
θ_m	9.84327

Table 21: Estimated parameter values

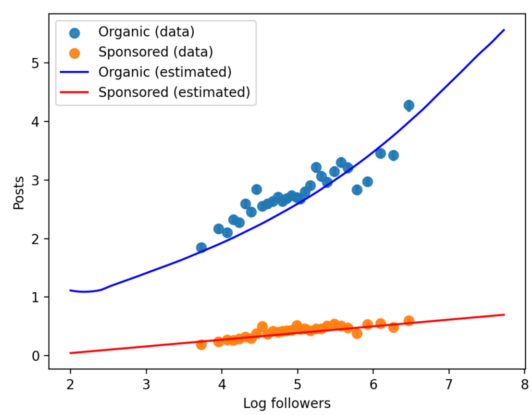


Figure 19: Optimal policy vs data (estimated parameters)

Table 21 shows the estimation results, and Figure 19 plots the corresponding optimal policy along with the data.

8 Counterfactuals

8.1 Negative impact of sponsored content on follower growth

A key difference between existing theoretical literature and my simulations is that sponsored content does not reduce follower growth. How would creator behavior change if it did? Figure 20 shows the optimal policy using the parameters from Table 20 except τ_s, which I set to -0.00194 to match the negative impact of sponsored content estimated in Cheng and Zhang (2022). Compared to 13, there is almost no change in sponsored content production, while there is a small decrease in organic content. The fraction of sponsored content therefore goes up. This behavior is similar to that described in Mitchell (2021). Increasing the

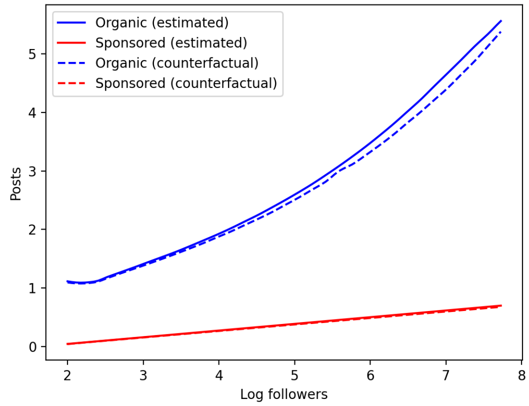


Figure 20: Optimal policy vs data: dynamic cost for sponsored content

growth penalty for sponsored content lowers the return to organic content because the influencer's incentive to produce organic content is future growth.

Figure 21 plots the marginal cost and marginal benefit of organic and sponsored posts under the original estimated parameter values and under the new values (setting τ_s to -0.00194). When plotting each curve, I

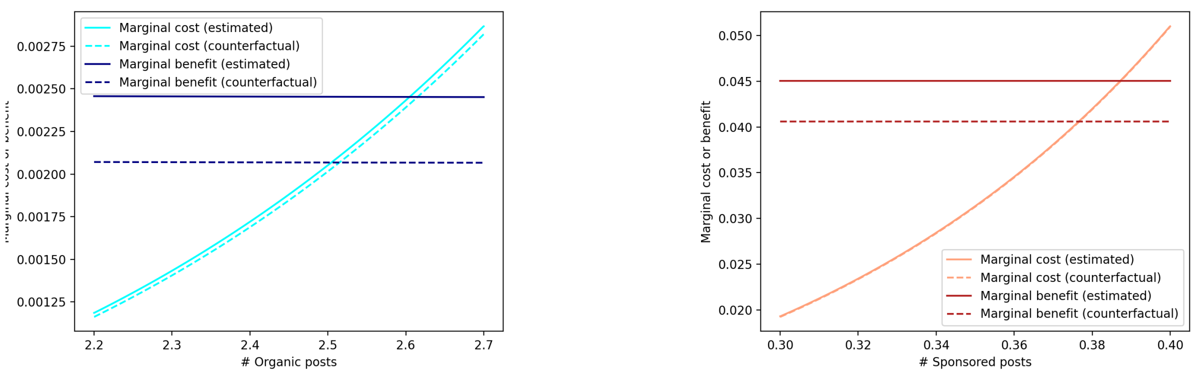


Figure 21: Marginal cost/marginal benefit: dynamic cost for sponsored content

fix follower count at 100,000, and I use the optimal policy to fix the number of sponsored posts. Then, I vary the number of organic posts and calculate marginal cost and marginal benefit. Because the only component of flow utility that depends on the number of organic posts is the effort cost c_e, the marginal cost of an organic post is

$$\frac{\partial}{\partial o} c_e(o, s) = \theta_e \eta (o + s)^{\eta - 1} \tag{6}$$

Organic posts do not generate utility in the current period, but they increase follower count in the future which yields future payment. The marginal benefit of an organic post is therefore the derivative of the expected next-period value function:

$$\frac{\partial}{\partial o} \left[\beta \int_2^9 V(x) \phi(x; f, o, s, \sigma_f) \, dx \right] \tag{7}$$

I approximate this gradient numerically. Similarly, the marginal cost of a sponsored post is

$$\frac{\partial}{\partial s}\left[c_e(o,s)+c_m(s)\right]=\theta_e\eta(o+s)^{\eta-1}-\theta_m e^{\theta_m s} \tag{8}$$

Sponsored posts generate utility both in the current period (through payment) and in the future (through follower growth). The marginal benefit is therefore

$$\frac{\partial}{\partial s}\left[\alpha(f)s+\beta\int_2^9 V(x)\phi(x;f,o,s,\sigma_f)\ dx\right] \tag{9}$$

which I again approximate numerically.

Although neither θ_e nor θ_m changes in this counterfactual, the marginal cost curves shift downward slightly because the optimal number of sponsored posts given $100,000$ followers decreases. Sponsored posts generate negative follower growth, so their marginal benefit is reduced. The marginal benefit remains positive because pay for a sponsored post outweighs the loss of future followers. The marginal cost curve for sponsored posts is steep, so the reduction in marginal benefit does not significantly change the influencer's optimal choice of sponsored posts, as is apparent in Figure 20.

Given the transition equation 1, organic posts increase follower growth in percentage terms. Let

$$\log f'_{t+1}=\log f_t+\tau_o(o_t+1)+\tau_s s_t+\tau_\ell \ell(o_t+1,s_t)+\varepsilon_t$$
$$\log f_{t+1}=\log f_t+\tau_o(o_t)+\tau_s s_t+\tau_\ell \ell(o_t,s_t)+\varepsilon_t$$

Then for a fixed follower count and a fixed number of sponsored posts, the increase in next period followers from one additional organic post is

$$\log f'_{t+1}-\log f_{t+1}=\tau_o+\tau_\ell(\ell(o_t+1,s_t)-\ell(o_t,s_t)) \tag{10}$$

where $\ell(o_t,s_t)$ is likes calculated from 4. The ratio of f'_{t+1} to f_{t+1} does not depend on τ_s, but their values do. Reducing τ_s reduces f_{t+1}, so applying the growth rate from 10 yields a smaller change in raw follower count (e.g. a 1% increase from 1000 followers is 10 additional followers while a 1% increase from 10,000 followers is 100 additional followers). The smaller raw increase means a smaller increase in future payments for sponsored content, so the marginal benefit of an organic post goes down when τ_s decreases. The marginal cost curve for organic posts is flatter than for sponsored posts, so the reduction in marginal benefit generates a larger change in the optimal choice of organic posts.

Organic posts affecting the percentage change in follower count makes sense given the ways potential followers might discover an influencer. Some proportion of existing followers will share the post with friends, and some of those friends will become followers. Posts from influencers with more followers tend to get more likes, and posts with more likes are probably spread more widely by content distribution algorithms. Overall, the number of new followers acquired from a post is approximately some fraction of current follower count. Influencers with more followers therefore see larger raw increases in follower count when they post.

8.2 Increased match costs

On the other hand, how do influencers respond to an increase in the cost of producing sponsored content? The FTC currently requires influencers to clearly disclose all sponsored posts, but studies suggest that proper disclosure is rare (Ershov et al. (2023)). More stringent monitoring would likely increase the cost of producing a sponsored post for influencers. For example, the FTC might require influencers to log their brand partnerships in a database, so sponsored content would take more time to make. In my model this amounts to an increase in θ_m because there is an increase in the marginal cost of a sponsored post but not of an organic post. Figure 22 shows the optimal policy after increasing θ_m from its estimated value

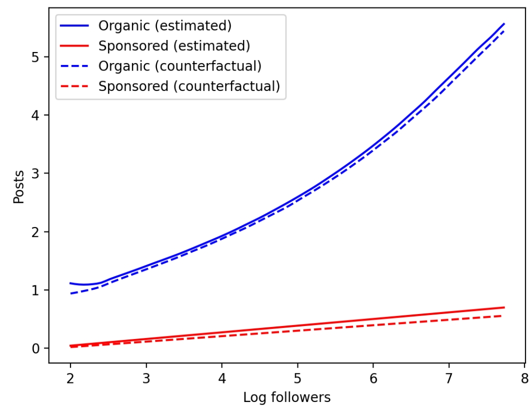


Figure 22: Optimal policy vs data: increased cost of sponsored content

(9.84) to 12, about a 20% increase. A creator with 100,000 followers produced about 0.4 sponsored posts per week under the original parameter values; they now produce about 0.3 per week, a 25% reduction. Organic content declines slightly, from about 2.6 posts per week to about 2.55. The fraction of sponsored content decreases slightly (Figure 24). Since sponsored posts are more costly and pay has not changed, profits for the influencer decrease. Organic posts generate follower growth that yields future profits, so the return to an organic post also decreases.

Figure 23 plots the marginal benefit and marginal cost curves for organic and sponsored posts under the estimated parameters and under the counterfactual with increased θ_m. The marginal benefit from an

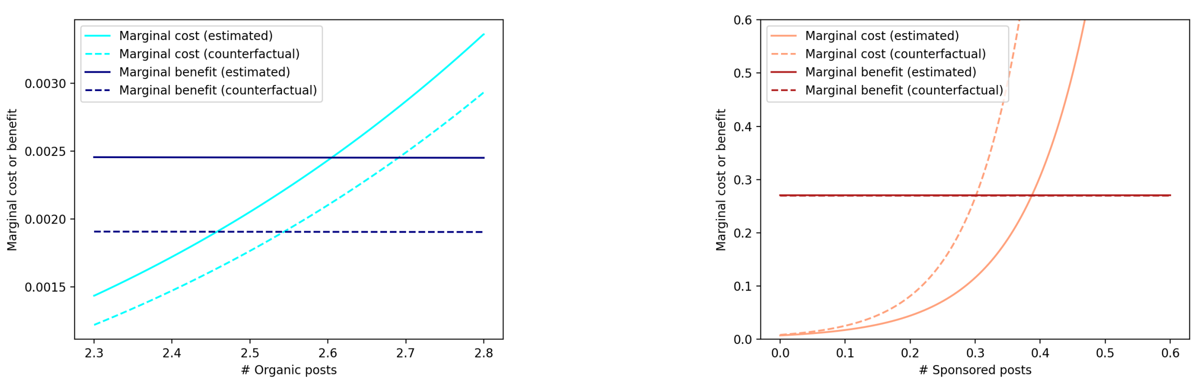


Figure 23: Marginal cost/marginal benefit: increased match cost

organic post increases slightly because organic posts generate future followers from whom the influencer now makes less profit. On the other hand, since sponsored content production decreases, the influencer moves leftward along the effort cost curve c_e, so the marginal cost of an organic post decreases. The net effect (for an influencer with 100,000 followers) is a slight decrease in the number of organic posts produced. The marginal cost curve for sponsored content shifts upward because of the increase in θ_m, so the optimal number of sponsored posts decreases.

This policy change generates a decrease in revenue for influencers since they make more sponsored posts and profit more from each one. The impact on consumer welfare depends on consumer preferences for organic vs. sponsored content, since they see fewer sponsored posts but also fewer organic posts. However, a primary concern the theoretical literature (e.g. Mitchell (2021)) raises is that increasing the cost of sponsored content could significantly reduce production of non-sponsored content because of dynamic incentives. While this concern does arise in my model, its effects are small because dynamic incentives are small in the data.

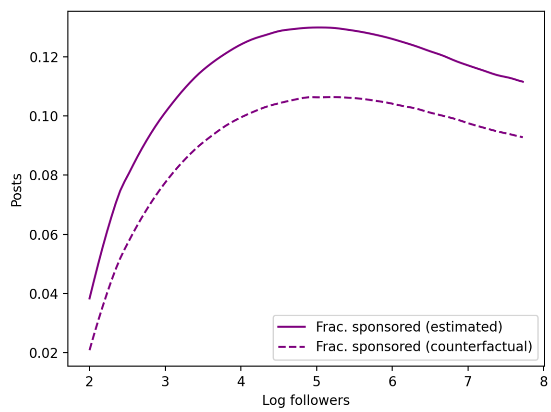


Figure 24: Optimal frac. sponsored content vs data: increased cost of sponsored content

8.3 Static model

Influencers might not actually behave dynamially. One influencer told me they do not consider long term career concerns when choosing whether to accept sponsorship offers. To test this idea, I run a counterfactual setting the discount factor β to zero, so the influencer only cares about current period utility, not the future. As expected, influencers stop producing organic posts (Figure 25). The only incentive to create them in my model is follower growth that translates into future revenue. When the influencer does not care about future revenue, there is no reason to make non-sponsored content.

Sponsored content production does not change, which means the tradeoff between pay and production costs entirely determines the influencers choice. Although sponsored posts generate positive follower growth, the effect is negligible relative to short-run considerations. A single sponsored post has little effect on follower growth given the estimate in Table 12. For example, one additional sponsored post turns a 1% increase in followers into a 1.024% increase in followers, or 24 additional followers for an influencer with 100,000 followers. Those 24 followers increase pay per deliverable by about five cents according to Table 1. The same influencer earns about $447 per sponsored post. The marginal cost curve for sponsored content is also very steep given the exponential form of c_m. The magnitudes of current-period costs and benefits of sponsored content are much larger than the future payoff, so the influencer puts almost zero weight on the future even when the discount factor is not zero.

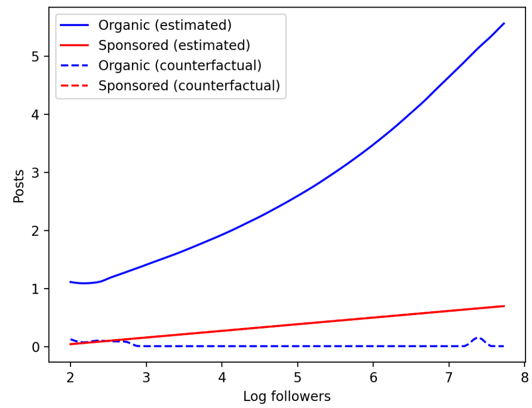


Figure 25: Optimal policy vs data: static model

References

Bairathi, Mimansa and Anja Lambrecht (2023), "Influencer marketing: Sponsorship disclosure and authenticity." *Available at SSRN 4679776*.

Baklanov, Nick (2021), "Our survey data of 1865 instagram influencers." URL `https://hypeauditor.com/blog/data-from-our-study-of-1865-instagram-influencers/`.

Berger, Jonah and Katherine L Milkman (2012), "What makes online content viral?" *Journal of marketing research*, 49, 192–205.

Cascio Rizzo, Giovanni Luca, Francisco Villarroel Ordenes, Rumen Pozharliev, Matteo De Angelis, and Michele Costabile (2024), "How high-arousal language shapes micro-versus macro-influencers' impact." *Journal of Marketing*, 00222429231207636.

Cheng, Mengjie Magie and Shunyuan Zhang (2022), "Reputation burning: Analyzing the impact of brand sponsorship on social influencers." *Available at SSRN 4071188*.

Ershov, Daniel, Yanting He, and Stephan Seiler (2023), "How much influencer marketing is undisclosed? evidence from twitter."

Ershov, Daniel and Matthew Mitchell (2020), "The effects of influencer advertising disclosure regulations: Evidence from instagram." In *Proceedings of the 21st ACM Conference on Economics and Computation*, 73–74.

FYPM (2024), "Frequently asked questions." URL `https://info.fypm.vip/faq`.

Geyser, Werner (2022), "Influencer rates: How much do influencers really cost in 2022?" URL `https://influencermarketinghub.com/influencer-rates/`.

Giuffredi-Kähr, Andrea, Alisa Petrova, and Lucia Malär (2022), "Sponsorship disclosure of influencers–a curse or a blessing?" *Journal of Interactive Marketing*, 57, 18–34.

Gong, Zheng (2021), "Growing influence." Technical report, working paper, University of Toronto.

Hughes, Christian, Vanitha Swaminathan, and Gillian Brooks (2019), "Driving brand engagement through online social influencers: An empirical investigation of sponsored blogging campaigns." *Journal of marketing*, 83, 78–96.

Judd, Kenneth L (1998), *Numerical methods in economics.* MIT press.

Leung, Fine F, Flora F Gu, Yiwei Li, Jonathan Z Zhang, and Robert W Palmatier (2022), "Influencer marketing effectiveness." *Journal of marketing*, 86, 93–115.

Li, Ivan (2023), "Sponsorships in livestreaming: Monetization and disclosure behavior of influencers on twitch."

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

Nelson, Phillip (1974), "Advertising as information." *Journal of political economy*, 82, 729–754.

Nistor, Cristina, Matthew Selove, and J Miguel Villas-Boas (2024), "Influencers: To grow or to monetize." *Available at SSRN 4708107*.

Novacic, Ines (2019), ""it's kinda crazy": Kid influencers make big money on social media, and few rules apply." URL `https://www.cbsnews.com/news/kid-influencers-instagram-youtube-few-rules-big-money-cbsn-originals`.

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

Řehůřek, Radim and Petr Sojka (2010), "Software Framework for Topic Modelling with Large Corpora." In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, 45–50, ELRA, Valletta, Malta. `http://is.muni.cz/publication/884893/en`.

Sahni, Navdeep S and Harikesh S Nair (2020), "Does advertising serve as a signal? evidence from a field experiment in mobile search." *The Review of Economic Studies*, 87, 1529–1564.

Statista (2024a), "Daily time spent on social networking by internet users worldwide from 2012 to 2024." URL `https://www.statista.com/statistics/433871/daily-social-media-usage-worldwide/`.

Statista (2024b), "Social media advertising." URL `https://www.statista.com/outlook/dmo/digital-advertising/social-media-advertising/worldwide`.

Tang, Qian, Bin Gu, and Andrew B Whinston (2012), "Content contribution for revenue sharing and reputation in social media: A dynamic structural model." *Journal of Management Information Systems*, 29, 41–76.

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

Valsesia, Francesca, Davide Proserpio, and Joseph C Nunes (2020), "The positive effect of not following others on social media." *Journal of Marketing Research*, 57, 1152–1168.

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

9 Appendix

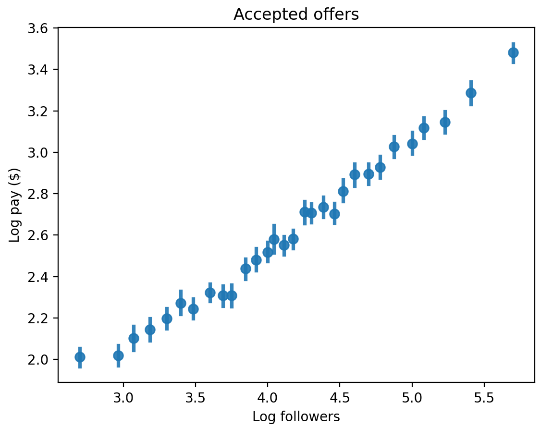


Figure 26: Pay vs follower count

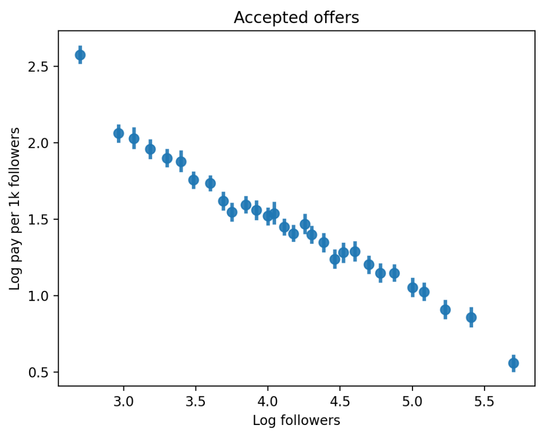


Figure 27: Pay per follower vs follower count

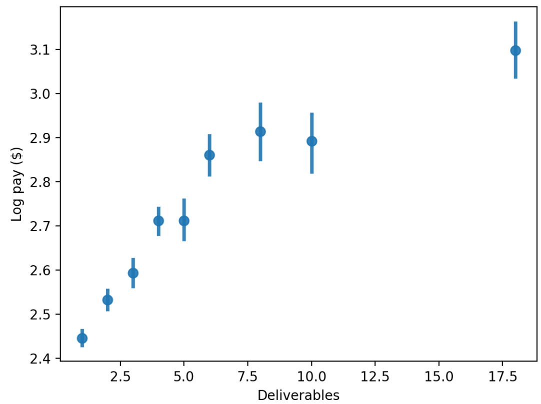


Figure 28: Pay vs deliverables, accepted offers only

	Log pay per deliverable	Log pay per deliverable
Log Instagram followers	0.509***	0.502***
	(0.007)	(0.007)
Intercept	0.114***	0.607
	(0.031)	(0.414)
Date FE	No	Yes
N	11,044	11,044
R2	0.298	0.311

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 22: Dependence of pay on Instagram follower count

Word
#ad
gift
gifted
patron
ambassad
ambassador
collaboration
partnership
collab
sponsored
sponsor
promo
partner
publicit
promotion
advertisement
publi

Table 23: Ershov and Mitchell (2020) manual “disclosed” classification

Word
.com
@
available
link
must
have
shop
buy
now
code
%
$
contest
even
launch
tonight
ship
hotel
campa
follow
until
official
video
new
thank
thanks
diet
shake
detox
smoothi
supplement
protein
tea
drink
health

Table 24: Ershov and Mitchell (2020) manual "sponsored" classification

	Change log followers (I)	Change log followers (I)
Posted * Log likes	0.001077***	0.001098***
	(0.000243)	(0.000241)
Has additional organic post (I)	-0.001112*	-0.001054*
	(0.000453)	(0.000449)
Change log followers (T)		0.059037***
		(0.008356)
Creator FE	Yes	Yes
Date FE	Yes	Yes
N	3,836	3,836
R2	0.00766	0.0221

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 25: Effect of an organic post on follower growth

	Change log followers (I)	Change log followers (I)
Posted * Log likes	0.000972***	0.000776***
	(0.000166)	(0.000170)
Has additional sponsored post (I)	-0.001761**	-0.001395*
	(0.000635)	(0.000631)
Change log followers (T)		0.082097***
		(0.018164)
N	625	625
R2	0.079	0.108

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 26: Effect of a sponsored post on follower growth