Controlling Fake Reviews

Yuta Yasui

August 25, 2020

Introduction

- Rating systems play key roles in platform markets:
  - Hollenbeck et al (2019): ratings vs advertisement in hotel industry
  - Reimers and Waldfogel (2020): ratings vs professtional reviews for books
- At the same time, the incentive to make fake reviews is growing.
- Effort to reduce fake reviews:
  - Amazon is strictly prohibiting incentivized reviews since 2016.
  - In 2019, FTC filed the first case challenging fake paid reviews: Cure Encapsulations:
    - selling a weight loss pill ($12.8 million in sales on Amazon)
    - paid AmazonVerifiedReview.com for fake reviews

Question: (How) Should a platform reduce fake reviews?

- Are fake reviews harmful?
  - Rational buyers might not be fooled by the fake reviews.
  - Costly fake reviews might work as a signal of good quality.
    - It might pay off only for high quality sellers through future sales. (Nelson; 70,74)
- Instruments of the platform:
  1. intensity of censorship on fake reviews
  2. weights on previous reviews and new reviews,

Overview

- The number of fake reviews is increasing in quality, decreasing in reputation.
- The stringent censorship reduces
  - the number of fake reviews in expectation,
  - the effects of fake reviews.
- For rational consumers:
  - a rating with fake reviews can be more informative than one without fake reviews
  - transition speed of the rating should be slower than the optimal level without fake reviews.
- For credulous consumers:
  - the stringent censorship reduces bias for the credulous consumers as long as positive number of fake reviews are observed.

Literature

Design of Rating Systems

- **[certification]** Lizzeri (1999), Harbaugh and Rasmusen (2018), DeMarzo, Kremer, Skrzypacz (2019), Hopenhayn and Saeedi (2019), Hui et al (2018), Zapechelnyuk (2020)
- **[scoring][one-shot]** Ball (2019), **[dynamic]** Vellodi (2019): entry/exit, directed search;

  Horner and Lambert (2018), Bonatti and Cisternas (2020), **this paper**: **signal jamming**

	platform controls..			
	weights	censorship	eqm action	credulous consumer
HL2018	Y	(Y)	constant	N
BC2020	Y	N	Markov	N
This paper	**Y**	**Y**	**Markov**	**Y**

Promotion and Signaling (Q: The higher quality, the more promotion?)

- Nelson (1970, 1974), Kihlstrom and Riordan (1984), Milgrom and Roberts (1986), Horstmann and Moorthy (2003), Mayzlin (2006), Dellarocas (2006): **One shot promotion**
- Horstmann and MacDonald (1994), Saraiva (2020) [numerical/empirical], and **this paper**: **Repeated promotions**; **true quality** and **reputation** play different/interactive roles

Motivating example

Fake reviews with "verified purchase" on Amazon

1. Fake reviewers are **refunded** by the seller. Refunds are done outside of of the platform.

   - cf) official review programs on Amazon
     - *Early Reviewer Program*: Amazon offers \$1-3 for a review of a previously purchased product
     - *Vine Voice*: Reviews for free not-yet-released products (invitation only)

2. The platform takes a **transaction fee** from each transaction
3. The platform can detect a part of fake reviews.

Model (1/3)

- Time: $t \in [0, \infty)$
- Players: a long lived seller, many short lived buyers
  - (a platform can control parameters before the game starts)
- Action at time t
  - Seller:
    - (sell one unit of the product: **fixed**)
    - choose the amount of the fake reviews: $F_t \in \mathbb{R}$
  - Buyers:
    - buy the product, or not
      $\rightarrow$ form the equilibrium price: $p_t = E[\theta_t | Y_t] \equiv M_t$ Details
- State:
  - θ_t: seller's type (quality of the product) at t
  - Y_t: seller's rating at t
- Information:
  - Seller at time t: the whole history so far $= (\theta_s, Y_s, F_s, p_s)_{s \in [0, t]}$
  - Buyers at time t: current rating $= Y_t$

Model (2/3)

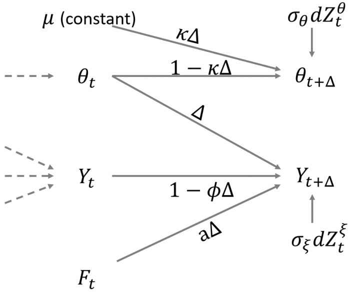
- State transition:
  - quality θ_t follows

$$d\theta_t = -\kappa\,(\theta_t - \mu)\,dt + \sigma_\theta dZ_t^\theta$$

    exogenous for players (seller/buyers) and for the platform
  - rating Y_t follows

$$dY_t = -\phi Y_t dt + aF_t dt + \theta_t dt + \sigma_\xi dZ_t^\xi$$

    $a > 0$: effectiveness of fake reviews. (low a = stringent censorship)



Model (3/3)

- Seller's instantaneous payoff:

$$\pi_t = \underbrace{(1-\tau)\,p_t\,(1+F_t)}_{\text{revenue}} - \underbrace{p_t \cdot F_t}_{\text{costs of refund}} - \underbrace{\frac{c}{2}F_t^2}_{\text{additional costs}}$$
$$= (1-\tau)\,p_t - \tau p_t \cdot F_t - \frac{c}{2}F_t^2$$

  - τ: transaction fee imposed by the platform.
- The market determines $p_t = E\left[\theta_t | Y_t\right] \equiv M_t$

$$\pi_t = (1-\tau)\,M_t - \tau M_t \cdot F_t - \frac{c}{2}F_t^2$$

  - $\tau = 0$: a. la. Holmstrom (1999), a special case of Horner and Lambert (2018)

Definition of Equilibrium

Stationary Linear Markov equilibrium

Definition

A linear Markov strategy $F = (F_t)_{t\geq 0}$ s.t. $F_t = \alpha\theta_t + \beta Y_t + \gamma$ is a stationary linear Markov equilibrium if

1. $F = \arg\max_{(\tilde{F}_t)_{t\geq 0}} E_0\left[\int_0^\infty e^{-tr}\left(\left(1-\tau\right)M_t - \tau M_t \cdot \tilde{F}_t - \frac{c}{2}\tilde{F}_t^2\right)\right]$
2. $M_t = E^F\left[\theta_t | Y_t\right]$
3. $(\theta_t,\ Y_t)_{t\geq 0}$ induced by F is stationary Gaussian

Stationarity of Equilibrium

- Transition of $(\theta_t,\ Y_t)$ (in discrete analogue):

$$\theta_{t+dt} = \theta_t \left(1 - \kappa dt\right) + \mu\kappa dt + \sigma_\theta dZ_t^\theta$$

$$Y_{t+dt} = Y_t \left(1 - (\phi - a\beta) dt\right) + \theta_t \left(1 + a\alpha\right) dt + a\gamma dt + \sigma_\xi dZ_t^\xi$$

  - $(\theta_t,\ Y_t)$ is stationary Gaussian if $\phi - a\beta > 0$
- When $(\theta_t,\ Y_t)$ is stationary Gaussian, then

$$M_t \equiv E\left[\theta_t | Y_t\right] = \underbrace{E\left[\theta_t\right]}_{\equiv\mu} + \underbrace{\frac{Cov\left(\theta_t,\ Y_t\right)}{Var\left(Y_t\right)}}_{\equiv\lambda}\left[Y_t - \underbrace{E\left[Y_t\right]}_{\equiv\bar{Y}}\right]$$

Characterize Equilibrium

- HJB equation:

$$
\begin{aligned}
rV(\theta, M) = \sup_{F \in \mathbb{R}} & (1-\tau) M \cdot q - \tau M \cdot F - \frac{c}{2} F^2 \\
& - \kappa (\theta - \mu) V_\theta + \{-\phi Y_t + a\lambda F + \theta\} V_M \\
& + \frac{\sigma_\theta^2}{2} V_{\theta\theta} + \frac{\sigma_\xi^2}{2} V_{YY} \\
\text{s.t. } & M = \mu + \lambda [Y - \bar{Y}]
\end{aligned}
$$

- Note: θ appears in the transition of states
- The equilibrium is characterized by guess-and-verify of
  - $F = \alpha\theta + \beta Y + \gamma$ (linear strategy)
  - $V = v_0 + v_1\theta + v_2 Y + v_3\theta^2 + v_4 Y^2 + v_5 Y\theta$ (quadratic value function)
  - $\phi - a\beta > 0$ (stationarity)

Theorem (Existence and uniqueness)

There exists a stationary linear Markov equilibrium. In this equilibrium, $\alpha > 0$, $\beta < 0$, $\lambda > 0$. The equilibrium is unique and continuously differentiable in parameters if a loose condition in parameters holds.

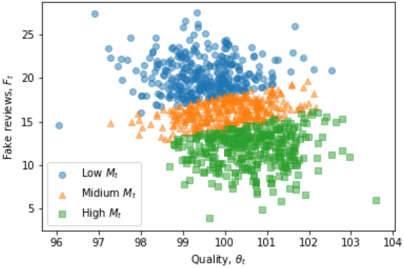
- Reminder: $F_t = \alpha\theta_t + \beta Y_t + \gamma$
- Uniqueness holds if $\phi > \kappa$ (rating evolves faster than underlying quality).
- $\beta < 0$:
  - Intuition: Higher rating, higher price, higher marginal cost of fake reviews, less fake reviews ▸ Details
- $\alpha > 0$:
  - Note: $\alpha > 0$ is **not from** the incentive to increase **revenue**
    - Higher θ_t, higher p_{t+dt} even without fake reviews
    - $\frac{dp_{t+dt}}{dF_t}$ does not depend on θ_t
  - Higher quality, more **cost-saving** in the future
    - Once its rating is boosted, the future self will reduce the fake reviews.
    - This effect remains for a long time, given high θ_t. ▸ Details

Consistency to data:

- $\beta < 0$ is consistent with Luca and Zervas (2016)
  - More manipulation after a drop of a rating

Implication to empirical literature:

- Hard to capture positive relationship b/w promotion level and quality.
  1. The rating should **not** be used as a proxy for quality
  2. Even with true quality index, researcher needs to control reputation.



Assume that the platform can change a and ϕ

- Recall:
  - Rating: $Y_{t+dt} = Y_t \times (1 - \phi dt) + d\xi_t$ (higher ϕ, a higher weight on today's review & faster transition)
  - New review:$d\xi_t = aF_t dt + \theta_t dt + \sigma_\xi dZ_t^\xi$ (smaller a, more stringent filtering)
- [In the paper, I am working on comparative statics about τ and σ_ξ]

Proposition

$E[F_t]$ is increasing in a. $E[F_t] \geq 0$ for sufficiently large a.

Proposition (The effects of fake reviews)

$a \cdot \alpha$, $a \cdot \beta$, $a \cdot \gamma$ goes to zero as $a \to 0$.

- Reminder: $aF_t = a\alpha\theta_t + a\beta Y_t + a\gamma =$ the effect of fake reviews
- Stringent censorship can **reduce** the **expected amount** and the **effects** of fake reviews.
- Note: $(\alpha, \beta, \gamma) \nrightarrow 0$ even when $E[F_t] \to 0$ or $(a\alpha, a\beta, a\gamma) \to 0$

Q: **Should** the platform reduce fake reviews?

Criteria: $\rho^2 = \frac{Cov(\theta_t, Y_t)^2}{Var(\theta_t)Var(Y_t)}$

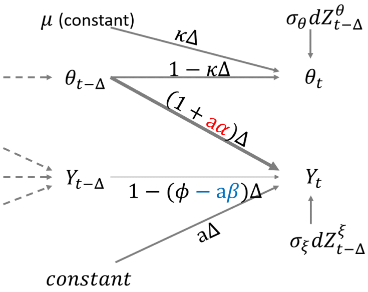
- Motivation:
- Regulators often want to make rating sytems informative.
- For the platform, if the rating system is not informative, the sellers and buyers might move out to other platforms.
  - Maximization of ρ^2 is equivalent to minimizing $Var(\theta|Y)$

$$Var(\theta|Y) = \underbrace{Var(\theta)}_{\text{exogenous}} \left(1 - \rho^2\right)$$

  - Note: M_t is an unbiased estimate of θ_t $(E[E[\theta_t|Y_t]] = E[\theta_t])$

Q: Should the platform reduce fake reviews?

Criteria: $\rho^2 = \frac{(\phi - a\beta)}{(\kappa+\phi-a\beta)} \frac{(a\alpha+1)^2}{((a\alpha+1)^2+\kappa(\sigma_\xi/\sigma_\theta)^2(\kappa+\phi-a\beta))}$ (given any α, β, δ)



- Impacts of the fake reviews:
  1. $a \cdot \alpha > 0$ enhances the positive relationship between the true quality θ_t and the rating Y_t.
  2. $-a\beta > 0$ increases the transition speed of the rating Y_t.

Proposition (Informativeness of fake reviews)

*The **equilibrium** strategy is **more informative than no-fake** strategy under a set of parameters such that*

1. *a is sufficiently large, or*
2. *(i)a is sufficiently small and*
   (ii) $\phi^2 < \frac{\sigma_\theta^2}{\sigma_\xi^2} + \kappa^2$

Intuition:

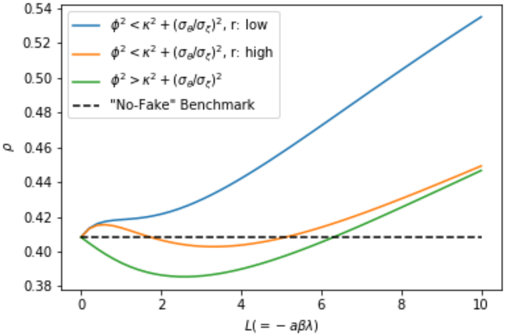
1. Higher a, higher $a \cdot \alpha$, more informative.
2. $a\beta < 0$ makes the reputation transition speed $\phi - a\beta$ faster.
   - This is good if ϕ was too small, and
   - bad if ϕ was already sufficiently high.

- First effect dominates for large a, and second effect dominate for small a.

Sketch of the proof:

- $\rho^2 = \frac{(\phi+L)}{(\kappa+\phi+L)} \frac{(A(L)+1)^2}{((A(L)+1)^2+\kappa(\sigma_\xi/\sigma_\theta)^2(\kappa+\phi+L))} = \rho^2(L, \phi)$ (given **eqm** α, β, δ)
  - L (eqm effect on the transition speed) is positive and increasing in a.

1. $\lim_{L\to\infty} \rho^2 = 1$
2. $\frac{\partial \rho^2}{\partial L}|_{L=0} > 0$ iff $\phi^2 < \frac{\sigma_\theta^2}{\sigma_\xi^2} + \kappa^2$

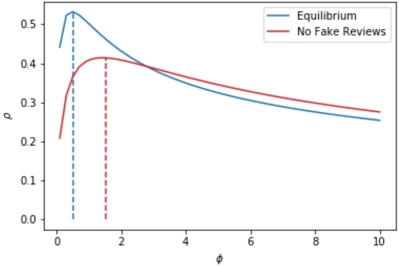


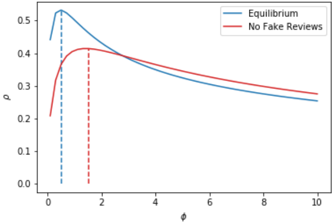
Q: How does the optimal ϕ (transition speed of the rating) change due to the fake reviews?

- Comparison with the optimal ϕ without fake reviews, ϕ^0
- $\phi^0 = \arg\max_\phi \rho^2 (0, \phi)$
- $\phi^* = \arg\max_\phi \rho^2 (L(\phi), \phi)$

Proposition

$\rho^2 (L(\phi), \phi)$ is decreasing in ϕ at $\phi = \phi^0$.
Furthermore, $\rho^2 (0, \phi^0) < \rho^2 (L(\phi^*), \phi^*)$ for sufficiently small r.





Intuition

- $\rho^2(L(\phi), \phi)$ is decreasing in ϕ at $\phi = \phi^0$.
  - w.o./ fake reviews: the platform can control the transition speed ϕ
  - w/ fake reviews: effective transition speed is $\phi + L$
  - $\rightarrow$ the platform should adjust ϕ downward.
- $\rho^2\left(0, \phi^0\right) < \rho^2\left(L\left(\phi^L\right), \phi^L\right)$ for sufficiently small r
  - small $r \Rightarrow$ high weight on the future $\Rightarrow$ high $\alpha > 0 \Rightarrow$ informative (ie, $\rho^2\left(0, \phi^0\right) < \rho^2\left(L\left(\phi^0\right), \phi^0\right)$ given small r)
  - By definition of ϕ^*, $\rho^2\left(L\left(\phi^0\right), \phi^0\right) < \rho^2\left(L\left(\phi^*\right), \phi^*\right)$

Credulous Consumers

- Credulous consumers believe that
  - they face a stationary Gaussian distribution of (θ_t, Y_t)
  - there is no fake reviews by the seller. (ie, assume $\alpha = \beta = \gamma = 0$)
- Reputation:
  - rational consumers: $M_t = E^F[\theta_t | Y_t] = \mu + \lambda(\alpha, \beta)\left[Y_t - \bar{Y}(\alpha, \beta, \gamma)\right]$
  - credulous consumers: $\widetilde{M}_t = \widetilde{E}[\theta_t | Y_t] = \mu + \lambda(0, 0)\left[Y_t - \bar{Y}(0, 0, 0)\right]$
    - belief based an wrong joint distribution of (θ_t, Y_t)
- Seller's payoff:
  - $\pi_t = (1-\tau)p_t - \tau p_t \cdot F_t - \frac{c}{2}F_t^2$
  - $p_t = \eta M_t + (1-\eta)\widetilde{M}_t$ where $\eta \in [0, 1]$
  - Interpretation:
    - η captures the rationality of all consumers
    - η is the ratio of rational consumers in the market. Details

Theorem

Existence and uniqueness given the same condition as the baseline model

Proposition

Existence of the credulous consumers decreases $E[F_t]$.

- Intuition:
  - Credulous consumers are less sensitive to the rating than rational consumers.
    - Rational consumers regard the rating infomative because of $a\alpha > 0$.
  - Less marginal benefit with credulous consumers.
  - Less fake reviews with credulous consumers.

Criteria for the credulous consumers:

$$Bias = E\left[\widetilde{E}\left[\theta_t | Y_t\right] - \theta_t\right]$$

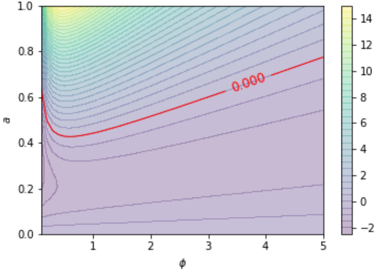
Lemma

$Bias \geq 0$ iff $E[F_t] \geq 0$.

Suppose there are only credulous consumers in the market.

Proposition

Stringent censorship policy reduces Bias as long as $E[F_t] \geq 0$.



Summary

Positive Analysis:

- The number of fake reviews is increasing in quality, decreasing in reputation.
- The stringent censorship
  - reduces fake reviews in expectation, but
  - reduces the effects of fake reviews.

Normative Analysis:

- For rational consumers:
  - a rating with fake reviews can be more informative than without fake reviews
  - Transition speed of the rating should be slower than the optimal level without fake reviews.
- For credulous consumers:
  - As long as $E[F_t] \geq 0$, the more stringent censorship, the less bias for the credulous consumers.

Intuition of the Equilibrium Strategy

Back to Theorem

- Reminder: $V = v_0 + v_1\theta + v_2 Y + v_3\theta^2 + v_4 Y^2 + v_5\theta Y$
- FOC: $F_t = -\frac{\tau}{c} M_t + \frac{a}{c}\left\{v_2 + 2Y_t v_4 + \theta v_5\right\}$
- $\beta < 0$
  - $\beta = -\frac{\tau}{c}\lambda + 2\frac{a}{c} v_4 = -\frac{\tau}{c}\lambda + \frac{a}{c}\frac{-\beta\lambda\tau}{(-a\beta+r+2\phi)}$
  - $\beta < 0$ since today's cost saving incentive dominates.
- $v_4 = \frac{-\beta\lambda\tau}{2(-a\beta\lambda+r+2\phi)} > 0$
  - Higher reputation, less promotion, less costly.
- $\alpha > 0$
  - $\alpha = \frac{a}{c} v_5$
  - $v_5 = \frac{1}{\kappa+r+\phi}\{\underbrace{2(a\alpha + bq)}_{\text{feedback to future } Y_t} v_4 - \alpha\lambda\tau\}$
    - Driving Force: Higher quality, higher reputation in the future, cost reduction in the future.
    - Counteracting effect: Higher quality, more promotion today/in the near future (if $\alpha > 0$)

Microfoundation of the price: $p_t = M_t$

Back to Model

- (Reminder: $M_t \equiv E\left[\theta_t \middle| Y_t\right]$)
- Suppose there is a mass (2) of buyers.
- Consumer $i \in [0, 2]$ feels

$$u_i = \begin{cases} \theta + \epsilon_i - p & \text{if the consumer purchase the product} \\ 0 & \text{otherwise} \end{cases}$$

  - $\epsilon_i \sim_{i.i.d.} F\left(\cdot\right)$ where $F\left(\cdot\right)$ is symmetric around zero
- Given Y, rational consumer purchases iff $M + \epsilon_i - p \geq 0$
- Market clearing

$$\begin{aligned} 1 &= 2q \cdot \left(1 - F\left(p - M\right)\right) \\ \Leftrightarrow p &= M \end{aligned}$$

Mixture of the rational/credulous consumers

▸ Back to Model

- $M = E[\theta|Y]$: rational consumer's belief (on the seller's quality)
- $\widetilde{M} = \widetilde{E}[\theta|Y]$: credulous consumer's belief (on the seller's quality)

Ratinale:

- 2η rational consumers and $2(1-\eta)$ credulous consumers in mkt
- Consumer $i \in [0, 2]$ feels

$$u_i = \begin{cases} \theta + \epsilon_i - p & \text{if the consumer purchase the product} \\ 0 & \text{otherwise} \end{cases}$$

  - $\epsilon_i \sim U(-C, C)$: iid over the consumer types.
- Rational consumer purchases iff $M_t + \epsilon_i - p \geq 0$
- Credulous consumer purchases iff $\widetilde{M}_t + \epsilon_i - p \geq 0$
- Market clearing

$$1 = 2\eta \cdot (1 - F(p - M)) + 2(1-\eta) \cdot \left(1 - F\left(p - \widetilde{M}\right)\right)$$

$$\Leftrightarrow p = \eta M + (1-\eta)\widetilde{M}$$