IIMB-WP N0. 580/2022

WORKING PAPER NO: 580

Influencer Marketing with Fake Followers

Abhinav Anand
Assistant Professor
Finance and Accounting
Indian Institute of Management Bangalore
Bannerghatta Road, Bangalore – 5600 76
abhinav.anand@iimb.ac.in

Souvik Dutta
Assistant Professor
Social Sciences
Indraprastha Institute of Information Technology
Delhi - 110020
souvik@iiitd.ac.in

Prithwiraj Mukherjee
Assistant Professor
Marketing
Indian Institute of Management Bangalore
Bannerghatta Road, Bangalore – 5600 76
pmukherjee@iimb.ac.in

Year of Publication – March 2022

Influencer marketing with fake followers*

Abhinav Anand† Souvik Dutta‡ Prithwiraj Mukherjee§

March 16, 2022

Abstract

Influencer marketing is a practice in which an advertiser pays a popular social media user (influencer) in exchange for brand endorsement. We develop a contract-theoretic model of an advertiser and an influencer. The influencer can inflate her publicly displayed follower count by buying fake followers, and can take a hidden action to legitimately increase her true number of followers. An imperfect audit can detect fraud, which imposes a high cost on the influencer. We show that the optimal contract exhibits high faking for influencers with intermediate follower counts, while faking levels are low for those with very small or very large true follower counts. Audits deter fraud only when accompanied by high penalties, but perversely, restitution paid to the advertiser encourage more fraud.

Keywords: Digital marketing, social media, influencer marketing, fake followers, optimal control, contract theory, moral hazard.

*We thank Bruno Badia, Manaswini Bhalla, Pradeep Chintagunta, Tirthatanmoy Das, Sreelata Jonnalagedda, Manish Kacker, Ashish Kumar, Sanjog Misra and participants at the Chicago Booth-India Quantitative Marketing Conference (2018) at IIM Bangalore.

†Indian Institute of Management Bangalore,

‡Indraprastha Institute of Information Technology Delhi,

§Indian Institute of Management Bangalore.

"At Unilever, we believe influencers are an important way to reach consumers and grow our brands. Their power comes from a deep, authentic and direct connection with people, but certain practices like buying followers can easily undermine these relationships." — Keith Weed, Chief Marketing and Communications Officer, Unilever (Stewart, 2018)

1 Introduction

Advertisers often pay popular social media users known as "influencers" to endorse their products online. Many of these influencers have large numbers of self-selected followers who share their interests (travel, fashion, cooking, etc.) and look up to them for advice in these domains. According to The Economist (2016), YouTube influencers with over 7 million followers command upto $300,000 per sponsored post, while the corresponding figures for Instagram, Facebook and Twitter are $150,000, $187,500 and $60,000 respectively, allowing social media followings to be monetized lucratively. Even influencers with fewer than 250,000 followers can make hundreds of dollars per sponsored post. Figure 1 shows some typical compensations for influencers on various platforms versus their follower counts. A Linqia (2018) survey across sectors including consumer packaged goods, food and beverage and retail in the US finds that 86% of marketers surveyed used some form of influencer marketing in 2017, and out of them, 92% reported finding it effective. 39% of those surveyed planned to increase their influencer marketing budgets. Similar trends reported by eMarketer (2017) and IRI (2018) suggest that influencer marketing is growing.

Insert figure 1 about here

Influencer marketing has led to the emergence of shady businesses called "click farms" which offer fake followers to influencers for a price, inflate the number of "likes" on their fan pages, and post spurious comments on their posts. Influencers use these services to fraudulently command higher fees from advertisers for promotional posts. A New York Times exposé finds that several influencers have bought fake followers from a prominent click farm (Confessore et al., 2018).

Sway Ops, an influencer marketing agency estimates the total magnitude of influencer fraud to be about $1 billion (Pathak, 2017). They find that in a single day, of 118,007 comments sampled on #sponsored or #ad tagged Instagram posts, less than 18% were made by genuine users. Another study by the Points North Group finds that influencers hired by Ritz-Carlton have 78% fake followers (Neff, 2018). The corresponding numbers for Procter and Gamble's Pampers and Olay brands are 32% and 19% respectively.

In a very short time, influencer marketing has emerged as a mainstream digital marketing practice. Naturally its many nuances have not been studied in depth yet. Not only are naïve advertisers cheated of potentially millions of dollars of ad budgets, but this practice also calls into question the promise of behavioral targeting and supposedly more accurate ad metrics (reach, engagement, impressions) that digital ads are supposed to deliver over their traditional offline counterparts (Gordon et al., 2020). High profile executives like Estée Lauder's chief executive officer (Stewart, 2019), Unilever's chief marketing officer (Stewart, 2018) and Kellogg's social media lead in UK & Ireland (Joseph, 2018), have all expressed concern about the fake follower problem, seeking mechanisms for marketers to understand and deal with it.

Much extant research on the fake follower problem is devoted to machine learning approaches to detect them (e.g. Cresci et al., 2015; Zhang and Lu, 2016; Khalil et al., 2017). Our approach however is complementary; we study the fake follower problem from economic first principles. Our game unfolds in a principal (advertiser)-agent (influencer) setup, where the former proposes to use the latter's social media endorsement based on the promise of earnings conditioned on her publicly displayed (possibly inflated) follower count. Such faking is costly to the influencer, not only because of the cost of buying fake followers, but extra expenditures needed to sustain the deception by buying fake engagement (likes, shares, comments etc.). Further, the influencer can also expend resources on unobservable effort (hidden action) where she invests in analytics or enhances her content quality to legitimately increase her follower count. The advertiser must take into account this moral hazard, and incentivize high levels of this hidden action accordingly.

All of this happens in the presence of an imperfect, non-strategic third party audit by a regulator. If the agent is caught, the regulator imposes fines, some of which is returned to the advertiser as restitution. The influencer also suffers additional reputational losses on detection of her fraud.

Drawing from analogous literature on insurance and accounting fraud, we use control theory to solve for the optimal *ex ante* contract between the advertiser and influencer. We find the existence of widespread faking across all types of influencers except the highest and lowest types. We complement our game-theoretic model with extensive policy simulations to demonstrate how audits and accompanying penalties deter fraud. We show that accurate audits only work when accompanied by sufficient penalties associated with them. Further, the advertiser is better off if these penalties are appropriated by the regulator rather than returned to it as restitution. Perversely, a restitution-to-advertiser-on-cheating clause in the contract encourages more fraud by the influencer.

The rest of this paper is organized as follows. Section 2 provides a brief overview of influencer marketing and related issues, section 3 outlines our principal (advertiser)-agent (influencer) setup and solution for the optimal contract, and section 4 outlines extensive policy simulations to show audits accuracy, penalties and principal's restitution affect the influencer's fraud level. Finally, section 5 discusses the model's implications for various stakeholders in the influencer marketing ecosystem.

2 Overview

Since the practice of influencer marketing is relatively new, we present a brief overview of this domain, linking it to established theories of endorsement and social influence. We divide this into four parts where we, (i) cover the general practice of influencer marketing, (ii) contextualize it within extant research on influence in social networks, (iii) discuss the role of follower count in influencer marketing, and (iv) describe how it leads to the specific problem of fake followers.

2.1 Influencer marketing

Influencer marketing is a promotion method that features brand endorsement using popular social media users. This makes it a hybrid of advertising and word-of-mouth. Nowadays, celebrities like actor Priyanka Chopra (Tiffany and co. jewelry), racing professional Danica Patrick (Lyft ride sharing), rapper Snoop Dogg (Tanqueray gin) star in online influencer campaigns alongside expert influencers like gamer H2ODelirious (Ubisoft gaming) and fashion blogger Jaclyn Hill (Becca cosmetics), and lay influencers like blogger Kelly Lund's pet Loki the Wolf Dog (Mercedes-Benz, Toyota automobiles). Specialist agencies like Bzzagent also enlist ordinary netizens as "nano-influencers" to promote usually inexpensive consumer brands like chocolates and yoghurt in return for free product samples (Berger, 2016). Figure 2 provides snapshots of some famous influencer campaigns on social media.

Insert figure 2 about here

As influencer marketing is often promotional material in the guise of word-of-mouth, disclosure of sponsored content is a major issue. The US Federal Trade Commission has taken cognizance of the possibility of consumers being misled by influencer marketing, and has mandated that sponsored posts must now clearly mention the relationship between the influencer and brand, usually as a hashtag such as #sponsored or #ad. Platforms like Instagram have implemented algorithms to automatically detect and tag paid influencer posts (FTC, 2019).

2.2 Social networks and influence

The study of influence in social networks is another well established research area relevant to influencer marketing. The effects of network structure, personality and contextual factors are extensively documented in several fields like marketing, sociology and network science. Influentials on social media are characterized by high reach and spread of messages, as well as high engagement on their content, leading to social influence (Marsden and Friedkin, 1993; Kumar and Mirchandani, 2012). Aral and Walker (2012) find that men are more influential online than women,

and that women influence men more than they influence other women. Kumar et al. (2013) develop an influence score for social media users using their social network data and word-of-mouth flow, and demonstrate its efficacy in promoting Hokey Pokey, an Indian ice cream brand. Their method of identifying influentials and incentivizing them to spread word-of-mouth is among the more sophisticated applications of influencer marketing. Agencies like Klout (now discontinued), and PeerIndex have also provided multidimensional influence metrics with mixed success. However, a more common practice today, as in figure 1, is to identify people with large follower counts on social media, and given their personas, pay them to endorse relevant brands. Recent research by Mallipeddi et al. (2021) provides a handy data-driven framework of influencer selection based on multiple criteria. In the next subsection, we outline how online follower count relates to influence.

2.3 Follower count and influence

A social media user with a larger follower count represents an advertising medium with a larger reach. While the reach of traditional media like TV, radio, hoardings and newspapers can only be approximately estimated, it is reasonable to expect a naive advertiser unaware of fake followers, to place a high amount of trust in seemingly accurate measures like follower counts, number of likes on sponsored posts, retweets and impressions offered by web analytics tools. In a large scale study of 74 million Twitter information cascades encompassing over 1.6 million Twitter users, Bakshy et al. (2011) show a positive link between follower count and online influence. Agent-based studies of targeted new product seeding like Libai et al. (2013) demonstrate that targeting hubs with high numbers of acquaintances speeds up the adoption process, while Yoganarasimhan (2012), in an empirical analysis, suggests that the popularity of a social media message over and above an individual's immediate neighborhood is driven by her follower count. Additionally, influencers with higher follower counts are perceived to be more credible (Jin and Phua, 2014) and likeable (De Veirman et al., 2017).

While follower count is not the only means of determining social influence, it certainly is a popular metric used by digital marketers today to identify influentials

on social networks. The above discussions shed some light on why advertisers pay more to influencers with higher follower counts, in turn generating incentives for influencers to boost their own follower counts via unethical means like buying fake followers.

2.4 The fake follower problem

Influencer marketing today is plagued with a fake follower problem, possibly because endorsers are paid as per their follower count. A quick Google search for "buy instagram followers" for example yields hundreds of results, where anyone with a credit card or Paypal account can buy hundreds of thousands of fake followers instantly (see figure 3). The rates for follows, retweets, likes and comments are different based on the platform in question. Paquet-Clouston et al. (2017) report that click farm clients pay an average of $49 for every 1,000 YouTube followers. The corresponding figures are $34 for Facebook, $16 for Instagram and $15 for Twitter. Average prices for 1,000 likes on these platforms are $50, $20, $14 and $15 respectively.

Insert figure 3 about here

A recent audit by the Institute of Contemporary Music Performances, of the Twitter and Instagram followers of some of the most popular social media celebrities indicates significant numbers of fake followers, mostly above 40% and going up to 57% for Ellen Degeneres, a popular American entertainer (ICMP, 2019). Social media platforms sometimes purge fake followers, though this reduces their user numbers considerably, and often elicits outrage from aggrieved users. Of note is an incident where Amitabh Bachchan, a veteran Indian actor with a Twitter following of over 38 million, berated the platform publicly for reducing his follower count by purging many bot accounts following him (Mathew, 2018). Note however, that the mere presence of fake followers does not always imply that a person has bought them—many fake profiles follow well-known celebrities as part of their online facade. We discuss the implications of this for audits in section 4.4.

3 The model

We adopt a contract theoretic approach inspired by a vast literature on insurance fraud (Crocker and Morgan, 1998; Crocker and Tennyson, 2002; Doherty and Smetters, 2005; Dionne et al., 2009), corporate tax evasion (Crocker and Slemrod, 2005) and earnings inflation (Crocker and Slemrod, 2007; Sun, 2014). These papers use adverse selection and moral hazard models to study fraud in principal-agent problems using contract theory and optimal control theory. Table 1 compares our work with some relevant analogous work in this domain.

Insert table 1 about here

We now describe our model setup in detail. A risk-neutral advertiser (principal) wishes to enter into an agreement with a risk-neutral influencer (agent) to leverage her social media endorsement.[1] The true follower count (type) of an influencer is denoted by n which is distributed over the interval $[n_L, n_H]$ according to the cumulative distribution function $F(n|a)$. $a > 0$ is an action taken by the influencer which can legitimately grow her follower base by investing in analytics, better content, more engagement with followers, etc. (e.g. Hutto et al., 2013; Patel, 2018; Newberry, 2019; Caro and Martínez-de Albéniz, 2020). As the hidden action increases follower count, we make the following assumptions on F, in line with Crocker and Slemrod (2007):

1. $F_a < 0 \ \forall n \in (n_L, n_H)$: first order stochastic dominance—an increase in a shifts F to the right
2. The support $[n_L, n_H]$ of $F(n|a)$, does not vary with a
3. $F_{aa}(n|a) > 0$, $\forall n \in (n_L, n_H)$ which implies $F_a(n_L|a) = F_a(n_H|a) = 0$

The advertiser observes neither the influencer's hidden action a, nor her true follower count n. The advertiser observes the reported follower count $u(n)$ which is possibly inflated where the extent of fraud by the influencer is given by $u(n) - n$.

[1] Table 2 presents all algebraic notation used in the model.

As discussed in section 2.3, influencers with more followers are perceived to be more likable and trustworthy, and have a larger reach, leading to more effective campaigns for advertisers. The advertiser earns a benefit according to the function $R(n)$ with $R' > 0$, for brand endorsement via social media posts.[2]

Insert table 2 about here

The problem is to design an optimal payment mechanism for the influencer. In return for the influencer's endorsement, the advertiser agrees to pay her a fixed component w and a variable component $v(u)$, conditioned on the observable follower count u which is potentially inflated. The influencer incurs a cost $h(a)$ for her action a with $h(a)$ satisfying the following features:

1. $h(0) = 0$
2. $h'(0) = 0$
3. $h'(a) > 0 \; \forall a > 0$
4. $h''(a) > 0 \; \forall a > 0$

The influencer can fraudulently increase her follower count to $u(n)$ by using click farms, at a cost $c(u(n)-n)$ which depends on the number of fake followers $u(n)-n$. We make the following assumptions on this cost function:

1. $c(0) = 0$ (no faking is costless)
2. $c'(0) = 0$ (no faking incurs minimum cost)
3. $c'(z) > 0 \; \forall z > 0$ (faking cost increases with the level of faking)
4. $c'' > 0 \; \forall z \geq 0$ (faking cost is convex)

The last assumption of convex faking cost is worth discussing in detail. Click farms usually provide linear cost structures, where the per-follower cost is constant, or feature bulk discounts. The click farm can do so because it has neither

[2]It suffices mathematically to assume that $R(n)$ is increasing. We acknowledge that the actual measurement of $R(n)$ is complicated (Gordon et al., 2020).

a reputation to maintain, nor significant marginal costs of producing more bot accounts. However, it is important to note that an influencer (whose business depends on her reputation) cannot just get away with buying fake followers. To pull off the fraud convincingly, she must also pay for recurrent engagement such as fake replies, fake comments and fake likes on her social media posts, all of which open her up to increased scrutiny as the magnitude of fraud increases. Unlike the purchase of fake followers which may be a one-time expense, buying fake online engagement from click farms is a recurrent expense. Additionally, social media platforms regularly purge fake followers, and the deception needs to be maintained with regular monitoring and maintenance of the "follower" base. Keeping these considerations in mind, we assume that the cost of deception is the sum of a linear cost of buying fakes and a convex cost of hiding them. Without loss of generality, this leads to a composite cost function $c(u(n) - n)$ which is convex in the general contract theoretic tradition of falsification with costly state verification as outlined in Lacker and Weinberg (1989). In short, our convex cost assumption reflects the aphorism, "to hide a lie, a thousand lies are needed."

Unlike the hidden action which is legitimate, inflating one's follower count is a fraudulent act that can backfire on the influencer if she is exposed. While regulations for influencer marketing are still works in progress, it is reasonable to expect two components to any retributive action when a fraud is uncovered: (i) restitution to the cheated principal analogous to similar penalties levied by the Federal Trade Commission on John Beck's Amazing Products (FTC, 2012) for misleading infomercials, on Uber for driver earnings inflation (FTC, 2017), and on Facebook for video viewership inflation (Morris, 2019); and (ii) extra fines collected by the regulator, plus further reputation costs incurred by the agent that are *not* passed on as restitution to the principal. One of our goals is to explore the policy implications of various penalty structures on the influencer's potentially fraudulent behavior.

The penalty incurred is usually commensurate with the scale of fraud perpetrated. We assume that a non-strategic third party audit by a regulator such as the Federal Trade Commission can detect this fraud with an exogenous proba-

bility γ. If this happens, the influencer's cost gets escalated by a factor $\delta > 1$ to $\delta c(u(n)-n)$. We note that out of this, the influencer has already spent $c(u(n)-n)$ in her initial faking and cover-up, and a further $(\delta-1)c(u(n)-n)$ is incurred on getting exposed. Out of this, an exogenous fraction $\rho < 1$ is returned to the cheated advertiser, while the remaining is incurred as damages which are not paid to the advertiser, but still incurred by the influencer. Figure 4 illustrates these costs and payments in detail. This structure is similar to Crocker and Slemrod (2005), but with three key differences as stated in table 1: (i) we decompose the probability of detection and severity of associated penalties into two separate terms while Crocker and Slemrod (2005) bundle them together. This approach allows for a more granular policy analysis, (ii) Crocker and Slemrod (2005) do not include restitution to the cheated principal—rather the principal and agent both suffer if the latter is caught, and (iii) Crocker and Slemrod (2005) do not have moral hazard in their model. Our moral hazard component is adapted from Crocker and Slemrod (2007), into which we incorporate the audits of Crocker and Slemrod (2005) leading to the rich set of policy simulations of section 4.

Insert figure 4 about here

The informational structure of our model is depicted by the sequence of events in figure 5. At the beginning, a contract $\{v(u), w\}$ is selected for the influencer in anticipation of the sequence of events. Next, the influencer chooses the hidden action a which is not observable to the advertiser. The influencer chooses this action before she knows her true follower count. The distribution F and how it behaves with a is however common knowledge. In the next stage, the influencer observes her true follower count which is also not observable to the advertiser and hence constitutes hidden information. Next the influencer chooses to report $u(n)$ after optimally choosing a and observing n. This implicitly determines the extent of fraud $u(n)-n$. At the last stage, audits are carried out, the contract $\{v(u), w\}$ is implemented and payoffs are realized.

Insert figure 5 about here

While in practice such contracts would be conditioned on the displayed follower count $u(n)$, we use the revelation principle to look for direct mechanisms instead (Myerson, 1979). This is applicable because the contract is implemented after the influencer possesses private information which leads to the tuple $\{v(n), u(n), w\}$ to be conditioned on private information.

3.1 Payoff functions

The advertiser chooses payments $v(n)$ and w to maximize expected payoff $\int_{n_L}^{n_H} \Pi f dn$, where the payoff function Π is:

$$\begin{aligned}\Pi &= (1-\gamma)[R(n) - v(n) - w)] + \gamma[R(n) - v(n) - w + \rho(\delta - 1)c(u(n) - n))] \\ &= R(n) - v(n) - w + \gamma\rho(\delta - 1)c(u(n) - n) \end{aligned} \tag{1}$$

while the influencer's payoff function is:

$$\begin{aligned}Y &= (1-\gamma)[v(n) + w - h(a) - c(u(n) - n)] + \gamma[v(n) + w - h(a) - \delta c(u(n) - n))] \\ &= v(n) + w - h(a) - (1 - \gamma + \gamma\delta)c(u(n) - n)) \end{aligned} \tag{2}$$

with her expected payoff being $\int_{n_L}^{n_H} Y f dn$.

In this contract, the advertiser (principal) maximizes expected profit, after taking into consideration the influencer's (agent) incentive compatibility, individual rationality and delegation constraints respectively, which we outline below.

3.2 Incentive compatibility

In order for the equilibrium to be incentive compatible, it must be that at the optimal v^*, u^*, w^*, there is no incentive for the influencer to not act according to her own type. This happens only if:

$$Y(v^*(n), u^*(n), w^*) \geq Y(v^*(\bar{n}), u^*(\bar{n}), w^*) \qquad \forall \bar{n} \neq n \in [n_L, n_H] \tag{3}$$

For brevity, we denote the optimal value function $Y(v^*, u^*, w^*) \equiv Y^*$ and note that since $Y^*(\cdot)$ is optimal, its derivative with respect to the arguments v and u must be 0:

$$\left.\frac{\partial Y^*}{\partial v}\right|_{v=v^*} = \left.\frac{\partial Y^*}{\partial u}\right|_{u=u^*} = 0 \tag{4}$$

Using the envelope theorem, by means of the total derivative, we establish the dependence of the optimal value function Y^* on the parameter n by:

$$\frac{dY^*}{dn} = \frac{\partial Y^*}{\partial v^*}\frac{dv^*}{dn} + \frac{\partial Y^*}{\partial u^*}\frac{du^*}{dn} + \frac{\partial Y^*}{\partial n} \cdot 1 \tag{5}$$

This leads to the standard envelope condition:

$$\frac{dY^*}{dn} = \frac{\partial Y^*}{\partial n} = (1 - \gamma + \gamma\delta)c'(u(n) - n) \tag{6}$$

3.3 Individual rationality

Because the contract is devised when the influencer does not yet know her true number of followers n, in order to find participation worthwhile it must be that her *ex ante* payoff is at least zero, which is what she gets from not participating. This leads to the following participation constraint:

$$\int_{n_L}^{n_H} Y(v, u, w) f(n|a) dn \geq 0 \tag{7}$$

3.4 Delegation constraint

Because the influencer selects action a before observing n, she will pick the action that maximizes her payoff:

$$\max_a \int_{n_L}^{n_H} Y(v, u, w) f(n|a) dn \tag{8}$$

Using the fact that $Y_a = -h'(a)$, leads to the following first order condition

which we refer to as the delegation constraint:

$$\frac{d}{da}\int_{n_L}^{n_H} Y(v,u,w)f(n|a)dn = 0 \Rightarrow \int_{n_L}^{n_H}(Yf_a - h'f)dn = 0 \tag{9}$$

The delegation constraint, while necessary, is not sufficient for optimality. For sufficiency, we need the following condition.

Lemma. *$F_{aa}(n|a) > 0$ is sufficient for the action a to maximize the influencer's expected payoff $\int_{n_L}^{n_H} Y(v,u,w)f(n|a)dn$*

Proof. See Appendix A □

3.5 Optimal control problem

The advertiser wishes to maximize its expected profit under the incentive compatibility, individual rationality, and delegation constraints of the influencer. We express it in terms of the following optimal control program:

$$\max_{v,u,a,w}\left(\int_{n_L}^{n_H}\Pi f dn\right) \text{ subject to} \tag{10}$$

$$\frac{dY}{dn} = \frac{\partial Y}{\partial n} \tag{11}$$

$$\int_{n_L}^{n_H} Yf dn \geq 0 \tag{12}$$

$$\int_{n_L}^{n_H}(Yf_a - h'f)dn = 0 \tag{13}$$

The expected profit function (10) under the incentive compatibility constraint (11), the participation constraint (12) and the delegation constraint (13) can be combined into the following Hamiltonian:

$$\mathbb{H} = \Pi f + \lambda(n)Y_n + \mu Yf + \psi(Yf_a - h'f) \tag{14}$$

In the above formulation, $Y(\cdot)$, the influencer's payoff function is the state variable with its equation of motion represented by condition (11). The control

variable is $u(\cdot)$; $\lambda(n)$ is the co-state variable corresponding to the incentive compatibility constraint (11); and μ and ψ are the Lagrangian multipliers associated with the participation constraint (12) and delegation constraint (13) respectively.

3.6 Optimal contract

The proposition below characterizes the necessary conditions for the optimal contract:

Proposition. *The necessary conditions which characterize the optimal contract are as follows:*

$$\frac{c'(u(n)-n)}{c''(u(n)-n)} = -\frac{\psi F_a(n|a)}{f(n|a)} \cdot \frac{(1-\gamma+\gamma\delta)}{[1-\gamma+\gamma\delta-\gamma\rho(\delta-1)]} \tag{15}$$

$$\int_{n_L}^{n_H} Y f(n|a)dn = 0 \tag{16}$$

$$\int_{n_L}^{n_H} Y f_a(n|a)dn = h'(a) \tag{17}$$

$$\int_{n_L}^{n_H} [R-(1-\gamma+\gamma\delta-\rho\gamma(\delta-1))c(u(n)-n)]f_a(n|a)dn$$
$$-h'(a)+\psi\left[\int_{n_L}^{n_H} Y f_{aa}(n|a)dn - h''(a)\right] = 0 \tag{18}$$

$$Y_n = (1-\gamma+\gamma\delta)c'(u(n)-n) \tag{19}$$

Proof. See appendix B □

This proposition yields a key insight via equation (15)—that the lowest and highest type influencers do not fraudulently inflate their follower count, while all other types do. Appendix B.1 presents a detailed analysis of why this is so, and we illustrate in section 4, that the level of faking: $u(n)-n$ displays an inverted U shape as a function of type n. Intuitively, this means that faking levels are low for low types *and* for high types, but high for intermediate types. Anecdotally too, investigations like Confessore et al. (2018) uncover that buyers of fake followers

tend to be people who are neither completely unknown, nor A-list celebrities but those who are moderately famous management gurus, TV personalities, fashion models and specialist influencers as click farm clients.

Corollary. *The optimal payment to the influencer is given by:*

$$v(n) = (1-\gamma+\gamma\delta)\left[c(u(n)-n)+\int_{n_L}^{n} c'(u(t)-t)dt\right] \tag{20}$$

$$w = h(a)-(1-\gamma+\gamma\delta)\int_{n_L}^{n_H}\left[\int_{n_L}^{n} c'(u(t)-t)dt\right] f(n|a)dn \tag{21}$$

Proof. See appendix B.2 □

3.6.1 Implementability and sufficiency

Implementability requires that:

$$\frac{\partial}{\partial n}\left(\frac{Y_u}{Y_v}\right)\cdot\frac{du}{dn}\geq 0 \tag{22}$$

$$\frac{\partial}{\partial n}\left(\frac{-(1-\gamma+\gamma\delta)c'(u(n)-n)}{1}\right)\frac{du}{dn}\geq 0 \tag{23}$$

The first term in (22) is the Spence-Mirrlees single-crossing condition. Based on the inequalities above, $u'>0$ is required for implementability which implies that the displayed follower count must increase with the true number of followers of the influencer.

In particular, if the cost function for hiring fakes is assumed to be quadratic, (15) leads to:

$$u = n-\frac{\psi F_a(n|a)}{f(n|a)}\cdot\left(\frac{1-\gamma+\gamma\delta}{1-\gamma+\gamma\delta-\gamma\rho(\delta-1)}\right) \tag{24}$$

$$\frac{du}{dn} = 1-\left(\frac{1-\gamma+\gamma\delta}{1-\gamma+\gamma\delta-\gamma\rho(\delta-1)}\right)\frac{d}{dn}\left(\psi\frac{F_a(n|a)}{f(n|a)}\right) \tag{25}$$

Clearly, for $u'>0$, it is sufficient that:

$$\frac{d}{dn}\left(\frac{\psi F_a(n|a)}{f(n|a)}\right)<1-\frac{\gamma\rho(\delta-1)}{(1-\gamma+\gamma\delta)}$$

4 Policy simulations

In our main model, we have three exogenous parameters γ, δ and ρ which characterize the accuracy, severity and potential restitutive compensation to the advertiser respectively. This specification leads to endogenous outcomes $u(n), v(n), a$ and ψ which constitute the payoff functions Π and Y. In an ideal scenario, we would like to present comparative statics of the form $du/d\gamma, du/d\delta$, etc., to ascertain the relative effects of changes in exogenous parameters on faking levels and effort (i.e. hidden action). Such an exercise is ideally the best and most general way to evaluate policy implications for an analytical model such as ours.

Unfortunately, the complexity of the equations characterizing the optimal contract renders such an analytical exercise infeasible. As equation (15) indicates, the faking level $u(n) - n$ and the variable payment $v(n)$ depend on the endogenous Lagrangian multiplier ψ and the optimal action a, apart from the exogenous parameters α, γ, δ and ρ. Further, the endogenous outcomes ψ and a themselves are simultaneously determined via equations (17) and (18) in the proposition. All of this renders analytical attempts at comparative statics difficult and unintuitive.

In lieu of analytical comparisons, we present a detailed numerical simulation analysis (emulating Mesak et al., 2016; Ha and Sibert, 1997; Judd, 1998). Here, we make reasonable assumptions about the functional forms of $R(n), c(u(n)-n), h(a)$ and $F(n|a)$. We then take multiple combinations of the exogenous parameter values α, γ, δ and ρ in a full factorial experimental design, and numercially compute the endogenous outcomes a, ψ, and in turn use them to calculate the display function $u(n)$ and hence faking level $u(n) - n$. As this is a randomized experiment with a full factorial design, we then run ordinary least squares regressions on these dependent variables, using the exogenous parameters γ, δ and ρ to evaluate the relative impacts of each on the faking level and hidden action.

4.1 Numerical setup

We design our numerical sensitivity analysis on the lines of the brief illustration in Crocker and Slemrod (2007). We assume the following functions:

$$R(n) = n \tag{26}$$
$$c(u(n)-n) = \frac{(u(n)-n)^2}{2} \tag{27}$$
$$h(a) = \frac{a^3}{3} \tag{28}$$
$$F(n|a) = n^a;\ n \in [0,1] \tag{29}$$

which further leads to:

$$F_a(n|a) = n^a \log n \tag{30}$$
$$F_{aa}(n|a) = n^a (\log n)^2 \tag{31}$$
$$f(n|a) = a n^{a-1} \tag{32}$$
$$f_a(n|a) = n^{a-1}(1 + a \log n) \tag{33}$$
$$f_{aa}(n|a) = n^{a-1} \log n (2 + a \log n) \tag{34}$$

This formulation allows us to now solve for ψ and a simultaneously from equations (17) and (18) which yield:[3]

$$\left[\frac{1}{(1+a)^2} - \frac{2\psi^2(1-a)}{a^2(2+a)^4} \cdot \frac{(1-\gamma+\gamma\delta)^2}{\{1-\gamma+\gamma\delta-\rho\gamma(\delta-1)\}}\right] - a^2 + \psi\left[\frac{-6\psi(1-\gamma+\gamma\delta)^2}{a(2+a)^4\{1-\gamma+\gamma\delta-\gamma\rho(\delta-1)\}} - 2a\right] = 0 \tag{35}$$
$$\frac{2\psi}{a(2+a)^3} \cdot \frac{(1-\gamma+\gamma\delta)^2}{\{1-\gamma+\gamma\delta-\gamma\rho(\delta-1)\}} - a^2 = 0 \tag{36}$$

[3]Given the tedious nature of the algebra and integrals involved, we suggest using Wolfram Mathematica to derive (35)–(38).

As is evident, there is no closed form solution to a, ψ from the above, and these must be solved numerically. Equation (15) yields the display function $u(n)$ as:

$$u(n) = n - \frac{\psi}{a} n \log n \cdot \frac{(1-\gamma+\gamma\delta)}{[1-\gamma+\gamma\delta-\gamma\rho(\delta-1)]} \tag{37}$$

From equation (20) we calculate $v(n)$ as:

$$v(n) = (1-\gamma+\gamma\delta)\left[\frac{(1-\gamma+\gamma\delta)^2\psi^2 n^2(\log n)^2}{2a^2[1-\gamma+\gamma\delta-\gamma\rho(\delta-1)]^2} + \frac{(1-\gamma+\gamma\delta)\psi n^2(1-2\log n)}{4a(1-\gamma+\gamma\delta)}\right] \tag{38}$$

4.2 Simulation design

We now outline the design of our simulations based on the setup described in section 4.1. We vary the parameters γ and ρ over 11 levels each between 0 and 1 with increments of 0.1, and the parameter δ over 9 levels between 1 and 5 with increments of 0.5 in a full factorial experimental design yielding 1,089 combinations of γ, δ, ρ. For each of these, we solve equations (35) and (36) simultaneously using a numeric solver. We then vary n at 11 levels 0 and 1 in increments of 0.1, and compute $u(n)$ for each using equation (15). Finally, we use these values to calculate $v(n)$ from equation (20). The above procedure yields us a synthetic data matrix with 11,979 observations. Note that while outcomes a, ψ are computed independent of n, the faking level $u(n) - n$ depends on n.

4.3 Policy implications

Table 3 presents the results of ordinary least squares regressions that ascertain the effects of exogenous parameters on faking levels and hidden action. Note that while faking levels are type dependent, the action is not.

Insert table 3 about here

For each of the dependent variables $u(n) - n$ and a, we present two regressions, one with only main effects of the exogenous parameters, and one with theoretically

relevant interaction effects. As the proposition predicts a non-monotonic relationship between the faking level and type (see appendix B.1 for a technical discussion on this), the first two models incorporate n and $\log n$. Figure 6 illustrates these results as well, with a few selected curves depicting the faking level's sensitivity to exogenous policy parameters.

Insert figure 6 about here

The regressions demonstrate our key results. First, the audit accuracy γ has a significant negative main effect on faking levels and a significant positive main effect on the hidden action, only when interaction terms are *not* considered (models 1 and 3). With the relevant interactions, the main effect of γ on the faking level vanishes (model 2). This is as per intuition—accurate audits only work as fraud deterrents when they are accompanied by a harsh penalty on the offender. Note that δ is capped because of the advertiser's implicit participation constraint.

Our second observation is counter-intuitive and a key finding of this paper. While the restitution factor ρ has a significant and positive main effect on the hidden action a (model 3), it also encourages more faking, as indicated by the positive and significant main effect in model 1. This vanishes when interaction terms are included (models 2 and 4 for faking level and hidden action respectively). However, the interaction terms on $\gamma\rho$ and $\gamma\delta\rho$ both have positive significant coefficients in model 2 (for faking level) and only for $\gamma\delta\rho$ in model 4 (hidden action). Figure 6 (bottom) illustrates this.

We interpret the above result thus: ρ is the fraction of the escalated costs $(\delta-1)c(u(n)-n)$ that the influencer incurs if caught with probability γ. It may be tempting from an intuitive angle, or even from the point of view of fairness,[4] for a third party regulator like the Federal Trade Commission to direct the influencer to pay a large part of its fine as compensation to the cheated advertiser. However, the deterrence effects of having such a clause in an *ex ante* contract are questionable. In fact such a policy backfires, causing the agent to fake her follower count even

[4]In an analogous scenario, when Uber was fined $20 million for inflating driver earnings in its promotions, the Federal Trade Commission's Director of Consumer Protection remarked, "This settlement will put millions of dollars back in Uber drivers' pockets." (FTC, 2017)

more. This is because with expected restitution as indicated in equation (1), the principal is willing to pay more in terms of $v(n) + w$, which in turn generates both legitimate incentives for the influencer on her hidden action, as well as perverse incentives for her to fraudulently inflate her follower count and cover this up.

The policy implications of this observation can be summed up in one line: *as a regulator, collect a reasonably high fine commensurate with the degree of fraud you have incontrovertible evidence for, but do not disburse any fraction back to the cheated advertiser.* This indicates designing an *ex ante* contract with a value of δ which is high enough, but ideally no restitution to the advertiser if fraud is caught.

4.4 Interpreting audits

While we note that our policy simulations cover the whole range of $\gamma \in [0, 1]$, in reality we expect this accuracy to be low. As mentioned in section 2.4, audits such as those in Mathew (2018), ICMP (2019) and Stanley (2019) uncover the existence of several fake followers of many well-known social media handles. However these audits do not provide any evidence that the owners of these popular handles have themselves bought fake followers. In fact, it is common practice for click farms to train both their human employees and automated bots to follow popular social media handles. On the other hand, thorough investigations like that by The New York Times (Confessore et al., 2018) are rare, given the huge hurdles required to validate damaging claims against subjects of such investigations.

The Confessore et al. (2018) exposé is unique because it manages to get actual proof of purchase of fake followers, primarily by moderately well-known personalities who are neither completely obscure, nor A-list celebrities like Justin Bieber, Kim Kardashian or Ellen Degeneres (named in ICMP, 2019), official Twitter handles of Liverpool, Manchester United and Arsenal (named in Stanley, 2019) or Amitabh Bachchan (named in Mathew, 2018). Confessore et al.'s evidence consists of credit card details of these influencers (or their social media managers) found in a well-known click farm's database—a true smoking gun that meets the high evidence standards required to establish the actual purchase of fake followers and fake engagement (likes, shares, comments). In another prominent case, the

police of Mumbai, India claim to have evidence for some B-list celebrities who have allegedly bought fake followers (Khan and Singh, 2020).

Such evidence is almost impossible for even platforms like Twitter and Instagram to provide without external investigation—while their internal databases and machine learning algorithms have sophisticated fake follower detection techniques that can detect both suspicious bot behavior and inexplicable spikes in follower count, unambiguously proving that these fake followers were paid for by a the owner of a given social media handle is non-trivial. Managers and regulators should keep this mind while commissioning and interpreting audits.

5 Concluding remarks

Our work addresses the concerns of a growing number of managers who now realise the scale of the fake follower problem in influencer marketing. Given that the click farms are getting more sophisticated than ever before, with bot behavior being camouflaged with advanced machine learning algorithms modeled on real online user behavior (Silverman, 2018), we expect the fake follower problem to persist, with click farms developing more sophisticated methods to avoid fake follower detection, both by human investigators and machine learning algorithms. Our work therefore breaks from the machine learning tradition of identifying fake followers, and instead studies the fake follower problem from the fundamental incentives associated with influencer marketing.

Our model maps the influencer's faking level with her true follower count, demonstrating an inverted U-shaped relationship with no faking at the highest and lowest types, and more faking at intermediate types instead. In simple words, we find that the very obscure "nano-influencers" and high-profile celebrities are unlikely to fake much. Rather, it is the moderately famous influencers who are more likely to display inflated follower counts by buying fake followers.

Our model also derives optimal payment schemes, demonstrating that accurate audits need to be coupled by commensurate penalties to deter influencer fraud. However, such penalties are better collected by third party industry regulators

than given back as restitution to cheated advertisers.

Methodologically, we use contract theory coupled with optimal control. Using analogous scenarios from the insurance and earnings management literature, we model the advertiser-influencer game in a principal-agent setting, simultaneously including moral hazard and audits together.

Our results are consistent with observed trends in influencer marketing, illuminating a dark corner of this unfortunately murky world.

6 Appendix

A Proof of Lemma

Proof. The delegation constraint requires that the action a chosen by the influencer must optimize her expected payoff:

$$\max_a \int_{n_L}^{n_H} Y f dn \tag{A1}$$

This requires two conditions, first that the derivative of the expected payoff with respect to a be 0:

$$\frac{d}{da} \int_{n_L}^{n_H} Y f dn = 0 \tag{A2}$$

and the second that the second derivative at the optimal a be negative:

$$\frac{d^2}{da^2} \int_{n_L}^{n_H} Y f dn < 0 \tag{A3}$$

$$\frac{d}{da} \int_{n_L}^{n_H} (Y f_a - h' f)\, dn < 0 \tag{A4}$$

$$\int_{n_L}^{n_H} (Y f_{aa} - h' f_a - h'' f - h' f_a)\, dn < 0 \tag{A5}$$

$$\int_{n_L}^{n_H} Y f_{aa} dn - h'' < 0 \tag{A6}$$

where the last line follows from $\int_{n_L}^{n_H} f_a dn = F_a(n_H|a) = F_a(n_L|a) = 0 - 0 = 0$. Substituting the definition of Y we get:

$$\int_{n_L}^{n_H} [v(n) + w - h(a) - (1 - \gamma + \gamma\delta)c(u - n)] f_{aa} dn - h'' < 0 \quad \text{(A7)}$$

$$(w - h(a)) \int_{n_L}^{n_H} f_{aa} dn + \int_{n_L}^{n_H} [v(n) - (1 - \gamma + \gamma\delta)c(u - n)] f_{aa} dn - h'' < 0 \quad \text{(A8)}$$

Noting that $\int_{n_L}^{n_H} f_{aa} dn = F_{aa}(n_H|a) - F_{aa}(n_L|a) = 0 - 0 = 0$ and using integration by parts for the second term, we get:

$$\begin{aligned} &[v(n) - (1 - \gamma + \gamma\delta)c(u(n) - n)] F_{aa}(n|a) \Big|_{n_L}^{n_H} \\ &- \int_{n_L}^{n_H} [v'(n) - (1 - \gamma + \gamma\delta)(c' \cdot (u' - 1))] F_{aa} dn - h'' \quad < 0 \end{aligned} \quad \text{(A9)}$$

Again the first term vanishes since $F_{aa}(n_H|a) - F_{aa}(n_L|a) = 0 - 0 = 0$. Further, noting that

$$\frac{dY}{dn} = v' - (1 - \gamma + \gamma\delta)c' \cdot (u' - 1) = \frac{\partial Y}{\partial n} = (1 - \gamma + \gamma\delta)c' \quad \text{(A10)}$$

Substituting the above result, we obtain:

$$-(1 - \gamma + \gamma\delta) \left(\int_{n_L}^{n_H} c' F_{aa} dn \right) - h'' < 0 \quad \text{(A11)}$$

Clearly, the above inequality holds if $F_{aa} > 0$. □

B Proof of Proposition

Proof. The Hamiltonian for the optimal control program is

$$\mathbb{H} = \Pi f + \lambda(n) Y_n + \mu Y f + \psi(Y f_a - h' f) \quad \text{(B1)}$$

Using the definition of the payoff function Π in equation (1); and the value of Y_n from equation (6), we get

$$\mathbb{H} = (R-h-[1-\gamma+\gamma\delta-\gamma\rho(\delta-1)]c-Y)f+\lambda(n)(1-\gamma+\gamma\delta)c'+\mu Yf+\psi(Yf_a-h'f) \tag{B2}$$

The Pontryagin first order conditions are:

1. Optimality condition:

$$\max_u \mathbb{H} \qquad \forall n \in [n_L, n_H] \equiv \frac{\partial \mathbb{H}}{\partial u} = 0$$

2. Equation of motion for state:

$$\frac{dY}{dn} = \frac{\partial \mathbb{H}}{\partial \lambda} = Y_n$$

3. Equation of motion for costate:

$$\frac{d\lambda}{dn} = -\frac{\partial \mathbb{H}}{\partial Y}$$

4. Transversality condition for state:

$$\lambda(n_L) = 0$$

Using the optimality condition:

$$\frac{\partial \mathbb{H}}{\partial u} = 0 \Rightarrow -\left[1-\gamma+\gamma\delta-\gamma\rho(\delta-1)\right]c'f+\lambda(1-\gamma+\gamma\delta)c'' = 0 \tag{B3}$$

$$\frac{c'}{c''} = \frac{\lambda}{f}\left[\frac{1-\gamma+\gamma\delta}{1-\gamma+\gamma\delta-\gamma\rho(\delta-1)}\right] \tag{B4}$$

Using the equation of motion for costate:

$$\frac{d\lambda}{dn} = -\frac{\partial \mathbb{H}}{\partial Y} = -\left[-f+\mu f+\psi f_a\right] \tag{B5}$$

$$= (1-\mu)f-\psi f_a \tag{B6}$$

Moreover, since a and w are independent of n, they must constitute the argmax of the following objective function:

$$\max_{a,w} \int_{n_L}^{n_H} [\Pi f + \mu Y f + \psi(Y f_a - h' f)]\, dn \tag{B7}$$

Thus the derivative of the objective function above with respect to w must be 0. Further, using the fact that $\Pi_w = -1$ and $Y_w = 1$, we get

$$\frac{d}{dw} \int_{n_L}^{n_H} [\Pi f + \mu Y f + \psi(Y f_a - h' f)]\, dn = 0 \tag{B8}$$

$$\int_{n_L}^{n_H} [\Pi_w f + \mu Y_w f + \psi Y_w f_a]\, dn = 0 \tag{B9}$$

$$(-1 + \mu) \int_{n_L}^{n_H} f dn + \psi \int_{n_L}^{n_H} f_a dn = 0 \tag{B10}$$

We note that $\int_{n_L}^{n_H} f dn = 1$ and that $\int_{n_L}^{n_H} f_a dn = F_a(n_H|a) - F_a(n_L|a) = 0 - 0 = 0$, which leads to

$$-1 + \mu = 0 \Rightarrow \mu = 1 \tag{B11}$$

Substituting this value of μ in (B6) and using the transversality condition $\lambda(n_L) = 0$ and the fact that $F_a(n_L|a) = 0$ we get the expression for λ:

$$\frac{d\lambda}{dn} = -\psi f_a \Rightarrow \lambda(n) - \lambda(n_L) = -\psi \int_{n_L}^{n} f_a dn \tag{B12}$$

$$\lambda(n) = -\psi F_a(n|a) \tag{B13}$$

Finally, we substitute the above expression for λ to derive (15):

$$\frac{c'(u(n) - n)}{c''(u(n) - n)} = -\frac{\psi F_a(n|a)}{f(n|a)} \left[\frac{1 - \gamma + \gamma\delta}{1 - \gamma + \gamma\delta - \gamma\rho(\delta - 1)}\right] \tag{B14}$$

We can make some quick observations regarding the display function $u(n)$ from the equation above. First, at $n = n_L$, since $F_a(n_L|a) = 0$, the RHS equals 0, which in turn implies that the LHS is 0 leading to the implication that $c'(u(n_L) - n_L) = 0 \Rightarrow u(n_L) = n_L$. Similary, $u(n_H) = n_H$. Second, for all types $n_L < n < n_H$,

the RHS is strictly positive, since $F_a(n|a) < 0$ and the tightness of the delegation constraint necessitates that $\psi > 0$. This in turn implies that the LHS be strictly positive, leading to $u(n) > n$.

Returning to the argument that posits that the action undertaken to increase the follower count is independent of n, we require that the derivative of the objective function with respect to a be 0 and using $\mu = 1$ from (B11), we get:

$$\frac{d}{da}\int_{n_L}^{n_H} \left[(\Pi + Y)\, f + \psi(Y f_a - h' f)\right] dn = 0$$

Using the definition of Π and Y and substituting, we get:

$$\frac{d}{da}\int_{n_L}^{n_H} \left(R - \left[1 - \gamma + \gamma\delta - \gamma\rho(\delta - 1)\right] c - h\right) f dn + \psi \frac{d}{da}\int_{n_L}^{n_H} \left(Y f_a - h' f\right) dn = 0$$

We note that $Y_a = -h'(a)$ and expand the above expression:

$$\int_{n_L}^{n_H} \left(R - [1 - \gamma + \gamma\delta - \gamma\rho(\delta - 1)]c\right) f_a dn - h' \int_{n_L}^{n_H} f dn - h \int_{n_L}^{n_H} f_a dn + \psi\left(\int_{n_L}^{n_H} Y f_{aa} dn - h' \int_{n_L}^{n_H} f_a dn - h' \int_{n_L}^{n_H} f_a dn - h'' \int_{n_L}^{n_H} f dn\right) = 0$$

Noting that $\int_{n_L}^{n_H} f dn = 1$ and that $\int_{n_L}^{n_H} f_a dn = 0$, we simplify the above to derive (18)

$$\int_{n_L}^{n_H} \left[R - (1 - \gamma + \gamma\delta - \gamma\rho(\delta - 1))c\right] f_a dn - h' + \psi\left(\int_{n_L}^{n_H} Y f_{aa} dn - h''\right) = 0$$

Finally, partially differentiating Y with respect to n, we obtain equation (19)

$$\begin{aligned} Y &= v(n) + w - h(a) - (1 - \gamma + \gamma\delta)c(u(n) - n) \\ Y_n &= (1 - \gamma + \gamma\delta)c'(u(n) - n) \end{aligned} \tag{B15}$$

□

B.1 Implications of the proposition

This proposition yields a key insight via equation (15)—that the lowest and highest type influencers do not fraudulently inflate their follower count. Appendix presents a detailed technical analysis of this. This is because $F_a(n_L|a) = F_a(n_H|a) = 0$ rendering the right hand side zero. By implication, the left hand side is also zero, leading to $c'(u(n_L) - n_L) = c'(u(n_H) - n_H) = 0$ which in turn implies that $u(n_L) = n_L$ and $u(n_H) = n_H$ given the assumptions on the cost function. For every other type in (n_L, n_H), the right hand side of equation (15) is positive, because $F_a(n) < 0$ by assumption; and $\psi > 0$ since the participation constraint is tight [equation (16)]. This is a key feature of our model which states that the faking level $u(n) - n$ is a non-monotonic function of n. Moreover, as we show in section 4, the level of faking assumes an inverted U shape.

B.2 Proof of Corollary 3.6

Proof. In order to derive the variable payment schedule we use incentive compatibility (11)

$$\frac{dY}{dt} = \frac{\partial Y}{\partial t} \tag{B16}$$

$$\int_{n_L}^{n} \frac{dY}{dt} dt = \int_{n_L}^{n} \frac{\partial Y}{\partial t} dt \tag{B17}$$

$$Y(n) - Y(n_L) = \int_{n_L}^{n} \frac{\partial}{\partial t}(v(t) + w - h(a) - (1 - \gamma + \gamma\delta)c(u(t) - t))dt \tag{B18}$$

Using the fact that $c(u(n_L) - n_L) = 0$ and simplifying

$$v(n) - v(n_L) - (1 - \gamma + \gamma\delta)c(u(n) - n) = (1 - \gamma + \gamma\delta) \int_{n_L}^{n} c'(u(t) - t)dt \tag{B19}$$

Normalizing $v(n_L) = 0$ without loss of generality, we obtain:

$$v(n) = (1 - \gamma + \gamma\delta) \left[c(u(n) - n) + \int_{n_L}^{n} c'(u(t) - t)dt \right] \tag{B20}$$

Now we derive the expression for the fixed payment w. Using the fact that the expected ex-ante payoff for the influencer is 0, which follows since the corresponding multiplier $\mu = 1$ (B11), we get:

$$\int_{n_L}^{n_H} [v + w - h - (1 - \gamma + \gamma\delta)c] \, f dn = 0 \tag{B21}$$

$$(w - h) \int_{n_L}^{n_H} f(n|a) dn + \int_{n_L}^{n_H} [v - (1 - \gamma + \gamma\delta)c] f dn = 0 \tag{B22}$$

Using the value of the variable payment v from above and expanding, we get

$$w = h - \int_{n_L}^{n_H} \left[(1 - \gamma + \gamma\delta)c - (1 - \gamma + \gamma\delta) \left(c + \int_{n_L}^{n} c' dt \right) \right] f(n|a) dn \tag{B23}$$

$$w = h(a) - (1 - \gamma + \gamma\delta) \int_{n_L}^{n_H} \left(\int_{n_L}^{n} c'(u(t) - t) dt \right) f(n|a) dn \tag{B24}$$

□

References

Aral, S. and D. Walker (2012). Identifying influential and susceptible members of social networks. *Science 337*(6092), 337–341.

Bakshy, E., J. M. Hofman, W. A. Mason, and D. J. Watts (2011). Everyone's an influencer: quantifying influence on Twitter. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pp. 65–74. ACM.

Berger, J. (2016). *Contagious: Why things catch on*. Simon and Schuster.

Caro, F. and V. Martínez-de Albéniz (2020). Managing online content to build a follower base: Model and applications. *INFORMS Journal on Optimization 2*(1), 57–77.

Confessore, N., G. Dance, R. Harris, and M. Hansen (2018). The Follower Factory. *New York Times*.

Cresci, S., R. Di Pietro, M. Petrocchi, A. Spognardi, and M. Tesconi (2015). Fame for sale: Efficient detection of fake Twitter followers. *Decision Support Systems 80*, 56–71.

Crocker, K. J. and J. Morgan (1998). Is honesty the best policy? Curtailing insurance fraud through optimal incentive contracts. *Journal of Political Economy 106*(2), 355–375.

Crocker, K. J. and J. Slemrod (2005). Corporate tax evasion with agency costs. *Journal of Public Economics 89*(9-10), 1593–1610.

Crocker, K. J. and J. Slemrod (2007). The economics of earnings manipulation and managerial compensation. *RAND Journal of Economics 38*(3), 698–713.

Crocker, K. J. and S. Tennyson (2002). Insurance fraud and optimal claims settlement strategies. *Journal of Law and Economics 45*(2), 469–507.

De Veirman, M., V. Cauberghe, and L. Hudders (2017). Marketing through Instagram influencers: the impact of number of followers and product divergence on brand attitude. *International Journal of Advertising 36*(5), 798–828.

Dionne, G., F. Giuliano, and P. Picard (2009). Optimal auditing with scoring: Theory and application to insurance fraud. *Management Science 55*(1), 58–70.

Doherty, N. and K. Smetters (2005). Moral hazard in reinsurance markets. *Journal of Risk and Insurance 72*(3), 375–391.

eMarketer (2017). Attitudes Toward Influencer Marketing Among US Agency and Brand Marketers, Nov 2017 (% of respondents). Technical report.

FTC (2012). At FTC's request, US court hands down record $478 million judgment against marketers of massive get-rich-quick infomercial scams. *Federal Trade Commission*.

FTC (2017). Uber agrees to pay $20 million to settle FTC charges that it recruited prospective drivers with exaggerated earnings claims. *Federal Trade Commission*.

FTC (2019). Disclosures 101 for social media influencers. *Federal Trade Commission*.

Gordon, B. R., K. Jerath, Z. Katona, S. Narayanan, J. Shin, and K. C. Wilbur (2020). Inefficiencies in digital advertising markets. *Journal of Marketing 85*(1), 7–25.

Ha, J. and A. Sibert (1997). Strategic capital taxation in large open economies with mobile capital. *International Tax and Public Finance 4*, 243–262.

Hutto, C. J., S. Yardi, and E. Gilbert (2013). A longitudinal study of follow predictors on Twitter. In *Proceedings of the sigchi conference on human factors in computing systems*, pp. 821–830.

ICMP (2019). Who has the most fake followers?

IRI (2018). BzzAgent and IRI Find Everyday Influencer Marketing Programs Drive the Highest Return on Ad Spend. Technical report.

Jin, S.-A. A. and J. Phua (2014). Following celebrities' tweets about brands: The impact of twitter-based electronic word-of-mouth on consumers' source credibility perception, buying intention, and social identification with celebrities. *Journal of Advertising 43*(2), 181–195.

Joseph, S. (2018). 'we don't pay influencers on reach': How Kellogg's is combating influencer fraud. *Digiday*.

Judd, K. (1998). *Numerical Methods in Economics*. MIT press.

Khalil, A., H. Hajjdiab, and N. Al-Qirim (2017). Detecting fake followers in Twitter: A machine learning approach. *International Journal of Machine Learning and Computing 7*(6), 198–202.

Khan, F. and V. Singh (2020). Mumbai police probe reveals 10 celebrities paid dollars for fake tweets, Facebook likes. *Mid-Day*.

Kumar, V., V. Bhaskaran, R. Mirchandani, and M. Shah (2013). Practice prize winner—creating a measurable social media marketing strategy: increasing the value and ROI of intangibles and tangibles for hokey pokey. *Marketing Science 32*(2), 194–212.

Kumar, V. and R. Mirchandani (2012). Increasing the ROI of social media marketing. *MIT Sloan Management Review 54*(1), 55.

Lacker, J. M. and J. A. Weinberg (1989). Optimal contracts under costly state falsification. *Journal of Political Economy 97*(6), 1345–1363.

Libai, B., E. Muller, and R. Peres (2013). Decomposing the Value of Word-of-Mouth Seeding Programs: Acceleration Versus Expansion. *Journal of Marketing Research 50*(2), 161–176.

Linqia (2018). The state of influencer marketing. Technical report, Linqia.

Mallipeddi, R., S. Kumar, C. Sriskandarajah, and Y. Zhu (2021). A framework for analyzing influencer marketing in social networks: selection and scheduling of influencers. *Management Science*.

Marsden, P. V. and N. E. Friedkin (1993). Network studies of social influence. *Sociological Methods & Research 22*(1), 127–151.

Mathew, S. (2018). Oh my! Amitabh Bachchan is so touchy about his Twitter followers. *The Quint*.

Mesak, H., A. Bari, and R. Blackstock (2016). On the robustness and strategic implications of a parsimonious advertising – inventory competitive model with extensions to pricing competition. *International Journal of Production Economics 180*, 38–47.

Morris, C. (2019). Facebook will pay $40 million to settle advertiser lawsuits claiming it inflated video views by up to 900%. *Fortune*.

Myerson, R. B. (1979). Incentive compatibility and the bargaining problem. *Econometrica 47*, 61–73.

Neff, J. (2018). Study of influencer spenders finds big names, lots of fake followers. *AdAge India*.

Newberry, C. (2019). How to get free Instagram followers: 27 tips that actually work. *Hootsuite Blog*.

Paquet-Clouston, M., O. Bilodeau, and D. Décary-Hétu (2017). Can We Trust Social Media Data?: Social Network Manipulation by an IoT Botnet. In *Proceedings of the 8th International Conference on Social Media & Society*, #SMSociety17, New York, NY, USA, pp. 15:1–15:9. ACM.

Patel, N. (2018). How to get more social media followers (without creating content). *Neil Patel's Blog*.

Pathak, S. (2017). Cheatsheet: What you need to know about influencer fraud. *Digiday*.

Silverman, C. (2018). Apps Installed On Millions Of Android Phones Tracked User Behavior To Execute A Multimillion-Dollar Ad Fraud Scheme. *BuzzFeed*.

Stanley, A. (2019). REVEALED: The Premier League clubs with the most 'fake followers' – from lowest to highest. *talkSPORT*.

Stewart, R. (2018). Unilever's Keith Weed calls for 'urgent action' to tackle influencer fraud. *The Drum*.

Stewart, R. (2019). Estée Lauder now spends a huge portion of its marketing budget on influencers. *The Drum*.

Sun, B. (2014). Executive compensation and earnings management under moral hazard. *Journal of Economic Dynamics and Control 41*, 276–290.

The Economist (2016). Celebrities' endorsement earnings on social media - Daily chart.

Yoganarasimhan, H. (2012, March). Impact of social network structure on content propagation: A study using YouTube data. *Quantitative Marketing and Economics 10*(1), 111–150.

Zhang, Y. and J. Lu (2016). Discover millions of fake followers in Weibo. *Social Network Analysis and Mining 6*(1), 16.

Table 1: Comparison with extant work using contract theory coupled with optimal control. To the best of our knowledge, this method has not been used in the marketing literature thus far

	Crocker and Morgan (1998)	Crocker and Slemrod (2005)	Crocker and Slemrod (2007)	*This work*
Phenomenon	Sharecropper and insurance fraud	Corporate tax evasion	Company earnings misreporting	*Influencer marketing with fake followers*
Principal	Farm owner/ insurer	Shareholders	Company owner	*Advertiser*
Agent	Sharecropper/ insuree	CEO	Manager	*Influencer*
Moral hazard	No	No	Yes	*Yes*
Audit accuracy	No	No	No	*Yes*
Audit severity	No	Yes	No	*Yes*
Audit consequence	N.A.	Principal and agent both suffer if agent is caught	N.A.	*Agent compensates principal if caught*
Policy analysis	N.A.	Analytical	Numerical (illustrative)	*Numerical (extensive)*

Table 2: Summary of notation used in our model

Term	*Description*
n	True follower count of the influencer (influencer type)
$f(n), F(n)$	Probability density and cumulative distribution function of n
a	Action undertaken by the influencer
h	Disutility incurred in taking action
$[n_L, n_H]$	Support of the probability density function f
$R(n)$	Advertiser's benefit due to an influencer with n true followers
$u(n)$	Influencer's displayed number of followers (true + fake)
$v(n)$	Variable payment to the influencer
w	Fixed wage of influencer
$c(u(n) - n)$	Cost of displaying $u(n)$ followers when true followers are n
Π	Advertiser's payoff
Y	Influencer's payoff
$\mathbb{H}$	Hamiltonian
$\lambda(n)$	Co-state variable
μ	Lagrange multiplier
ψ	Lagrange multiplier
γ	Probability of successful detection of the influencer's fake followers
δ	Penalty factor imposed due to successful detection of fake followers
ρ	Fraction of influencer's escalated cost paid as restitution to advertiser

Table 3: Ordinary least squares regression results

VARIABLES	(1) $u(n)-n$	(2) $u(n)-n$	(3) a	(4) a
n	-0.273***	-0.273***		
	(0.00213)	(0.00209)		
$\log n$	0.0172***	0.0172***		
	(0.000117)	(0.000114)		
γ	-0.0740***	-0.00486	0.0820***	-0.00636***
	(0.00172)	(0.00797)	(0.000497)	(0.00133)
δ	-0.0187***	-0.00855***	0.0207***	0.00377***
	(0.000421)	(0.00144)	(0.000122)	(0.000242)
ρ	0.0275***	-0.00624	0.0341***	-0.00627***
	(0.00172)	(0.00797)	(0.000497)	(0.00133)
$\gamma\delta$		-0.0280***		0.0208***
		(0.00244)		(0.000409)
$\gamma\rho$		0.0212		0.00212
		(0.0135)		(0.00226)
$\delta\rho$		0.00629***		0.00483***
		(0.00244)		(0.000409)
$\gamma\delta\rho$		0.00285		0.0165***
		(0.00412)		(0.000691)
Constant	0.430***	0.407***	0.311***	0.375***
	(0.00223)	(0.00488)	(0.000531)	(0.000789)
Observations	11,979	11,979	11,979	11,979
R-squared	0.704	0.715	0.836	0.947

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

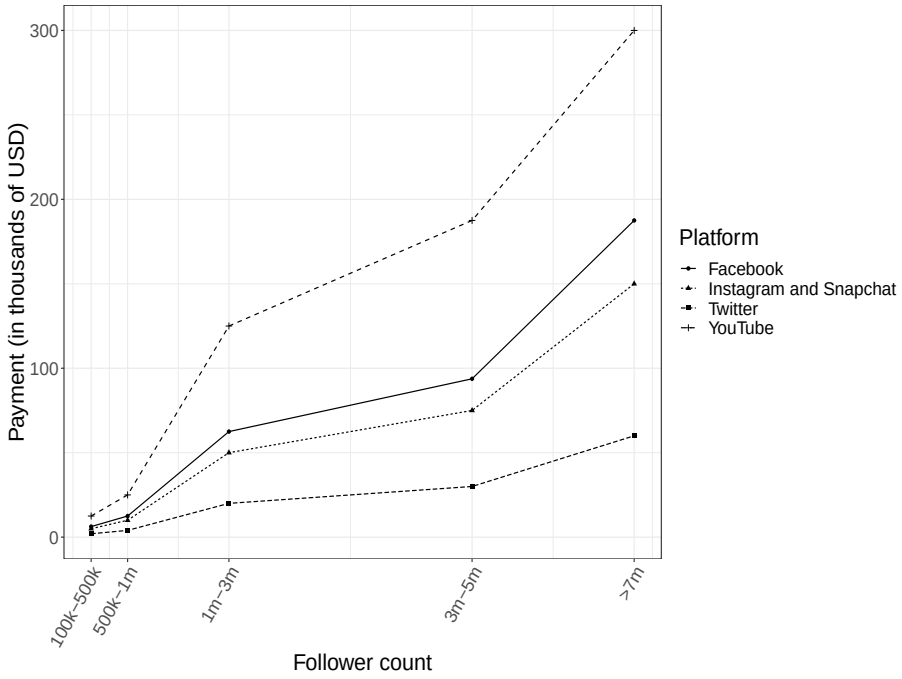

Figure 1: Typical compensation schemes for influencers versus follower counts on various platforms. Adapted from The Economist (2016)

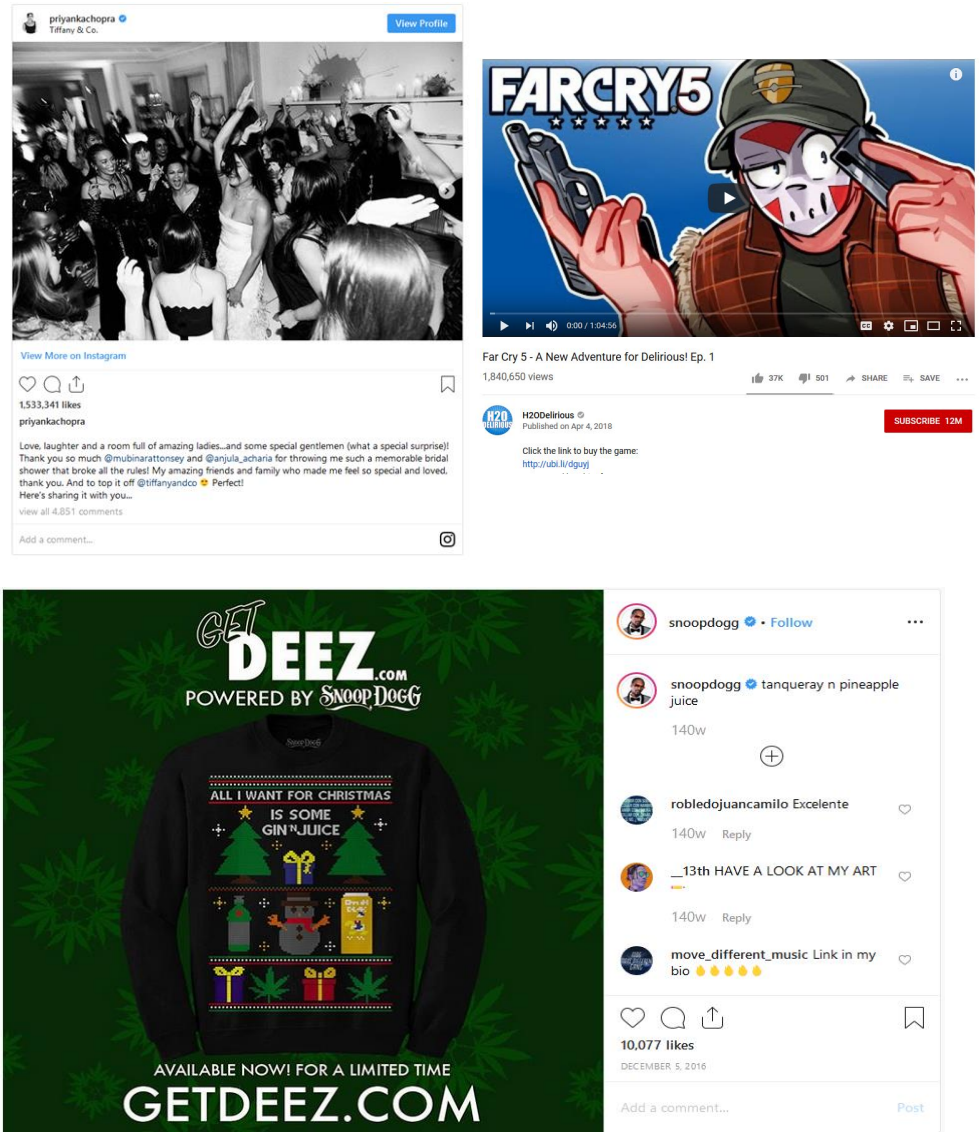

Figure 2: Some prominent influencer campaigns on social media. *Clockwise from top left:* Priyanka Chopra for Tiffany and co. jewelry (Instagram), H2ODelirious for Ubisoft gaming (YouTube) and Snoop Dogg for Tanqueray gin (Instagram)

Figure 3: Snapshots from some websites selling fake followers for different social media platforms

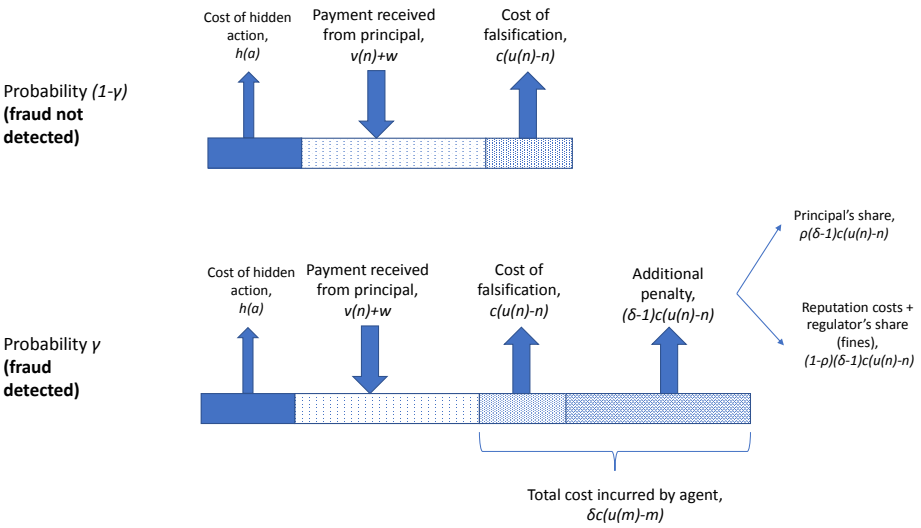

Figure 4: Illustrating the various costs and payoffs

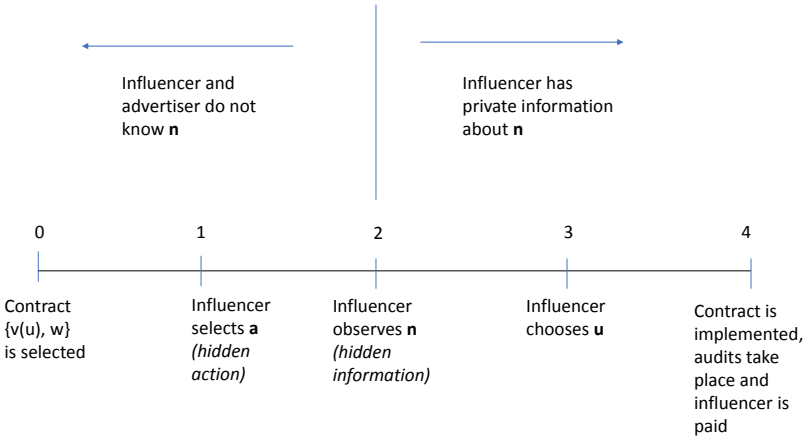

Figure 5: The game sequence between advertiser and influencer

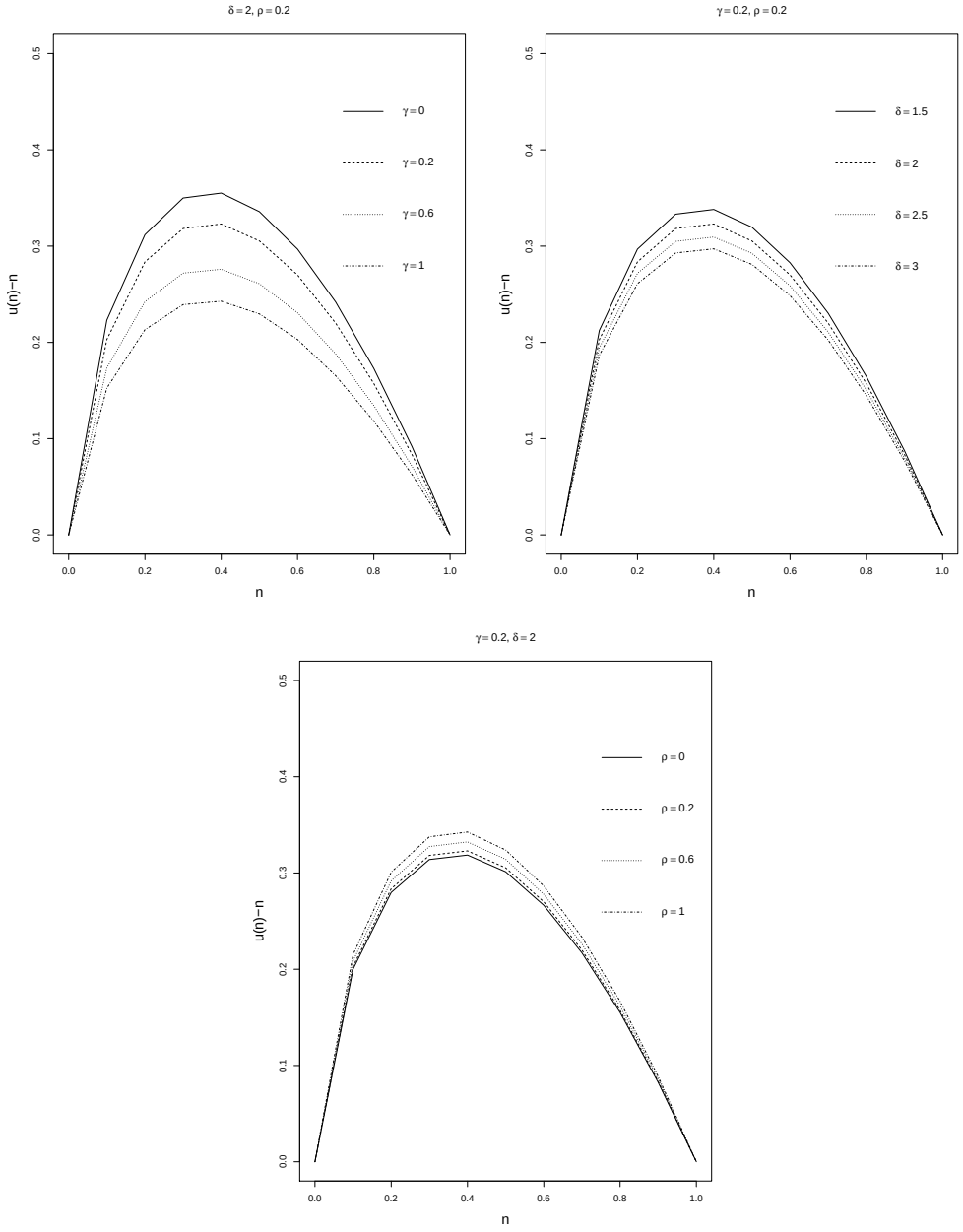

Figure 6: Impact of audits on faking levels. In each plot, we keep two exogenous parameters at the base level indicated at the top, and then plot $u(n) - n$ vs. n for four levels of the third parameter

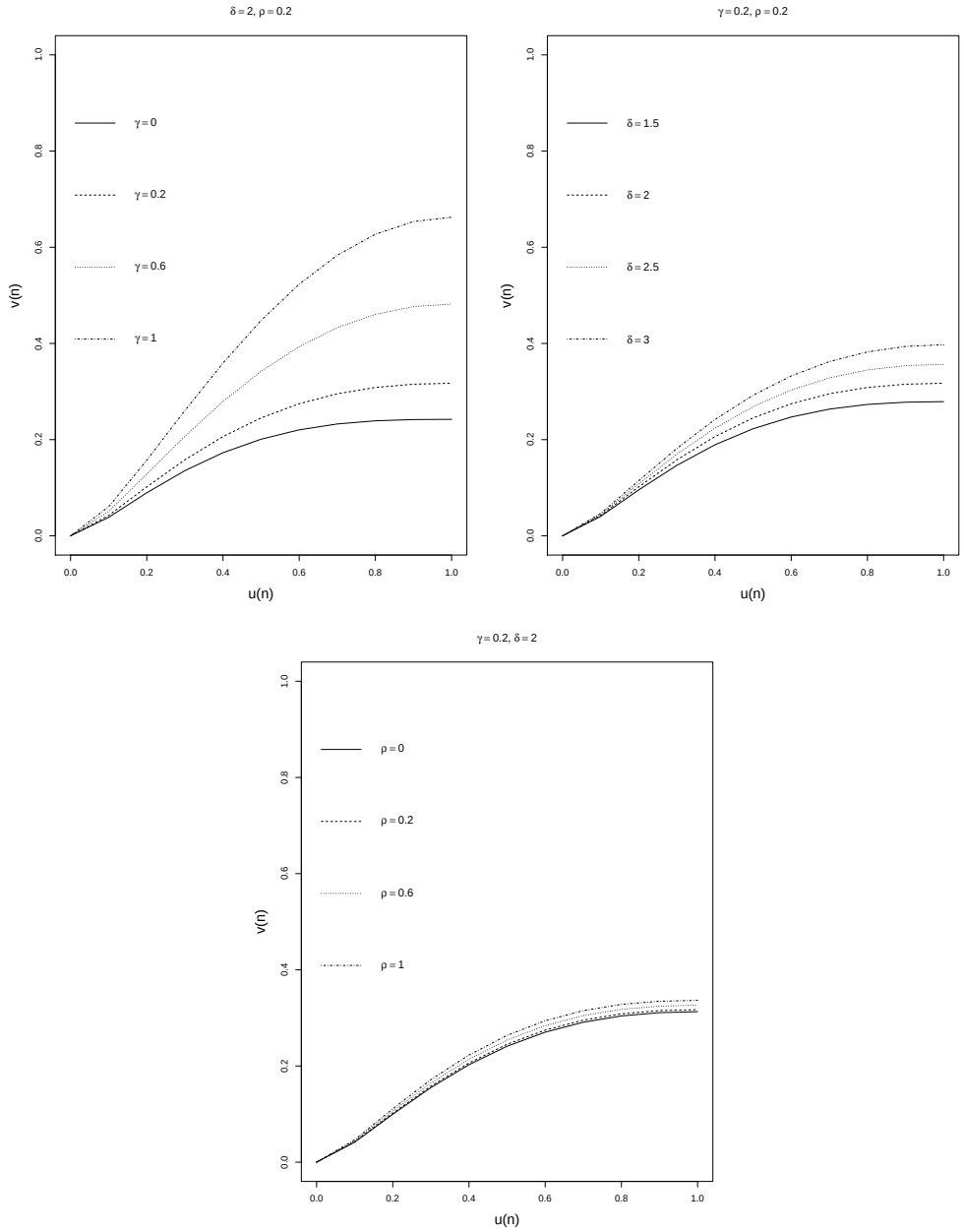

Figure 7: Impact of audits on payments. In each plot, we keep two exogenous parameters at the base level indicated at the top, and then plot the variable payment $v(n)$ vs. the publicly displayed follower count $u(n)$ for four levels of the third parameter