DYNAMIC INCENTIVES BEHIND MANIPULATED ONLINE REVIEWS

Hisayuki Yoshimoto and Andriy Zapechelnyuk

Abstract. We study the dynamic incentives behind the practice of firms writing potentially manipulated online reviews. First, we propose a stylized model with a Bellman equation, which characterizes marginal economic costs and gains for a firm to write inflated reviews. The model shows that, under some review score aggregation rules, the incentive to write manipulated reviews diminishes as the number of reviews increases. Second, to test the model predictions, we investigate online reviews among high-end restaurants. Specifically, we compare (1) online reviews, which anyone can write and are subject to manipulation, and (2) professional restaurant guidebook reviews, which are written by professional reviewers and less subject to manipulation. Notably, we exploit a discontinuity in the review score aggregation rule. We empirically find that restaurant review dynamics are heterogeneous, but some of them fall into the model-predicted behavior of review manipulation.

JEL Classification: L15, L81, L86,

Keywords: Product quality information, Online reviews, Costly review manipulation, Reputation, Platform

1. Introduction

When and in what economic environments are firms more prone to write manipulated reviews? With the theory-and-evidence approach, this article investigates dynamic reputational incentives to write manipulated reviews. We first propose a stylized model, characterizing firm-side manipulated-review-writing behavior, which comes with economic costs. We then examine the model prediction with restaurant online review scores and printed guidebook review scores among high-end restaurants in a specific country.

This study contributes to the literature by proposing a dynamic microeconomic model for analyzing the incentives behind manipulated reviews, and by testing the predictions of this model with restaurant review data. By exploiting the nature of dynamics and display discontinuities, this paper also contributes to the detection of manipulated online reviews, which is of interest to competition authorities.

With the advent of large-scale online review platforms, such as Amazon, Google, Expedia, Yelp, Booking.com, Apple Store, Uber, Lyft, etc., self-promoting manipulated online reviews have become a challenge for informed consumer choices. By writing self-promotional reviews for its own business, a firm could inflate the demand (willingness to pay) for its product and set a higher price to increase its revenue. The potential manipulability of online reviews is also of the public interest, due to consumer welfare protection.[1]

As reported in the literature review section, there are several preceding academic works, both in theory and empirics, examining the existence, scope, degree, and patterns of manipulated reviews, including the pioneering study of Mayzlin, Dover, and Chevalier [2014]. However, most of the preceding studies are based on static analyses, abstracting the dynamic incentives. To the best of our knowledge, only a few preceding studies investigate the dynamics of potentially manipulated reviews. This article aims to fill such a gap by proposing a dynamic

[1]For example, the Competition and Market Authority (CMA), the regulatory authority of the UK government for market competition and consumer protection, launched an investigation on potential fake reviews on Amazon and Google in 2021. Source: CMA to investigate Amazon and Google over fake reviews (2021): `https://www.gov.uk/government/news/cma-to-investigate-amazon-and-google-over-fake-reviews`.

model and examining its model-predicted behavior with high-end restaurant review data. As described below, our approach is in two steps, model analysis and empirical investigation.

First, we propose a dynamic Bellman equation model for potentially manipulated review writing activities, in which the firm (potentially manipulated review writer) engages in costly self-promoting review activity and makes dynamic tradeoffs. The dynamic economic tradeoff that the firm faces is costly manipulated review writing in a current period and future gain from an inflated review score. Notably, our model exploits a review score aggregation rule, the arithmetic average review score, displayed at the top of the online review webpage (for a specific restaurant), which we observe in our dataset.[2]

Our model has two state variables: an arithmetic-average score (e.g., the average score of 4.23 out of a maximum of 5.00 score scale) and an accumulating number of reviews (e.g., 538 past reviews already written). We use these model specifications, as it is globally common that an online review platform website (e.g., a restaurant review website for a specific restaurant or a platform website for a particular product) displays an aggregate score, which consists of some score aggregation rule. We focus on the arithmetic average score, as (1) this aggregation rule is observed in our dataset, and (2) it concisely captures the diminishing return from writing an additional self-promotional review: If a restaurant already has a large number of reviews (e.g., 2,649 reviews), the influence of an additionally manipulated review is relatively small, compared to that with a small number of reviews (e.g., only 195 reviews). Particularly, we drive several testable model predictions for empirical investigation.

Second, we examine the model predictions with an online review dataset of high-end restaurants. To do so, we use a dataset from a restaurant online review platform, as well as reviews from widely-circulated printed guidebooks, in a specific country. Specifically, we focus on high-end restaurants, as it allows us to directly compare online reviews and guidebook reviews. We exploit the gaps in costly manipulated review writing activities between two

[2]With the cost of modeling complication, our model could be extended to other score aggregation rules, such as the average of most recent two-year review scores or the last 100 reviews, although the number of state spaces and transition rules would become a challenge in economic analysis. For this reason, we focus on the simplest arithmetic average score, which the platform in our dataset displays.

different review formats: While it is relatively low cost to write manipulated online reviews, it is relatively more costly to manipulate scores in printed guidebooks. For instance, to write an additional self-promotional online review, a restaurant worker only needs to create a new account (which does not require an authenticity check, such as a driver's license check). On the other hand, to inflate a restaurant review score in a well-circulated guidebook, for which professional food critics report scores, a restaurant worker needs to approach guidebook reviewers and potentially bribe, which is at least relatively more costly, if not impossible.

To enable this comparison between online scores (for which anyone can write a review) and professional guidebook review scores, we collected scores from about 5,000 guidebook-listed restaurants' online reviews, as well as corresponding professional reviews in well-circulated guidebooks, during the period of 2008-2019 (spanning 12 years). We choose the starting year of 2008, as the major online-restaurant review platform launched its website in 2008 in this specific country.

Regarding the online reviews of guidebook-listed restaurants, on average, each restaurant has more than 200 reviews. Thus, the sample size of the online reviews exceeds more than one million (5,000 restaurants multiply by 200 reviews). Each online review consists of an individual user's review score, date of review, reviewer ID, and verbal review comments.[3] It is worth noting that these online reviews (of guidebook-listed restaurants) are happening on an ongoing basis. An online review can be written whenever an online reviewer wants to provide her/his opinions related to a specific restaurant, which we will exploit in both modeling and empirical analyses.

Regarding the (physically printed) professionally-rated-guidebook scores, we manually collected the restaurant food evaluation data from four types of well-circulated and branded guidebooks, which were annually published during 2008-2019 (for 12 years). Thus, our guidebook dataset is from 48 physically printed guidebooks (4 brands of guidebooks multiplied by 12 years). Notably, each of these guidebook series (types) is published by a different

[3]For example, a review may look like this: 4 out of 5 stars, July 4, 2017, johndoe67, "Fabulous... but expensive!".

publisher, which claims the independence of its restaurant evaluations, which allows us robustness checks in our empirical analyses. Also, note that these printed (and independently edited) guidebook series have annual publication cycles, typically a new edition of a branded guidebook is published in the autumn of the previous year (e.g., the 2015 edition of a guidebook is published in November of 2014). As we know the exact publication date of each guidebook, we can match the annual publication cycle and ongoing basis online review score patterns.

By exploiting a large number of online reviews, our empirical study strategy is as follows. In order to investigate the patterns of the potentially manipulated online scores in a linear regression, we construct the dependent variable of review score gaps: an online score (which is relatively costless to manipulate) subtracted by a guidebook score (which is relatively costly to manipulate). This normalization allows us to eliminate unobserved restaurant-level heterogeneity. The explanatory variables are the set of observable variables, including cuisine types and prices, as well as timing-related variables (e.g., post-new-guidebook publication period indicator). Notably, we exploit the discontinuity of displayed scores, which are explained in the data description section.

Our empirical analyses show that, although the dynamics of the online reviews are quite heterogeneous, a sizable portion of restaurants' dynamic score patterns fall into the model prediction of manipulated online reviews, especially when the number of accumulated reviews is relatively small.

2. Literture Review

The paper most related to ours is the seminal study of Mayzlin, Dover, and Chevalier [2014], in which the authors report a static analysis of manipulated online reviews. Specifically, they compare hotel reviews from Tripadvisor.com (where a booking confirmation is not required to write a review) and Expedia.com (where only customers who actually booked and stayed at a hotel can write reviews). The authors compare the proportion of negative reviews (1 and 2 ratings, on a 1-5 scale), as well as positive reviews (4 to 5 ratings) between these two websites. They discovered significant review manipulation, especially in the form of negative reviews.

Related to our regression discontinuity, Anderson and Magruder [2012] examine restaurant review data from Yelp.com to investigate the effect of online reviews on restaurant demand, measured by reservation availability. Similar to our study, the authors exploited the 0.5 star rounding rule for displayed scores, but they found little evidence for fake reviews in their sample from Yelp.com.

Using the data related to Amazon.com, He, Hollenbeck, and Proserpio [2022] empirically analyses the market for fake reviews, investigating the relations among product ratings, sales rankings, advertising, and pricing strategy. The author reports that the observed review-buying activities have links to the increases in average ratings and the number of reviews: After a firm stop buying fake reviews, the average ranking falls, and the proportion of one-star reviews increases significantly, notably among young products. With a Tripadvisor hotel review dataset, Hollenbeck, Moorthy, and Proserpio [2019] report that online reviews and advertising spendings are substitutes, as well as reporting such substitute relationship is stronger for independent (non-hotel-chain affiliated) hotels.

Mayzlin [2006] proposes a strategic interaction model of promotional chat and reviews provided by firms, in which consumers have uncertainty about product quality. In equilibrium, firms spend more economic resources on promoting low-quality products, which raises welfare concerns.

Regarding the reviews on large-size online platforms, Reimers and Waldfogel [2021], Belleflamme and Peitz [2021] and Belleflamme and Peitz provide broader views and economic insights, including welfare implications.

3. Industry Background and Data Descriptions

This study compares online restaurant reviews and professional guidebook reviews, observed in an anonymous country, called Country J. This section describes the industry background, and the next section provides detailed descriptions of the datasets. As outlined below, these four guidebooks have different evaluation systems, which are professionally rated. Our research design is based on the idea that the cuisine evaluation scores in the guidebooks are hard to manipulate due to the authenticity and reputational concerns of the guidebook publishers. Meanwhile, evaluation scores online are relatively easy to manipulate. Accordingly, we focus on the cuisine evaluation systems and other key observables in our empirical analyses.

3.1. **Online Restaurant Review Data.** This subsection briefly describes the online restaurant review data, extracted from online resstaurant review platofrorm. There are two types of web data: (a) restaurant-level and (b) review-level. We use the term "overall experience score", which provides an overall review for the quality of the restaurant, and which is different from the categorical scores described below.[4]

There is a major restaurant review website in Country J, which we label Website A. We also call this webite as online-restaurant-review platform A. Website A has a dominant position in the online restaurant review sector in this country, measured by search frequencies (Google Trends). Indeed, the second most popular website has a much lower search frequency. Thus, our main focus is the online review scores listed on Website A.

Similar to restaurant and hotel review websites in other countries, Website A provides a separate page for each restaurant, where consumers can post reviews. However, the requirement to be a reviewer is quite low. To be an online reviewer, Website A does not require any ID documents, such as a driver's license or passport. In addition, it does not require proof of an actual visit. Thus, anyone who uses a unique email address can write online reviews, and the creation of multiple accounts is reasonably considered to have a low cost.

[4]Upon submitting a review, reviewers on Website A are asked to submit scores for overall experience, food, service, value, and atmosphere, as well as a written review. The provision of photos is optional.

	Cuisine Score Evaluation System (Before Normalization)	Number of Listed Restaurants	Note and Other Listed Variables
Guidebook I	0-3 Scale	About 1,700	150 for 1-3 rating frestaurants and 1,550 for 0 rating restaurants
Guidebook II	1-10 Scale	About 900	Price, chef name(s), and other restaurant characteristics
Guidebook III	1-5 Scale	About 1,700	Chef name(s), owner name(s) (until 2016), and other restaurant characteristics
Guidebook IV	1-5 Scale	About 3,000	Price, ambience score, and service score

Table 1. Summary of Printed Guidebooks

For (a) restaurant-level data, our dataset consists of the contents from the top (front) page of each restaurant's on online-restaurant-review platform A, including restaurant name, address, phone number, overall experience score (arithmetic average, and rounded to 0.5), number of reviews, as well as categorical scores for food, service, value, and atmosphere respectively. The dataset also contains rating distribution information for the overall experience score, i.e. the number of reviews with "Excellent", "Very good", "Average", "Poor", and "Terrible" evaluations.

Regarding (b) review-level data, the dataset contains overall experience score, review date, written review, and date of visit.

3.2. **Guidebook Review Data.** Next, there are four widely circulated restaurant guidebooks in Country J, and we call them Guidebook I, II, III, and IV. At every fall, these guidebooks publish their latest editions. A summary of these professionally rated guidebooks is found in Table 1.

Guidebook I is the local edition of a globally recognized guidebook. The guidebook has a well-known 3-tier rating system, which represents excellence of food and other restaurant characteristics. Although it is less well-known, Guidebook I also lists other (0-tier) restaurants.

Specifically, Guidebook I lists about 150 restaurants rated 1-3, which are generally considered the "best of the best". It also lists about 1,500 0-tier restaurants. Our conversations with professional chefs reveal that it is quite prestigious to be listed in this globally recognized guidebook, and the attainment of a 1-3 rating is the ultimate honor. The chefs also confirmed that a decreased rating, or the loss of listed status, causes substantial business damage, primarily through the reputational channel. Another notable feature of this guidebook is that the number of listed restaurants (mostly in tier 0) has declined substantially in recent years.

Guidebook II is published by a local publisher, located in Country J. Although this guidebook is not a global brand, Google Trends reveals that it is comparably popular in Country J, and sometimes exceeds the globally recognized Guidebook I in terms of search frequency.

Each year, Guidebook II provides cuisine evaluations for about 900 restaurants. One of the notable features this guidebook is its fine-grained score system and detailed cuisine descriptions. It has a 1-10 scale (1 being lowest and 10 highest). Also, it should be noted that Guidebook II lists chef name(s) for each restaurant, which may motivate chefs to be included in this guidebook.

Guidebook III is also published by a local publisher in Country J. Each year, this guidebook lists cuisine evaluations of about 1,700 restaurants on a 1-5 scale. One of the variables extracted from this guidebook is the ownership status of the restaurants. Until the 2016 edition, this guidebook explicitly listed chef and owner names, which enabled us to create a variable for the employment status of the chefs. Notably, we distinguish between hired chefs, owner chefs, and partially hired chefs (i.e., a chef who is considered to be a residual claimant). In the empirical analysis section, we will exploit this ownership variable.

Finally, Guidebook IV is likewise published by a local publisher in Country J. Guidebook IV has the advantage for economic analysis that it prints cuisine evaluations for a large number of high-end restaurants, about 3,000 per year. In addition, this guidebook contains separate scores for food, ambience, and service, as well as price information.

Related to our research design, all these guidebooks claim that their cuisine evaluations are provided by professional inspectors and editors. The publishers are responsible for the authenticity of the guidebook contents, and any trace of manipulation would immediately damage the reputation of the guidebook, as well as the publisher itself. From the restaurant workers' viewpoint, it would be very costly to try to manipulate the cuisine evaluations in the guidebooks, as would involve contact and negotiation with the inspectors and editors. Accordingly, following the framework proposed by Mayzlin, Dover, and Chevalier [2014], this study uses the guidebook evaluations as a basement of comparison, which will be described in the data description section below.

Next, Guidebooks I, II, III, and IV have a clear hierarchical structure, not only in search popularity, but also in number of listed restaurants. In terms of search popularity, measured by Google Trends (see Appendix), Guidebook I is the most recognized one, due to its globally well-regarded brand name, followed by Guidebook II. Guidebooks III and IV have similar search engine frequencies, but they are much less searched compared to Guidebook I and II. Thus, the popularity ranking is: Guidebook I $>$ Guidebook II $>$ Guidebook III   Guidebook IV.

On the other hand, in terms of the number of listed restaurants, which measures the degree of exclusivity, as seen in Figures XI to XIV, the ranking is: Guidebook I (counting only restaurants rated 1-3) $<$ Guidebook II $<$ Guidebook I (counting only restaurants rated 0) $<$ Guidebook III $<$ Guidebook IV.

Roughly speaking, these four guidebooks have a reversed popularity vs. exclusivity structure: The more popular guidebooks are more selective in their choice of listed restaurants. Thus, the high-end restaurant sector could be interpreted as having a hierarchical structure, in which those listed in more selective guidebooks attract consumers with a higher willingness to pay. Specifically, if a restaurant is eliminated from a high-ranking guidebook, restaurant workers may wish to engage in manipulated online reviews to make up for their lost demand.

4. Displayed Arithmetic Average Online Scores

This section summarizes how Website A (in Country J, restaurant evaluation pages) displays the review scores on the top page of a specific restaurant. Before proceeding, it should be mentioned that, throughout this (sub)section, the term "top page" means the top page of the Website A website, which lists the reviews of a specific restaurant. We provide the detailed descriptions below, as the summarized score on the top page attracts much attention among webpage viewers, including potential consumers. However, in general, little is known about page summarized scores, not just about those listed on the Website A website, but for other product/service websites. In summary, the Website A summarized score is an arithmetic average score, which is empirically shown below.

This section is organized as follows: First, we illustrate the background of the Website A website, such as the top page structure, including webpage contents to list individual reviewers' evaluation scores and comments. Second, based on the individual Website A review-level data for those restaurants listed in prestigious printed guidebooks, we deductively explain that a summarized score displayed on the top page is an arithmetic average score. Third, we illustrate the properties of an arithmetic average score, which we exploit in our theoretical and empirical analysis. Before proceeding, it should also be clearly mentioned that there are two types of scores for each restaurant: an aggregated (and summarized) score and individual consumer review scores. The former is the aggregated version of the latter, and it is this aggregated version that is displayed on the top page. Throughout this section, we empirically define the relationship between these two (aggregated and individual) interrelated scores.

First, Website A has a common framework, typical of any product review website: On the top page, it lists restaurant information, notably the display of aggregated review score, and other relevant information (such as restaurant name, address, phone number, opening hours, photos, etc.). Additionally, the top page also lists a limited number of individual consumer reviews below the aggregate display score, including their evaluation scores and comments. However, it is also common that, on the top page, a consumer cannot see all past individual

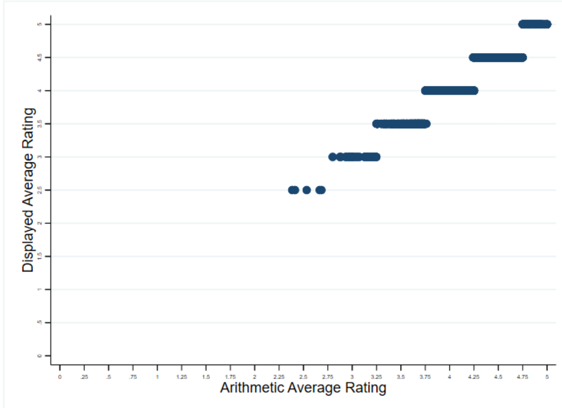

FIGURE 1. Displayed Scores (vertical) vs. Arithmetic Average Scores (horizontal)

reviews of the restaurant but instead must move to the send-top, third-top,..., and the last-page to see all of them, which involves a fair amount of time cost.

The implication of this top-page structure is that consumers, who search restaurant quality and other relevant information, are guided to pay attention to the contents appearing at the top of the top page, including the aggregated (and summarized) scores, which are typically displayed at a conspicuous position.[5]

Second, it is observed that Website A displays the rounded arithmetic average on the top page. Figure 1 empirically depicts the relationship between the top-page displayed and arithmetic average scores. The horizontal axis is the arithmetic average evaluation scores, which are calculated based on all individual customers' reviews (posted on the top, second-top, third-top, and all subsequent pages until the last page). For example, if a restaurant obtains 1,234 individual customer review scores posted on multiple pages (and each of them is scaled in a 0 to 5 review score), the horizontal axis is the arithmetic average score of 1,234 consumer reviews (such as 4.23). On the other hand, the vertical axis is an aggregated score, which is displayed prominently on the top page, which attracts much of the consumers' attention.

[5]To the best of the authors' knowledge, this structure (aggregated score display plus other relevant restaurant information [e.g., address, phone numbers, photos, link to restaurant's private website, etc.] on the top page, followed by individual reviews below) is common across many restaurant and other product review websites globally.

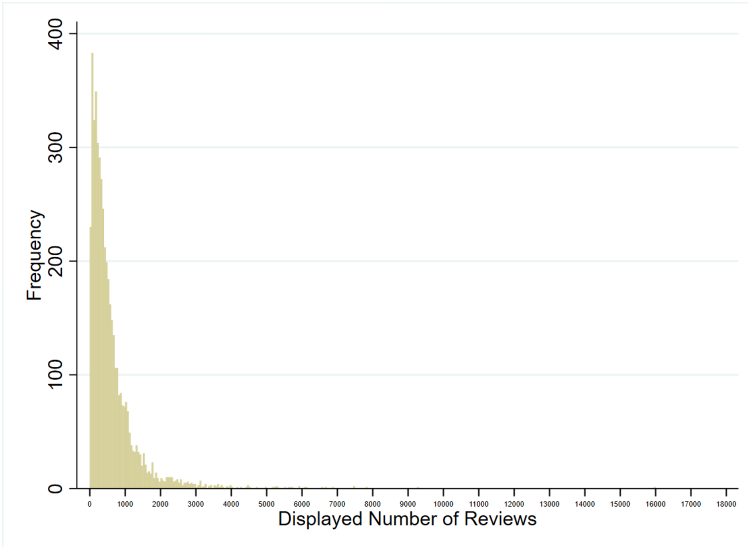

FIGURE 2. Displayed Number of Reviews

Notably, Figure 1 reveals a step function structure with the rounding rule based on 0.5 increment size (out of a 0 to 5 scale). For instance, if the arithmetic average score falls in the domain of [3.75, 4.25), the score of 4.0 is displayed on the top page.

Third, this arithmetic average score rule reveals the property of diminishing return from an additional manipulated review: The consequential change caused by an additional manipulated review is relatively large when the total number of individual reviews is small, while it diminishes as the number of individual reviews increases. This means, if restaurant workers engage in manipulated review activities, they have a better return when a restaurant is young, meaning that the number of accumulated reviews is relatively small (e.g., 50, 100, 200 reviews, etc.). However, when the restaurant already has a large number of reviews (e.g., 2,000 reviews), the additional manipulated review has relatively small influence. We theoretically and empirically take advantage of this simple property in the remainder of this article.

Forth, Figure 2 plots the distribution of the number of reviews for each guidebook-listed restaurant. As we can observe, the numbers of reviews have a long tail: Although the majority of the restaurant has less than 1000 reviews, a non-negligible proportion of guidebook-listed restaurants have more than 1,000 reviews.

5. Dynamic Model

In this section, we report the stylized model, which captures dynamic incentives for writing manipulated reviews. First, we provide explanations of model setups. Second, we derive testable model implications.

5.1. **Setup.** Consider a firm that sells one unit of product in each period $t = 1, 2, ...$ In a given period t, there are three characteristics of the product (state variables): price $p_t \geq 0$, quality $q_t \in [0, 1]$, and average review score $s_t \in [0, 1]$. We assume that price p_t depends on the last-period score:

$$p_t = \max\{0, \alpha + \beta s_{t-1}\},$$

where $\alpha \in \mathbb{R}$ and $\beta > 0$ are parameters. The parameter β captures the competitiveness of the environment. A greater β means that the firm faces fiercer competition and stands to lose more from a reduction in the average review score.

The quality q_t is given by

$$q_t = \gamma q_{t-1} + (1 - \gamma)x_t,$$

where $\gamma \in (0, 1)$ is a parameter that captures the quality decay with time, and $x_t \in [0, 1]$ is the investment effort to improve the quality.

The average review score is a simple average of past review ratings,

$$s_t = \frac{1}{t-1} \sum_{k=1}^{k-1} r_k,$$

where r_k is a review rating in period k. Each review rating r_t can take one of two values: $r_t = q_t$ if the review is honest and represents the true quality, and $r_t = 1$ if the review is fake and represents the highest possible quality. That is,

$$r_t = (1 - y_t)q_t + y_t,$$

where $y_t \in [0, 1]$ is the probability that the review in round t is fake.

The firm's present value in period t is given by

$$u_t = p_t - bx_t - cy_t + \delta u_{t+1},$$

where $b > 0$ and $c > 0$ are marginal costs associated with decisions x_t and y_t, respectively, and $\delta \in (0, 1)$ is a discount factor. The two control variables, $(x_t)_{t=1,2,\dots}$ and $(y_t)_{t=1,2,\dots}$, are chosen to solve the firm's dynamic maximization problem.

5.2. **Analysis.** Consider the first-order conditions:

$$\frac{\mathrm{d}u_t}{\mathrm{d}x_t} = -b + \delta\frac{\mathrm{d}u_{t+1}}{\mathrm{d}x_t} = -b + \frac{\delta\beta(1-\gamma)(1-y_t)}{t},$$

and

$$\frac{\mathrm{d}u_t}{\mathrm{d}y_t} = -c + \delta\frac{\mathrm{d}u_{t+1}}{\mathrm{d}y_t} = -c + \frac{\delta\beta(1-q_t)}{t} = -c + \frac{\delta\beta(1-\gamma q_{t-1} - (1-\gamma)x_t)}{t}.$$

We focus on pure-strategy solutions where y_t is either 0 or 1.

Observation 1. If $y_t = 1$, then $\frac{\mathrm{d}u_t}{\mathrm{d}x_t} = -b < 0$, so $x_t = 0$. That is, in any given period, the firm can choose to invest into the quality or into fake reviews, but never into both.

Observation 2. Irrespective of the parameters and initial conditions, for sufficiently large t both expressions in FOC become negative. So, the firm eventually stops investing into either quality of fake reviews, and lets the quality gradually deteriorate until the price becomes zero (at which point the firm shuts down).

Observation 3. The firm prefers to maximally invest into quality, $x_t = 1$, if there is a positive benefit from that:

$$\left.\frac{\mathrm{d}u_t}{\mathrm{d}x_t}\right|_{y_t=0} = -b + \frac{\delta\beta(1-\gamma)}{t} \geq 0, \tag{1}$$

and if investing into quality is preferred to investing into fake reviews:

$$\left.\frac{\mathrm{d}u_t}{\mathrm{d}y_t}\right|_{x_t=0} - \left.\frac{\mathrm{d}u_t}{\mathrm{d}x_t}\right|_{y_t=0} = b - c + \frac{\delta\beta\gamma(1-q_{t-1})}{t} \leq 0. \tag{2}$$

Note that if $x_t = 1$, then $q_t \geq q_{t-1}$, so (2) that holds in some period t will also hold in all periods $t + k$. In contrast, inequality (1) that holds in some period t will be violated after some period $t + k$.

To sum up this observation, if (1) and (2) holds in some period t, then the firm will maximally invest into quality, $x_{t'} = 1$, in periods $t' = t, t + 1, ..., t + k$ for some k, and then will stop investing altogether. This firm will never be interested in producing fake reviews.

Note that a necessary condition for the above to happen is that $b < c$, that is, the marginal cost of investment into quality is lower than the marginal cost of producing a fake review.

Observation 4. The firm prefers to invest into fake reviews, $y_t = 1$ (and $x_t = 0$) if there is a positive benefit from that:

$$\left.\frac{\mathrm{d}u_t}{\mathrm{d}y_t}\right|_{x_t=0} = -c + \frac{\delta\beta(1 - \gamma q_{t-1})}{t} \geq 0, \tag{3}$$

and if investing into fake review is preferred to investing into quality:

$$\left.\frac{\mathrm{d}u_t}{\mathrm{d}y_t}\right|_{x_t=0} - \left.\frac{\mathrm{d}u_t}{\mathrm{d}x_t}\right|_{y_t=0} = b - c + \frac{\delta\beta\gamma(1 - q_{t-1})}{t} \geq 0. \tag{4}$$

Note that if $b \geq c$, then (4) always holds. In contrast, inequality (3) that holds in some period t will be violated after some period $t + k$. To sum up this observation, if $b \geq c$ and (4) holds in some period t, then the firm will only invest into fake reviews, $y_{t'} = 1$, in periods $t' = t, t + 1, ..., t + k$ for some k, and then will stop investing altogether. This firm will never be interested in improving quality.

A more intricate case is $b < c$ and (3)–(4) hold. If $x_t = 0$, then $q_t < q_{t-1}$, so the expression $\frac{\delta\beta\gamma(1-q_{t-1})}{t}$ is generally nonmonotone. Note, however, that it is concave and eventually becomes very small, making (4) negative.

To sum up this observation, it is possible that the firm starts out with making no invesnment, $x_{t'} = 0$ and $y_{t'}$ in periods $t' = t, t + 1, ..., t + k$ for some k. During this period, the quality deteriorates, and the firm's incentive to produce fake reviews goes up. Then, between $t + k$

and $t + k'$, the firm will produce face reviews, and after that stop investing altogether. This firm will never be interested in improving quality.

so (2) that holds in some period t will also hold in all periods $t + k$. In contrast, inequality (1) that holds in some period t will be violated after some period $t + k$.

To sum up this observation, if (1) and (2) holds in some period t, then the firm will maximally invest into quality, $x_{t'} = 1$, in periods $t' = t, t + 1, ..., t + k$ for some k, and then will stop investing altogether. This firm will never be interested in producing fake reviews.

6. Empirical Specification

In this section, we illustrate our empirical linear regression specification. Following the precedent of Mayzlin, Dover, and Chevalier [2014], the dependent is the gap between the online review score (relatively low cost to manipulate) and the guidebook score (relatively costly to manipulate). In addition, the observation unit for the linear regression analyses below is an online review (technically, the observation unit is an online evaluation, which a reviewer on Website A provided for a specific guidebook-listed restaurant). As four focus of this study is economic incentives to write a self-promotional review, our emphasis is on the right-hand-side variables, which indicate the environment a restaurant faces, as well as on the gap construction of the dependent variable.

Specifically, ur econometric specification employs differenced and normalized scores, which exploit the hard-to-manipulate guidebook scores as a basement. Specifically, we have the following regression specification:

$$\begin{aligned} ORS_{i,r_i} - GRS_{i,t} = {} & \beta^{\text{Young}} D^{\text{Young}}_{i,r_i} + \beta^{\text{Middle}} D^{\text{Middle}}_{i,r_i} + \gamma^{\text{BelowThreshold}} D^{\text{BelowThreshold}}_{i,r_i} \\ & + \ \delta^{\text{Excluded}} D^{\text{Excluded}}_{i,t} + \delta^{\text{NegativeLastTen}} D^{\text{NegativeLastTen}}_{i,r_i} \\ & + \ X'_{i,t}\theta^{\text{Control}} + \alpha_i + \tau_i + u_{i,r_i}, \end{aligned} \tag{5}$$

where i is a restaurant index, r_i is the online review index for restaurant i. For instance, $r_i = 148$ indicates the 148-th review for guidebook-restaurant i. t was the guidebook year in which review r_i is written. In other words, the r_i-th online review was written for restaurant i, when the latest version of the guidebook was the year t edition.

Regarding the dependent variable, ORS_{i,r_i} is the (potentially manipulated) online review score, while $GRS_{i,t}$ is the guidebook review score. These raw scores are converted to a 0-5 scale. We calcualte the gap of these scores to eliminate unobserved heterogeneity, and the difference between these scores is the main interest of our empirical investigation.

Next, about the right-hand side varaibles, D_{i,r_i}^{Young} is the indicator variable for a young restaurant, which is defined as a restaurant with less than 200 online reviews (i.e., $r_i < 200$). Similarly, $D_{i,r_i}^{\text{Middle}}$ is the indicator for a mid-range restaurant, which has between 200 and 500 online reviews (i.e., $200 \leq r_i < 500$). The default basement is a restaurant with more than or equal to 500 reviews.

As previously defined, $AORS_{i,r_i}$ is an (arithmetic) average online review score at the time when restaurant i obtains its r_i-th online review. $D_{i,r_i}^{\text{BelowThreshold}}$ is an indicator variable which is equal to 1 when the arithmetic average of the online review scores is just below the thresholds for the next displayed score, within a 0.1 range (on the 0-5 scale). For example, this indicator is equal to 1 when restaurant i has $AORS_{i,t_i} \in [\text{Threshold} = 0.1, \text{Threshold})$.

$D_{i,t}^{\text{Excluded}}$ is an indicator varriable, which is equal to 1 when a restaurant is excluded from a relatively more prestigious guidebook (Guidebook I and II), but still listed in the relatively less-prestigious guidebooks (Guidebook III and IV).

$D_{i,r_i}^{NegativeLastTen}$ is the indicator for recently experiencing negative review(s), reflecting the fact that the online review site only publishes 10 reviews per page, in chronological order, on the restaurant's profile. This dummy variable is equal to 1 when a restaurant experienced at least one negative review, defined by a 1- or 2-point online review score, within 10 reviews preceding r_i.

$X_{i,t}$ is a vector of observables, which are extracted from the year t edition of the guidebook and other sources, including the number of chefs, restaurant capacity, cuisine category, etc. This vector also contains local variables, such as the number of guidebook-listed restaurants within a quarter-mile radius, as well as local population density, at year t.

α_i is the restaurant fixed effect, and τ_t is the guidebook-year fixed effect. The former captures the systematic time-invariant difference between the online and guidebook review scores, while the latter captures year-by-year variation in the online and guidebook restaurant scores.

Lastly, u_{i,r_i} is an idiosyncratic error term. Because our dataset contains a long series of online reviews for each restaurant, we use the HAC (Heteroskedasticity and Autocorrelation Consistent) estimator for the variance and covariance matrix.

References

Michael Anderson and Jeremy Magruder. Learning from the crowd: Regression discontinuity estimates of the effects of an online review database. *The Economic Journal*, 122(563): 957–989, 2012.

Paul Belleflamme and Martin Peitz. Inside the engine room of digital platforms: Reviews, ratings, and recommendations. *Economic Analysis of the Digital Revolution*, pages 75–114.

Paul Belleflamme and Martin Peitz. *The Economics of Platforms*. Cambridge University Press, 2021.

Sherry He, Brett Hollenbeck, and Davide Proserpio. The market for fake reviews. *Marketing Science*, 41(5):896–921, 2022.

Brett Hollenbeck, Sridhar Moorthy, and Davide Proserpio. Advertising strategy in the presence of reviews: An empirical analysis. *Marketing Science*, 38(5):793–811, 2019.

Dina Mayzlin. Promotional chat on the internet. *Marketing science*, 25(2):155–163, 2006.

Dina Mayzlin, Yaniv Dover, and Judith Chevalier. Promotional reviews: An empirical investigation of online review manipulation. *American Economic Review*, 104(8):2421–2455, 2014.

Imke Reimers and Joel Waldfogel. Digitization and pre-purchase information: the causal and welfare impacts of reviews and crowd ratings. *American Economic Review*, 111(6): 1944–71, 2021.