From Complaint to Action:

Technology-Enabled Quality Improvement from Consumer Reviews

Guangyu Cao Shenghao He Ginger Zhe Jin *

January 10, 2026

Abstract

This paper studies how the adoption of an automated review monitoring system (ARMS) enables firms to translate consumer reviews into operational quality improvements. ARMS provides automated reminders for negative reviews and facilitates back-end work-ticket management, reducing the costs of identifying, prioritizing, and addressing operational problems revealed by customer feedback. Using restaurant-level data on ARMS adoption and consumer reviews, we find that adoption leads to higher average star ratings, a lower share of negative reviews, and more positive sentiment in review text. We find no evidence that these improvements are driven by strategic manipulation of reviews. Instead, the effects are greater in operational dimensions that had been flagged by consumers prior to adoption, consistent with ARMS improving firms responsiveness to informative feedback. We further show that ARMS adoption crowds out front-end managerial responses to reviews, highlighting a substitution between operational remediation and public reply. The effects are strongest when staff exhibit a more reflective approach to handling negative reviews in the back end. Finally, ARMS adoption is associated with increased consumer engagement, as reflected in longer and more detailed reviews. Together, these findings show that online review systems, when complemented by operational technologies, can generate meaningful quality improvement rather than merely reputation management.

Keywords: Consumer Reviews, Restaurant, Quality, Automated Review Monitoring, Back-end Management, Front-end Response

JEL Codes: L86, L15, O33, D83

*Cao: School of Economics, Peking University (Email: cgy1117@pku.edu.cn); He: Guanghua School of Management, Peking University (Email: gsmhsh@pku.edu.cn); Jin: Department of Economics, University of Maryland & NBER (Email: ginger@umd.edu). We are grateful that the anonymous company that developed ARMS provided us detailed data, offered numerous technical support, and patiently answered many questions throughout our research. All errors are ours. The authors are listed in alphabetical order and contribute equally to the article. Cao and He acknowledge the support of the National Natural Science Foundation of China (Grant Nos. 72473003, 72192844, and 72203008).

1 Introduction

Consumer reviews posted on digital platforms have become indispensable to both buyers and sellers. By observing others' experiences, consumers learn about the quality of products and services (Fang, 2022; Lewis & Zervas, 2016; Moretti, 2011; Zhao et al., 2013), infer potential risks (Farronato & Zervas, 2022; Mejia et al., 2019), update their beliefs (Acemoglu et al., 2022; Shin et al., 2023; Wu et al., 2015), and ultimately make purchase decisions (Anderson & Magruder, 2012; Chevalier & Mayzlin, 2006; Luca, 2016). For sellers, consumer reviews constitute electronic word-of-mouth (eWOM) and shape their reputation that can substantially affects sales (Anenberg et al., 2022; Basuroy et al., 2003; Chevalier & Mayzlin, 2006; Hollenbeck, 2018; Luca, 2016; Zhu & Zhang, 2010). In practice, the economic importance of reviews has even induced some firms to engage in strategic and sometimes illicit reputation manipulation, such as bribing consumers to write positive reviews, posting fake negative reviews about competitors, or injecting fictitious positive feedback about their own products (Gandhi et al., 2024; He et al., 2022; Lappas et al., 2016; Luca & Zervas, 2016; Mayzlin et al., 2014; Wang et al., 2025).

Beyond the reputation mechanism that primarily affects consumer choice, consumer reviews, especially negative ones, also contain valuable information about a firm's operational shortcomings. If sellers could efficiently extract such information and respond to consumer-identified problems, they can make real quality improvement. However, this seemingly intuitive process is far from frictionless. Reviews are posted continuously over time and convey information in multiple formats, including star ratings, images, and textual content that may contain compliments, complaints, sarcasm, or narrative accounts. Dynamic monitoring and extracting actionable insights from such heterogeneous and unstructured information can be challenging. This difficulty is further exacerbated in settings such as the Chinese restaurant industry, which constitutes our empirical sample, where most establishments lack dedicated staff to monitor and analyze consumer reviews. Instead, review monitoring is typically handled sporadically by the outlet manager, often alongside routine operational responsibilities.

This paper documents how a new digital technology can transform consumer reviews into actionable insights for quality improvement. In particular, an automated review monitoring system (ARMS), developed by *Company C*, was introduced to a set of Chinese restaurants in 2019. ARMS enables managers to efficiently capture, interpret, and respond to consumer feedback in real time and at scale. The system continuously scrapes and archives reviews from *Dianping*, the largest review platform in China, and offers two core functions: (i) automated reminders for newly posted negative reviews, and (ii) a work ticket management tool that helps staff systematically address consumer complaints. By leveraging these tools, managers can convert dispersed online feedback into structured actions aimed

at improving restaurant quality.

The staggered rollout of ARMS enables a Difference-in-Differences (DID) analysis using detailed consumer reviews and work tickets collected by the system. We find that, after adopting ARMS, a restaurant's weekly average rating on *Dianping* increases significantly by 0.131 stars relative to a pre-adoption average of 4.265. The effect is even larger when measured by the decline in the proportion of negative reviews (a reduction of 0.033 from a pre-adoption mean of 0.081) or the increase in average sentiment score of review texts (an increase of 0.076 from a pre-adoption mean of 0.622). Estimates using both the conventional two-way fixed effects (TWFE) and recently developed DID estimators (Borusyak et al., 2024; Cengiz et al., 2019; Sun & Abraham, 2021), confirm that these improvements are not driven by differential pre-adoption trends between treatment and control restaurants.

A central concern in interpreting the observed quality improvements is that ARMS adoption might not have induced genuine enhancements; instead, the apparent improvements could reflect manipulative practices, such as bribing consumers to post positive reviews. To address this possibility, we examine two specific tests. First, if strategic manipulation were the main driver, 5-star reviews would be disproportionately affected. When frontline staff solicit favorable reviews offline, they almost invariably request the highest possible rating to maximize reputational benefits, with little reason to ask for 4- or 4.5-star ratings. Motivated by this, we exclude all 5-star reviews and replicate the analyses. The key results remain stable and significant, indicating that the main findings are not driven by 5-star review inflation. Second, even if consumers are persuaded to assign a high star rating on *Dianping*, they may still express dissatisfaction in the review text. Such cases, referred to as *masked negativity*, occur when a review appears positive in rating but conveys negative sentiment in its text. If strategic manipulation were predominant, we would expect a significant increase in masked negativity following ARMS adoption. However, the empirical evidence does not support this. Taken together, no evidence suggests that the observed improvement after ARMS adoption are primarily driven by strategic manipulation.

We next demonstrate that the quality improvements are robust across a range of alternative specifications, including clustering at different levels, restricting the analysis to more comparable subsamples, and controlling for restaurant demands. To explore the potential mechanisms underlying the effects of ARMS adoption, we further conduct two heterogeneity analyses. First, ARMS promotes convergence in restaurant quality by enabling worse-rated restaurants to achieve greater improvements through the effective use of consumer negative reviews. Second, examining 6 key dimensions of the dining experience: food taste, service, environment, waiting time, price, and food hygiene, we find that post-adoption improvements are more pronounced for dimensions in which restaurants performed relatively poorly prior to adoption.

The adoption of ARMS brings about an internal and operationally oriented way of handling consumer feedback that is not directly observable to the public, which we refer to as *back-end* management. We find that managerial response rates decline sharply after ARMS adoption, suggesting that back-end management partially substitutes for front-end responses. Moreover, using texts from work ticket data, we classify back-end staff attitudes to negative reviews as either *defensive* or *reflective*. Defensive attitudes attribute complaints to consumer-related factors, such as excessive expectations, whereas reflective attitudes focus on internal service or operational shortcomings. We find that ARMS generates significantly smaller improvements in restaurants with more defensive back-end staff attitudes, highlighting that the effectiveness of technology adoption depends on complementary organizational practices. Finally, restaurants with defensive back-end attitudes revealed by work tickets participate more actively in front-end responses, and they exhibit a smaller decline in managerial response rates than their reflective counterparts. This pattern suggests that front-end response behavior captures not only managerial attention to customer feedback but also a reputation-management motive aimed at shaping external perceptions, which may not necessarily reflect operational follow-ups or quality improvement.

Finally, we move beyond restaurant-level outcomes to examine whether ARMS-induced improvements correlate with changes in consumer review behavior. First, we find that consumer reviews became more multidimensional and detailed following ARMS adoption, as evidenced by an increase in the number of dimensions (out of the six) mentioned in each review and a decrease in the proportion of reviews that do not reference any specific dimension. Second, after dining in a restaurant that has adopted ARMS, consumers exerted greater effort when composing and posting reviews on *Dianping*, reflected in more frequent inclusion of pictures and longer texts. Collectively, these patterns indicate that consumers are more engaged in review writing following ARMS adoption, even if they do not observe whether a restaurant has adopted ARMS in the back end. Importantly, prior research has documented that review solicitation or incentivization can degrade the informational quality of consumer reviews (Gao et al., 2025; Khern-am-nuai et al., 2018). Therefore, the observed increase in consumer engagement provides further evidence against a manipulation-driven interpretation.

Our work contributes to several strands of the literature. First, we extend the research on the role of online consumer reviews. While the extant literature has predominantly viewed reviews as consumer-facing signals of quality that influence demand (Acemoglu et al., 2022; Anderson & Magruder, 2012; Chevalier & Mayzlin, 2006; Fang, 2022; Farronato & Zervas, 2022; Lewis & Zervas, 2016; Luca, 2016; Moretti, 2011; Shin et al., 2023; Zhao et al., 2013), we highlight their utility as supply-side feedback for operational improvement. While a subset of the literature has explored seller-side responses, these studies focus primarily on strategic manipulation, specifically, how firms may solicit inauthentic re-

views to artificially bolster their reputations (He et al., 2022; Lappas et al., 2016; Luca & Zervas, 2016; Mayzlin et al., 2014; Pocchiari et al., 2024; Wang et al., 2025). Our study diverges from this focus on "gaming" by demonstrating that sellers have strong incentives to address consumer complaints through substantive quality improvements, particularly when digital tools facilitate timely and systematic intervention.

Second, considering the enormous value of consumer reviews, the existing literature has documented three approaches to mitigate the underprovision of consumer reviews: (i) directly motivating consumers by financial incentives and social norms (Burtch et al., 2018; Cabral & Li, 2015; Chen et al., 2010; Li & Xiao, 2014; Yu et al., 2022); (ii) stimulating their desire to write reviews through appropriate platform mechanism design (Deng et al., 2022; Fradkin et al., 2021; Goes et al., 2014; Huang et al., 2019; Jiang et al., 2023); and (iii) guiding consumers to actively engage in review writing through managerial responses (Ananthakrishnan et al., 2023; Chen et al., 2019; Chevalier et al., 2018; Proserpio & Zervas, 2017; Yang et al., 2019). We add to this literature by providing evidence that proactive managerial efforts to make real improvement can also encourage consumers to engage and provide more informative reviews.

A couple of recent studies convey a similar insight that platform sellers utilize online consumer reviews to improve their products and services (e.g., Ananthakrishnan et al., 2023; Karanam et al., 2025). Our work differs from theirs in two key ways. First, whereas the existing literature primarily studies the front-end handling of consumer reviews, such as whether and how managers make public responses on the platform, we are the first to shed light on back-end management of consumer reviews, especially how consumer reviews are translated into operational follow-up actions to enhance quality improvement, and document the correlation between front- and back-end managerial actions. Second, we highlight the role of digital technology in enabling improvements through unstructured, high-volume online consumer reviews. Intuitively, sellers can learn from consumer reviews, but how they convert rapidly growing textual feedback into concrete operational practices and how this process changes following a major technology adoption remain largely underexplored. This paper seeks to fill this gap.

The remainder of this paper is organized as follows. Section 2 introduces the research background. Section 3 describes the data, variables, and sample construction. Section 4 presents the empirical setting. Section 5 reports the main empirical findings of the impact of ARMS adoption. Section 6 explores the interaction between front-end managerial responses and back-end operational improvement, as well as ripple effects on consumer review writing. Section 7 derives managerial implications, and Section 8 concludes.

2 Research Background

2.1 Dianping and Consumer Reviews

Many countries host their own dominant consumer review platforms. Yelp plays this role in the United States, while in China the corresponding platform is *Dianping*. Founded in 2003, *Dianping* merged with another major review platform, Meituan, in 2015 and has since become the largest and effectively the sole nationwide consumer review platform in China. By 2024, *Dianping* reported coverage of more than 8 million local businesses across virtually all major cities and thousands of smaller cities, and had accumulated 268 million genuine consumer reviews. In China, *Dianping* is deeply embedded in consumers' dining decisions. When planning to eat out, a typical consumer opens the *Dianping* app to search for restaurants, browses ratings and reviews to choose where to dine, and may subsequently leave her own review. As the largest repository of Chinese review texts worldwide, *Dianping* reviews capture consumers' perceptions of restaurants, including complaints, criticism, praise, and expectations, and therefore constitute a valuable source of information for restaurant operators.

Figure 1 displays an illustrative consumer review posted on *Dianping*. To submit a review, a user must provide a star rating on a 5-star scale, while providing a textual comment and attaching photographs are optional. In the sample analyzed in this paper, 97% of individual reviews contain both a star rating and text, while photographs are attached to 79% of individual reviews. Accordingly, the analysis below focuses primarily on variables constructed from star ratings and textual reviews. Importantly, there is no binding constraint linking the star rating to the content of the textual review. As a result, a reviewer may assign the maximum 5-star rating while expressing strong criticism in the accompanying text. This phenomenon, referred to as *masked negativity*, will be employed in the construction of our empirical tests.

2.2 Automated Review Monitoring System

The Automated Review Monitoring System (ARMS) developed by *Company C*[1] operates upon explicit authorization from business clients, who are usually the headquarter office of a national or regional restaurant chain. Once authorization is granted, ARMS accesses consumer reviews on *Dianping* of all the restaurant outlets owned and authorized by the client and stores all retrieved reviews in a private data warehouse. The collected reviews consist of two components. First, ARMS retrospectively retrieves and archives the entire stock of reviews that existed prior to authorization. Second, from the time of authorization, ARMS continuously monitors and captures newly posted reviews, which are scraped and stored on an hourly basis.

[1] *Company C* operates independently of *Dianping*; accordingly, ARMS is a third-party software system that is entirely independent of *Dianping*. At the company's request, we anonymize its identity and refer to it as *Company C* throughout the paper.

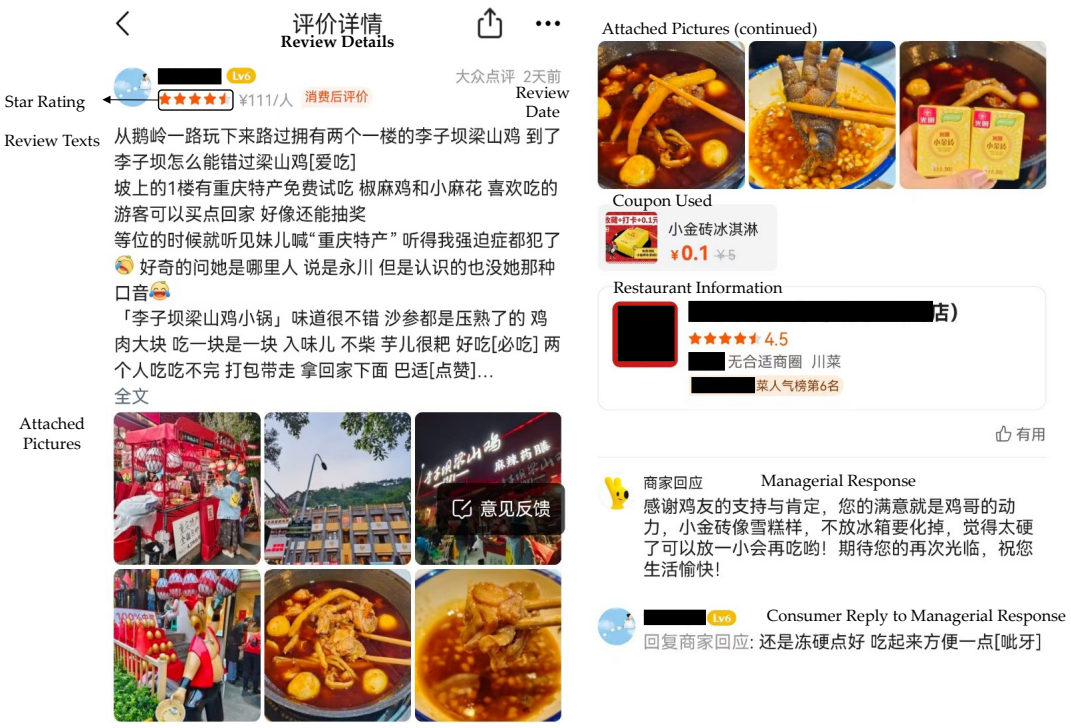


Figure 1. *An Illustrative Consumer Review on Dianping*

Based on the collected reviews, ARMS provides two core functions. First, it generates automated reminders for newly posted negative reviews, which are identified using both star ratings and textual analysis enabled by machine learning techniques. Second, it offers a work ticket management tool that facilitates restaurant staff in handling complaints raised by consumers. We elaborate on each function below.

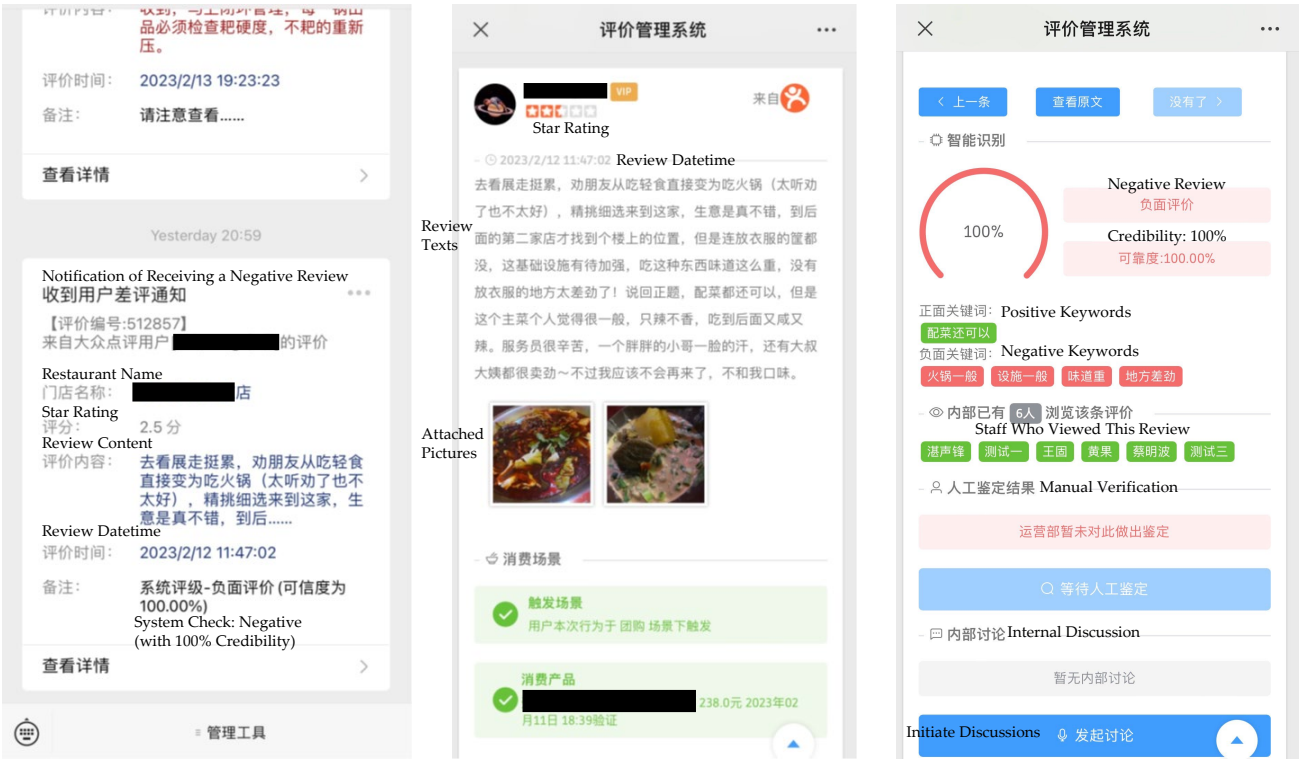
2.2.1 Automated Reminders for Negative Reviews

For each newly posted review, ARMS conducts an immediate assessment and sends a notification to the restaurant manager and staff if either of two conditions is satisfied. First, the review's star rating is below three stars. Second, the review text is classified as negative by a textual analysis algorithm developed by *Company C*.

This function addresses two practical constraints faced by restaurants. First, automatic alerts eliminate the need for manual and continuous monitoring of reviews on *Dianping*. Second, it enables systematic analysis of review text. While star ratings are easy to interpret, reading and evaluating textual reviews is time-consuming and subject to individual judgment. ARMS provides a scalable and consistent approach to identifying text-based, rather than purely star-based, negative feedback.

Figure 2 Panel (a) illustrates the interface displayed when restaurant staff receive a notification of a negative review. These notifications are pushed directly to staff members' personal mobile phones via WeChat, the dominant messaging application in China. Because WeChat is the primary platform for

daily communication and is integrated with multiple functionalities, such as electronic payments and QR code scanning, such notifications are highly salient and difficult to overlook. Upon tapping the notification, restaurant staff are directed to the interface corresponding to the second function, namely work ticket management.



(a) (b) (c)

Figure 2. *The ARMS Interface Displayed to Restaurant Staff*

2.2.2 Work Ticket Management

After a negative review is identified, ARMS not only sends an automated reminder but also initiates a work ticket in WeChat. Restaurant staff can first review the flagged content and verify whether the review is informative. If the review contains no actionable information, such as a 1-star rating without text, or if it was incorrectly classified as negative by the algorithm, the staff member may close the ticket without further action. If the review passes this manual verification, staff can initiate an internal discussion within the work ticket, and managers may assign specific tasks, such as re-cleaning a table, to designated employees. The restaurant manager can ultimately close the ticket after confirming that the issue has been addressed. For an overview, we draw an illustrative flowchart in Figure 3, which describes the whole workflow of ARMS.

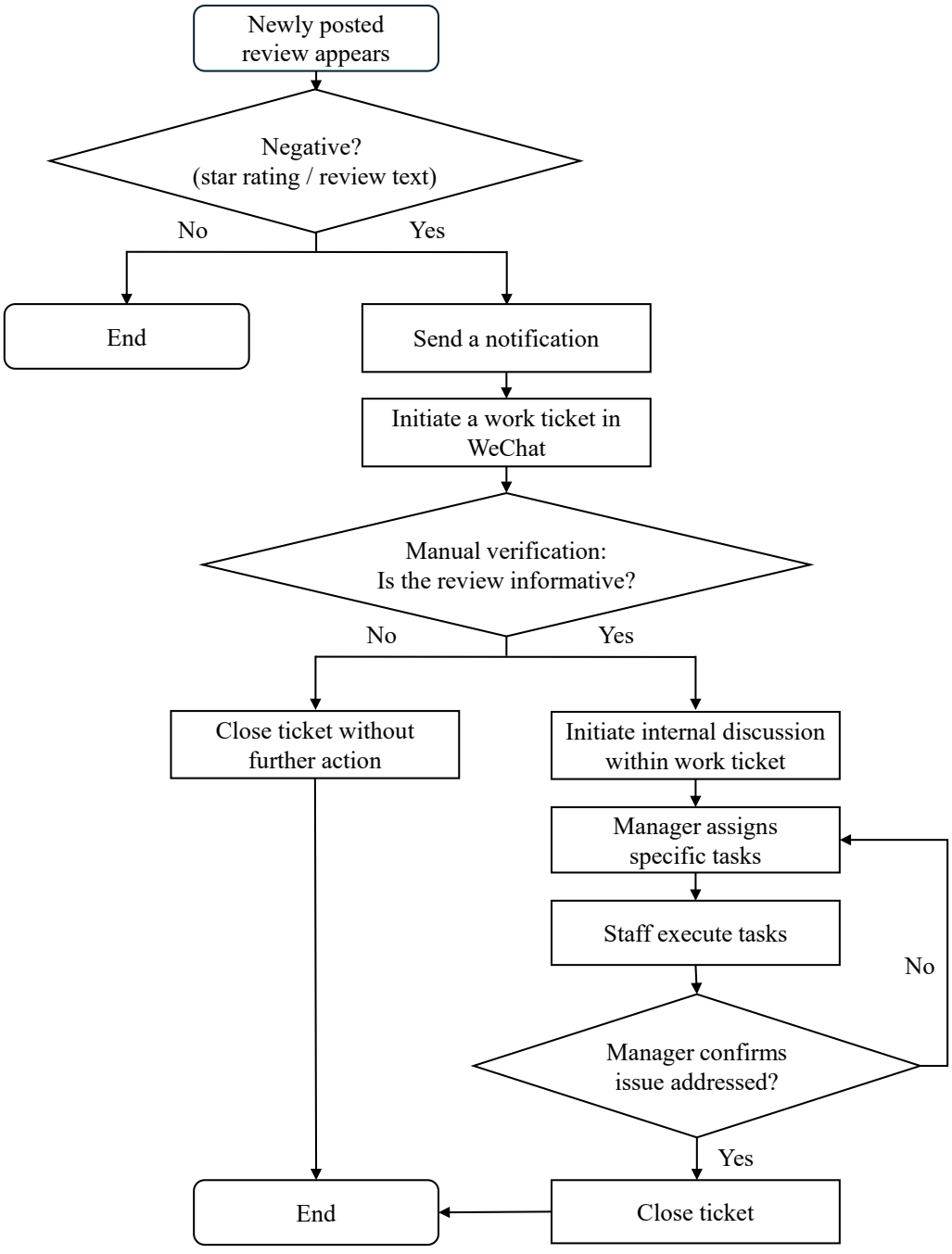


Figure 3. *An Illustrative Flowchart of ARMS*

2.2.3 The Process of ARMS Installation

Before turning to the data details, it is useful to clarify the ARMS installation process, as it determines data availability and sample composition. The collaboration between an ARMS-adopting firm and *Company C* involves two distinct milestones: authorization and adoption. Authorization refers to the firm granting *Company C* merchant API access on *Dianping*, which allows ARMS to retrieve all historical reviews and continuously capture newly posted reviews for the firm's restaurants thereafter. Importantly, authorization alone does not constitute a treatment. Granting data access to *Company C* does not alter restaurant operations or generate any system outputs.

Adoption, by contrast, is the treatment exploited in this study. Only when a restaurant adopts ARMS under a subscription plan do the system's functionalities become active. From that point onward, ARMS generates notifications for newly posted negative reviews and initiates corresponding work tickets. Reviews posted prior to adoption remain stored as historical records in the private data warehouse but do not trigger notifications or work tickets. For the firms that authorized API access but never adopted ARMS, *Company C* continues to retrieve and dynamically capture their reviews on a rolling basis, as the authorization remains valid throughout the sample period. These firms constitute the control group in the empirical analysis, for which review data are available but the treatment, namely ARMS adoption, is absent.

It is worth noting that ARMS adoption is a top-down decision inside each client. Usually, the headquarter of a restaurant chain decides whether to adopt ARMS, and if the answer is yes, whether to adopt the work ticket function in addition to the automated alert function. These decisions apply to all restaurant outlets governed by the headquarter. In this sense, ARMS adoption is not driven by the preference of individual restaurant managers but they must deal with whatever ARMS functions adopted by the headquarter.

One may argue that the chains that eventually adopt ARMS place more managerial emphasis on consumer reviews, are more concerned of problems revealed by consumer reviews, or have experienced more negative reviews prior to adoption. This potential selection bias is partially addressed by the fact that our control group consists of only the chains that have authorized *Company C* to access their review data, which in itself reflects the headquarter office's attention to and concern of consumer reviews. As detailed below, besides restaurant fixed effects, our empirical analysis controls for restaurant characteristics such as restaurant location and cuisine type, and allows differential trend or seasonality by these observables. We also test the parallel trends assumption between treatment and control groups and present a battery of robustness checks, in order to minimize the potential selection bias in ARMS adoption.

3 Data and Sample

We compile three main datasets for the empirical analysis in this paper. The first two are provided by *Company C* and include the universe of (i) consumer reviews posted on *Dianping* and (ii) records of work tickets initiated within ARMS. In addition, we collect a rich set of restaurant-level attributes to account for heterogeneity across establishments. We describe these datasets in detail below.

3.1 Consumer Reviews

As shown in Figure 1 and discussed in Section 2.1, for each consumer review posted on *Dianping*, we extract (i) the star rating, (ii) the full review text, (iii) a dummy variable indicating the presence of attached photographs, and (iv) a dummy variable indicating the presence of managerial response. Using both the quantitative and qualitative information contained in consumer reviews, we construct a set of measures of restaurant quality.

The simplest and most direct measure is the star rating itself. Star ratings range from 0.5 to 5 in increments of 0.5. Although imperfect, this measure provides a direct and widely used indicator of consumers' evaluations of restaurant quality. In addition, we construct a dummy variable equal to one if the star rating is less than 3, a threshold commonly used in the restaurant industry to classify "negative reviews." ARMS adopts the same rule, and reviews with ratings less than 3 stars automatically trigger a reminder, as described in Section 2.2.1.

In contrast to the quantitative information conveyed by star ratings, qualitative information contained in review texts typically provides more diagnostic and granular, yet less readily interpretable, signals (Archak et al., 2011; Büschken & Allenby, 2016; Chakraborty et al., 2022; Cho et al., 2022). To quantify this information, we apply a sentiment classification algorithm. The algorithm is a pre-trained model fine-tuned on Chinese restaurant reviews on *Dianping*.[2] The raw outputs of the model are logits, which we transform into probability scores using the softmax function, following standard practice. This procedure yields a sentiment score for each review, taking values between 0 and 1 and representing the probability that the review is positive. This text-based measure captures the overall tone of consumer narratives and provides a richer representation of consumer feedback beyond numeric star ratings.

In addition to a measure capturing overall restaurant quality, we further use review texts to construct measures of quality along six distinct dimensions. Based on consultations with experts in the restaurant industry and product managers from *Company C*, we identify six well-established dimensions that capture key aspects of the dining experience: food taste, service, environment, waiting time, price, and food hygiene. For each review, we employ the DeepSeek-V3 large language model to iden-

[2]The model is available at https://huggingface.co/uer/roberta-base-finetuned-dianping-chinese (accessed on September 15, 2025).

tify whether each dimension is mentioned and, if so, to classify the associated sentiment as positive or negative. This procedure allows us to derive dimension-specific measures of restaurant quality from review texts, complementing the aggregate evaluation.

It is worth noting that *Company C* uses its own machine-learning algorithm to gauge the sentiment of review text, incorporate it in a credibility score, and report the credibility score along with the full content of a negative review when it produces an automatic alert. In this process, it does not specify which dimension(s) the review focuses on, as *Company C* believes that the raw content of the highlighted review is more informative than a crude classification driven by algorithms.

Finally, we construct several additional review-based measures to facilitate subsequent analyses. We compute the managerial response rate to measure front-end managerial actions. Moreover, we construct the following variables to reflect consumer behaviors in review writing: (i) the average number of dimensions (out of the six) mentioned in a review, (ii) the proportion of reviews that do not mention any specific dimension, which we refer to as *topicless*, (iii) the proportion of picture-attached reviews, and (iv) the logarithm of average review length.

3.2 Work Ticket Data

The second dataset derived from ARMS consists of records of work tickets. As illustrated by the workflow in Figure 3, when a work ticket is triggered by a negative review, restaurant managers and staff can initiate internal discussions to address the issue. These discussions take place within the ARMS interface, and the associated communication texts are recorded and stored by the system.

Based on our own reading of the discussion texts and interviews with restaurant staff, we find that staff members adopt two distinct attribution and handling strategies when confronted with negative reviews. Some exhibit a *defensive* attitude, primarily attributing complaints to consumer-related factors such as excessive expectations or misunderstandings. Others exhibit a *reflective* attitude, focusing on their own service or operational shortcomings. We further employ the DeepSeek-V3 large language model to classify each discussion conversation into one of these two categories. This classification is used to examine how staff attitudes shape the impact of ARMS adoption, even though ARMS itself is a common and neutral tool.

3.3 Restaurant characteristics

The final dataset is self-constructed and contains a set of cross-sectional variables used to account for potential heterogeneity across restaurants. Specifically, we use ArcGIS to process raster data and point-of-interest (POI) data to compute (i) population density within a three-kilometer radius of each restaurant, (ii) nighttime light intensity within the same area, (iii) the distance from the restaurant to the city center, and (iv) the distance to the nearest subway station. All these variables are measured

in 2018 and therefore predate the sample period, which helps avoid confounding effects arising from ARMS adoption. In addition, following suggestions from staff at *Company C*, we construct a dummy variable indicating whether a restaurant specializes in hotpot, a highly distinctive and culturally salient dining format in Chinese cuisine. In the analysis that follows, we collectively refer to these variables as "restaurant attributes" or "restaurant characteristics."

3.4 Sample Construction

Our sample consists of 122 restaurants, of which 104 ultimately adopted ARMS, while the remaining 18 authorized *Company C* to access their review data but did not adopt ARMS. These two groups serve as the treatment and control groups, respectively, in the empirical analysis. Among the 104 restaurants in the treatment group, all activated the automated reminder function, while only 81 also activated the work ticket management tool. Accordingly, analyses that rely on work ticket data are conducted on a reduced sample of 99 (= 81 + 18) restaurants.

Unless otherwise stated, the baseline unit of analysis is restaurant-week. This aggregation smooths day-to-day volatility and day-of-week seasonality while retaining sufficient temporal granularity. Specifically, we first construct variables at the level of individual consumer reviews and then average them at the restaurant-week level. Table 1 reports summary statistics for the main variables used in the paper.

Restaurants in the treatment group have adopted ARMS in four waves: March 2019, February 2023, March 2023, and April 2023. *Company C* has granted us access to data over 24 weeks (roughly equivalent to half a year) before and after the adoption timing. For restaurants in the control group, we have obtained access to consumer review data in all calendar weeks that overlap with any treatment restaurant. That is, the time horizon for the control group observations is the union of time spans of the treatment group sample.

Table 1. *Summary Statistics*

Variable	Data granularity	Obs.	Mean	Std. dev.	Min.	Max.
Quality Measures:						
Average Star Rating	Restaurant-week	5,723	4.366	0.568	0.5	5
Proportion of Negative Reviews	Restaurant-week	5,723	0.065	0.137	0	1
Average Sentiment Score	Restaurant-week	5,612	0.700	0.200	0.004	0.996
Restaurant-specific Characteristics:						
Nearby Population Density in 2018 (persons/km^2)	Restaurant	122	15049.77	11519.45	280.43	56138.98
Nearby Nighttime Light Intensity in 2018	Restaurant	122	29.536	10.267	7.169	63.826
Distance to the Nearest Subway Station (m)	Restaurant	122	904.42	1172.56	30	3660
Distance to City Center (km)	Restaurant	122	8.070	7.495	0.115	33.991
Hotpot Dummy	Restaurant	122	0.475	0.501	0	1
Additional Review-based Measures:						
Managerial Response Rate	Restaurant-week	5,723	0.731	0.391	0	1
Average Number of Dimensions Mentioned	Restaurant-week	5,723	2.286	0.739	0	6
Proportion of Topicless Reviews	Restaurant-week	5,723	0.101	0.198	0	1
Proportion of Picture-attached Reviews	Restaurant-week	5,723	0.569	0.338	0	1
Average Length of Reviews (in logs)	Restaurant-week	5,612	4.396	0.557	0	6.380

Notes: This table reports summary statistics of the main variables. Measures derived from the review data are aggregated at the restaurant-week level. For restaurant-weeks consisting solely of empty-text reviews, it is not possible to define the weekly average sentiment score or review length. "Nearby" is defined to be within three-kilometer radius of the restaurant. Geographical characteristics are calculated based on geocoding the restaurants by authors. The hotpot dummy is collected from the homepage of each restaurant on *Dianping*. The raster data used to calculate population density and nighttime light intensity are from WorldPop and Chen et al. (2021), respectively.

4 Empirical Setting

4.1 Baseline Specification

To identify the effects of ARMS adoption on restaurant quality, we exploit the staggered rollout of ARMS and implement a Difference-in-Differences (DID) design. The baseline specification is given by:

$$Y_{it} = \alpha_i + \lambda_t + \beta ARMS_{it} + \Gamma(X_i \cdot f(t)) + \mu G_i \cdot t + \varepsilon_{it}, \tag{1}$$

where i and t index restaurants and calendar weeks, respectively. Y_{it} denotes measures of restaurant quality for restaurant i in week t, constructed as described in Section 3.1, including weekly averages of star ratings, the proportion of negative reviews, and sentiment scores. Restaurant and week fixed effects are denoted by α_i and λ_t respectively. The indicator variable $ARMS_{it}$ equals one if restaurant i has adopted ARMS by week t.

X_i represents predetermined restaurant characteristics, which are interacted with $f(t)$ to flexibly control for differential time trends across restaurants that may vary systematically with these characteristics. The function $f(t)$ is specified either as a third-order polynomial in linear time or as a full set of week fixed effects λ_t. G_i is a dummy variable indicating restaurants in the treatment group, which is interacted with a linear time trend to allow for group-specific trends. ε_{it} is an error term. The coefficient of interest, β, captures the effect of ARMS adoption on restaurant quality. Standard errors are clustered at the restaurant level to account for potential serial correlation.

4.2 Relative Time Model

The baseline DID model specified in Equation (1) relies on the parallel trends assumption, namely that, absent ARMS adoption, restaurants in the treatment and control groups would have followed similar outcome trends. To assess the plausibility of this assumption, we estimate a relative time model, in which the adoption indicator $ARMS_{it}$ is decomposed into a set of indicators for time windows relative to the adoption date. Specifically, we estimate the following model:

$$Y_{it} = \alpha_i + \lambda_t + \sum_j \tau_j Pre_{it}^j + \sum_k \omega_k Post_{it}^k + \Gamma(X_i \cdot f(t)) + \mu G_i \cdot t + \varepsilon_{it}. \tag{2}$$

Pre_{it}^j is an indicator variable that equals one if calendar week t is j weeks prior to the ARMS adoption week for restaurant i. Similarly, $Post_{it}^k$ equals one if week t is k weeks after ARMS adoption for restaurant i. All other variables are defined as in Equation (1). We bin all periods earlier than 11 weeks before adoption into a single indicator, and likewise collapse all periods beyond 11 weeks after adoption into one indicator. For model identification, we take $j = -1$ as the reference period and omit it from the estimation. Accordingly, the estimated pretreatment coefficients $\{\tau_j\}_{j=-2}^{-11}$ allow us to assess

whether treatment and control restaurants exhibited parallel trends prior to ARMS adoption, while the post-adoption coefficients $\{\omega_k\}_{k=1}^{11}$ capture the dynamic effects of ARMS adoption over time.

Recent advances in the econometric literature have pointed out that the TWFE-OLS estimator implied by Equation (2) may be biased in settings with staggered treatment adoption due to the negative weighting problem (Goodman-Bacon, 2021). To address this concern, we additionally estimate the relative time model using several recently developed DID estimators that are robust to such biases (Borusyak et al., 2024; Cengiz et al., 2019; Sun & Abraham, 2021).[3]

5 ARMS Adoption and Restaurant Quality Improvement

In this section, we present empirical evidence on how the adoption of ARMS affects restaurant quality. We begin by reporting estimates from the baseline specification in Equation (1), using both rating-based and text-based measures constructed from consumer reviews as outcome variables. We then turn to the relative time model specification to assess the plausibility of the parallel trends assumption. Next, we conduct a series of empirical tests to address the concern that the observed improvements may be driven by strategic manipulation, such as bribing consumers to post favorable reviews. We proceed with a set of robustness checks, including alternative clustering and specifications that restrict the sample to more comparable restaurants and control for restaurant demands. Finally, we conclude this section with a series of heterogeneity analyses to shed light on the potential mechanisms underlying the effects of ARMS adoption.

5.1 Quality Improvement

We begin by using the most straightforward measure of restaurant quality, the average star rating, as the outcome variable to estimate Equation (1). Table 2 reports the corresponding results. We progressively add controls across specifications to demonstrate the robustness of the estimated effects. Column (1) includes only restaurant and week fixed effects. Column (2) further adds a linear time trend specific to restaurants in the treatment group. Column (3) additionally controls for interactions between predetermined restaurant characteristics and third-order polynomial time trends. Finally, Column (4) replaces the time trends in these interactions with a full set of week dummies, thereby allowing for the most flexible form of differential time variation associated with restaurant attributes.

Across all specifications, we find a stable and statistically significant improvement in star ratings following the adoption of ARMS. The estimated coefficient β ranges from 0.101 to 0.191, corresponding to a 2.3% – 4.4% increase in weekly average star ratings relative to the sample mean of 4.366. Our preferred specification, Column (4), suggests that ARMS adoption leads to an approximately 3.0%

[3]When estimating the relative time model, the time trend component $f(t)$ is restricted to a third-order polynomial in linear time to accommodate the estimator proposed by Borusyak et al. (2024), as the corresponding Stata package `did_imputation` does not accommodate more flexible interactions such as $X_i \times \lambda_t$.

Table 2. *Baseline Results*

Dep. Var.	Average Star Rating			
	(1)	(2)	(3)	(4)
ARMS	0.101**	0.190***	0.191***	0.131**
	(0.046)	(0.042)	(0.042)	(0.052)
Dep. Var. Mean	4.366	4.366	4.366	4.366
Week FE	YES	YES	YES	YES
Restaurant FE	YES	YES	YES	YES
Treatment Linear Trend		YES	YES	YES
Covariates $\times f(t)$			YES	
Covariates $\times \lambda_t$				YES
Observations	5723	5723	5723	5723
Clusters	122	122	122	122

Notes: This table reports the baseline results. The dependent variable is weekly average star rating of restaurants. All regressions control for week and restaurant fixed effects. Columns (2), (3) and (4) include treatment-specific linear time trends to control for different time trends between treatment and control restaurants. Column (3) includes interactions of restaurant characteristics with a third-order polynomial of time trend, and Column (4) includes interactions of restaurant characteristics with week dummies, to absorb potential discrepancies due to restaurant traits. Standard errors are in parentheses and clustered at the restaurant level. *, ** and *** indicate statistical significance at 10%, 5% and 1% level, respectively.

increase in star ratings.

It is important to note that this improvement does not translate one-for-one into the star ratings displayed on *Dianping*. The estimated effect is based on newly posted reviews, whereas the displayed rating for each restaurant aggregates all historical reviews. Nonetheless, the results indicate a systematic upward shift in the flow of incoming consumer evaluations. Over a sufficiently long horizon, such improvements in newly posted reviews are likely to accumulate and eventually be reflected in the platform's displayed star ratings.

The preceding results provide preliminary evidence of a positive impact of ARMS adoption on restaurant quality. However, as emphasized in the existing literature, relying solely on star ratings as outcome measures may be problematic, as they can be subject to rating inflation and strategic distortions (Aziz et al., 2023; Bolton et al., 2019; Filippas et al., 2022; Fradkin et al., 2021; Horton & Golden, 2015; Tadelis, 2016). In light of this concern, we complement the analysis by employing alternative measures of restaurant quality.

First, we construct the proportion of negative reviews at the restaurant-week level, where negative reviews are defined as those with star ratings below 3. This measure more directly captures highly unsatisfactory consumer experiences. As reported in Column (1) of Table 3, the adoption of ARMS is associated with a statistically significant decline in the incidence of negative reviews. The estimated reduction of 3.3 percentage points is economically meaningful relative to the sample mean of 6.5%,

implying that ARMS adoption reduces the prevalence of negative reviews by roughly one half. This pronounced effect is consistent with the design of ARMS, which primarily operates through automated alerts triggered by negative reviews. By drawing managerial attention to service failures in a timely manner, ARMS is particularly effective at mitigating the occurrence of severe dissatisfaction among consumers.

Table 3. *ARMS Impact on Alternative Measures of Quality*

Dep. Var.	Proportion of Negative Reviews	Average Sentiment Score
	(1)	(2)
ARMS	-0.033**	0.076***
	(0.013)	(0.019)
Dep. Var. Mean	0.065	0.700
Week FE	YES	YES
Restaurant FE	YES	YES
Treatment Linear Trend	YES	YES
Covariates $\times \lambda_t$	YES	YES
Observations	5723	5612
Clusters	122	122

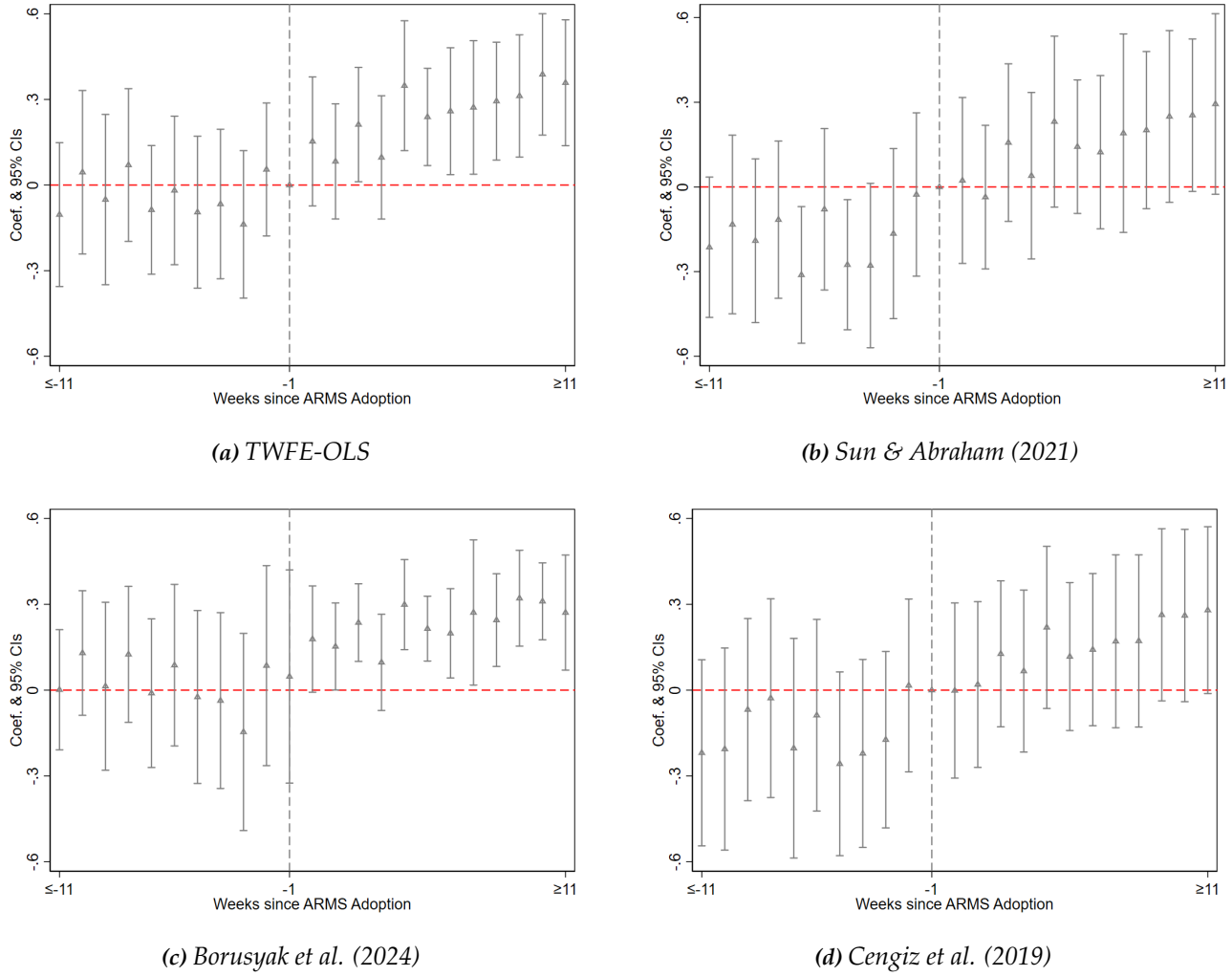
Notes: This table shows quality improvements following the adoption of ARMS using alternative measures. Column (1) replaces the dependent variable with the proportion of negative reviews, and Column (2) focuses on weekly average sentiment score. Standard errors are in parentheses and clustered at the restaurant level. *, ** and *** indicate statistical significance at 10%, 5% and 1% level, respectively.

Second, we employ an additional quality measure derived from review texts, namely the sentiment score. As reported in Column (2) of Table 3, the weekly average sentiment score increases by 0.076 following the adoption of ARMS. This result indicates that, when restaurant quality is evaluated based on the textual content of consumer reviews rather than numeric ratings alone, the overall tone of consumer feedback becomes systematically more favorable. Taken together, these findings suggest that ARMS contributes to genuine improvements in the perceived quality of treatment restaurants, rather than merely inducing mechanical changes in star ratings.

5.2 Tests of Parallel Trends Assumption

Next, we estimate the relative time model specified in Equation (2) to assess the plausibility of the parallel trends assumption. Figure 4 plots the estimated pre- and post-adoption coefficients, together with 95% confidence intervals, using star ratings as the outcome variable. Panels (a) – (d) correspond to the TWFE estimator and the alternative estimators proposed by Sun and Abraham (2021), Borusyak et al. (2024), and Cengiz et al. (2019), respectively. Although the estimates differ somewhat across methods in terms of magnitude and statistical significance, the results in Figure 4 consistently indicate that prior to ARMS adoption, there are no discernible differences in star rating trends between restaurants in the

treatment and control groups. By contrast, following adoption, the positive effects begin to emerge approximately one month (four weeks) after ARMS is implemented, suggesting that the system requires time to take effect. Over a longer horizon, these improvements persist, pointing to a sustained impact of ARMS. Figures A1 and A2 report analogous results when using the proportion of negative reviews and the sentiment score as outcome variables, respectively, revealing similar patterns in both pre-adoption trends and post-adoption dynamics.



(a) *TWFE-OLS*

(b) *Sun & Abraham (2021)*

(c) *Borusyak et al. (2024)*

(d) *Cengiz et al. (2019)*

Figure 4. *Dynamic Effect of ARMS on Star Rating*

Notes: This figure portrays estimates from the relative time model as specified in Equation (2). Panels (a) – (d) correspond to the TWFE estimator and the alternative estimators proposed by Sun and Abraham (2021), Borusyak et al. (2024), and Cengiz et al. (2019), respectively. The dependent variable is weekly average star rating. The x-axis represents the time period relative to ARMS adoption. Points represent coefficient estimates and vertical bars show 95% confidence intervals.

5.3 Addressing Concerns about Strategic Manipulation

So far in this section, we have shown that following the adoption of ARMS, restaurants experienced significant improvements across multiple measures, including star ratings, the proportion of nega-

tive reviews, and text-based sentiment scores. Taken together, these results suggest that ARMS helps restaurant staff more effectively process information contained in negative reviews and translate consumer feedback into genuine improvements in product and service quality.

However, a key threat to this interpretation, or an alternative explanation for the observed patterns, is strategic manipulation. Specifically, the installation of ARMS may not have led to real quality improvement; instead, the apparent improvement could reflect manipulative practices, such as bribing consumers to post positive reviews or otherwise distorting review content (He et al., 2022; Lappas et al., 2016; Luca & Zervas, 2016; Mayzlin et al., 2014; Wang et al., 2025). If this were the case, the implications of our empirical findings would be fundamentally different. In this section, we present a series of empirical tests that, while not fully conclusive, provide evidence against the hypothesis that the observed improvements are primarily driven by strategic manipulation.

Excluding 5-Star Reviews. Five-star reviews warrant particular attention because, if strategic manipulation is indeed present, it should disproportionately manifest in this rating category. The intuition is straightforward. When frontline staff solicit favorable reviews from customers through offline persuasion, they almost invariably request the highest possible rating in order to maximize the reputational benefits. There is little incentive for staff to ask for a 4- or 4.5-star rating, as such ratings do not fully serve the purpose of inflating visible reputation metrics.

Motivated by this observation, we conduct a test to mitigate the potential influence of strategic manipulation by excluding all 5-star reviews from the sample. We then recompute the average star rating, the proportion of negative reviews, and the text-based sentiment score, and replicate the main analyses using these adjusted measures. If the estimated effects of ARMS adoption persist after removing 5-star reviews, this would suggest that the observed quality improvements are unlikely to be driven primarily by strategic manipulation. Consistent with our expectation, Table 4 presents the corresponding results, which show that our findings remain robust even when 5-star reviews are excluded.

Masked Negativity in Consumer Reviews. Compared with manipulating star ratings, altering review text is substantially more costly, so the textual sentiment score provides a more credible reflection of consumers' authentic perceptions, which we have already employed as an outcome variable. This feature also offers a way to test for the potential influence of strategic manipulation. Even if consumers are persuaded to give a high star rating on *Dianping* and actually agree to do so to avoid offending restaurant staff, they may still express dissatisfaction in the review text. When this occurs, it results in *masked negativity*, defined as cases in which a review appears positive in star rating but conveys negative sentiment in the text. If the observed improvements were primarily driven by strategic manipulation, we would expect to see a substantial increase in masked negativity after ARMS adoption.

Table 4. *Analysis without 5-Star Reviews*

Dep. Var.	Average Star Rating	Proportion of Negative Reviews	Average Sentiment Score
	(1)	(2)	(3)
ARMS	0.201***	-0.067***	0.104***
	(0.073)	(0.025)	(0.026)
Dep. Var. Mean	3.708	0.126	0.569
Week FE	YES	YES	YES
Restaurant FE	YES	YES	YES
Treatment Linear Trend	YES	YES	YES
Covariates $\times \lambda_t$	YES	YES	YES
Observations	5348	5348	5264
Clusters	122	122	122

Notes: This table reports the results of analysis without 5-star reviews. Restaurant-weeks consisting entirely of 5-star reviews are thus excluded from the sample. Standard errors are in parentheses and clustered at the restaurant level. *, ** and *** indicate statistical significance at 10%, 5% and 1% level, respectively.

Concretely, we define *masked negativity* as reviews that exhibit a star rating above a specified threshold but a sentiment score below a given threshold. For star ratings, we consider both 5-star (corresponding to a "perfect experience" by *Dianping*'s definition) and 4-star (corresponding to a "good experience") as thresholds. For sentiment scores, we experiment with 0.3 and 0.7 as cutoffs. Under each combination of thresholds (i.e., star rating = 5 & sentiment score < 0.3, star rating $\geq$ 4 & sentiment score < 0.3, star rating = 5 & sentiment score < 0.7, star rating $\geq$ 4 & sentiment score < 0.7), we identify reviews meeting the criteria for masked negativity and compute their proportion at the restaurant-week level.

Using this measure, we estimate the impact of ARMS adoption on the proportion of masked negativity. As reported in Table 5, we do not find any significant increase in masked negativity following ARMS adoption. The only exception is Column (2), where the coefficient is significantly negative rather than positive, which contradicts the hypothesis that strategic manipulation drives our previous findings. Overall, these results provide no evidence that ARMS has induced manipulation of consumer reviews in the form of masked negativity.

5.4 Robustness Checks

We have also conducted a series of robustness checks, with results reported in Table A1. First, to account for potential intra-firm correlation among restaurants belonging to the same firm, we adjust the clustering level of the standard errors from the restaurant to the firm level. Given that the number of firms is small (10), we employ a wild bootstrap method to avoid biased standard errors and unreliable inference (Cameron et al., 2008). As shown in Column (1), the estimated effect remains statistically significant after switching the clustering unit.

Table 5. *Analysis on Masked Negativity*

Dep. Var.	Proportion of Masked Negativity in Positive Reviews			
	(1)	(2)	(3)	(4)
ARMS	-0.010	-0.039***	0.011	-0.015
	(0.014)	(0.012)	(0.025)	(0.026)
Rating Threshold	5	4	5	4
Sentiment Threshold	0.30	0.30	0.70	0.70
Dep. Var. Mean	0.034	0.070	0.138	0.221
Week FE	YES	YES	YES	YES
Restaurant FE	YES	YES	YES	YES
Treatment Linear Trend	YES	YES	YES	YES
Covariates $\times \lambda_t$	YES	YES	YES	YES
Observations	5165	5619	5165	5619
Clusters	122	122	122	122

Notes: This table investigates the change in masked negativity in consumer reviews. The dependent variables are defined by 4 different threshold combinations of star rating and sentiment score. Standard errors are in parentheses and clustered at the restaurant level. *, ** and *** indicate statistical significance at 10%, 5% and 1% level, respectively.

Second, although the relative time model confirms the general comparability between treatment and control restaurants, concerns may remain regarding residual differences between the two groups. To address this, we restrict the sample to the treatment group restaurants and perform a before-and-after comparison within this group. As shown in Column (2), after ruling out potential influences arising from differences between treatment and control restaurants, the estimated effects remain stable in both magnitude and statistical significance.

Third, another potentially omitted factor is demand shocks faced by restaurants, which could influence the consumer reviews they receive. To partially address this concern, we additionally control for the one-period lagged review volume, which serves as a proxy for demand-side patterns. As shown in Column (3), although the estimated effect slightly decreases in magnitude, the impact of ARMS adoption on restaurant quality remains positive and statistically significant.

5.5 Heterogeneity Analyses

5.5.1 Convergence in Restaurant Quality

At the time of ARMS adoption, restaurants may be operating at markedly different levels of business performance. For instance, some may be heavily burdened by negative reviews, whereas others may already perform relatively well in terms of consumer feedback. Given that ARMS primarily facilitates improvement by directing restaurant staff's attention to negative reviews, a natural hypothesis is that restaurants facing more severe negative feedback stand to benefit more from ARMS adoption.

To empirically examine this hypothesis, we ask *Company C* to compute each treatment restaurant's

Table 6. *Convergence in Quality with ARMS Adoption*

Dep. Var.	Average Star Rating	Proportion of Negative Reviews	Average Sentiment Score
	(1)	(2)	(3)
ARMS	0.149***	-0.036***	0.079***
	(0.051)	(0.013)	(0.019)
ARMS × PreRating	-0.160***	0.030***	-0.029**
	(0.025)	(0.007)	(0.013)
Dep. Var. Mean	4.366	0.065	0.700
Week FE	YES	YES	YES
Restaurant FE	YES	YES	YES
Treatment Linear Trend	YES	YES	YES
Covariates $\times\lambda_t$	YES	YES	YES
Observations	5723	5723	5612
Clusters	122	122	122

Notes: This table shows convergence in the quality of treatment restaurants. The explanatory variable is the interaction between ARMS adoption dummy and the initial quality of a restaurant. The initial quality is measured by standardized star rating prior to ARMS adoption. Standard errors are in parentheses and clustered at the restaurant level. *, ** and *** indicate statistical significance at 10%, 5% and 1% level, respectively.

pre-adoption star rating, calculated from all historical consumer reviews prior to ARMS adoption. We then standardize this measure by demeaning and dividing by its standard deviation, yielding a cross-sectional index denoted by $PreRating_i$. Using the same specification as Column (4) of Table 2, we further interact the treatment indicator $ARMS_{it}$ with $PreRating_i$. The coefficient on this interaction term captures heterogeneity in the effects of ARMS adoption as a function of restaurants' initial performance.

The regression results are reported in Table 6, where the outcome variables are, respectively, the star rating, the proportion of negative reviews, and the sentiment score. As indicated by the estimated coefficients on the interaction term across the three columns, restaurants with higher pre-adoption ratings benefit significantly less from ARMS adoption. By contrast, for restaurants starting from a relatively low level of initial quality (i.e., $PreRating_i = 0$), the improvements following ARMS adoption are both statistically significant and economically meaningful, as reflected in the coefficients on $ARMS_{it}$. The results indicate that ARMS promotes convergence in restaurant quality by enabling initially disadvantaged restaurants to achieve greater improvements through effective use of consumer negative reviews.

5.5.2 Dimensional Improvements

As introduced in Section 3.1, beyond the measure capturing overall restaurant quality, we identify whether each of 6 well-established dimensions is mentioned in a review and, if so, whether the associated sentiment is positive or negative: food taste, service, environment, waiting time, price, and food

hygiene. This more granular analysis of consumer review data allows us to examine heterogeneity along another dimension, namely, which aspects of restaurant operations benefit most from ARMS adoption.

Specifically, for each dimension k, we estimate the following equation:

$$Quality_{it}^{k} = \alpha_i + \lambda_t + \beta^k ARMS_{it} + \eta^k ARMS_{it} \times PreNegProp_i^k + \Gamma(X_i \cdot f(t)) + \mu G_i \cdot t + \varepsilon_{it}, \quad (3)$$

where the dependent variable $Quality_{it}^{k}$ is the dimensional quality measure, defined as the proportion of reviews containing positive comments within dimension k for restaurant i in week t. The variable $PreNegProp_i^k$ denotes the proportion of reviews containing negative comments within dimension k received by restaurant i prior to ARMS adoption, constructed in the same spirit as $PreRating_i$.

Table 7 reports the results, with each column corresponding to one dimension. Across all dimensions, we find positive effects of ARMS adoption, manifested as increases in the proportion of reviews expressing positive feedback along the corresponding dimension. Moreover, the coefficients on the interaction terms indicate that these improvements are dimension-specific. In particular, for dimensions in which restaurants performed relatively poorly prior to adoption (i.e., those associated with higher $PreNegProp_i^k$), the post-adoption improvements are more pronounced.

This pattern is weaker for the environment and food taste dimensions, as indicated by interaction estimates that are statistically indistinguishable from zero, although their signs remain positive. This is plausible, as improvements in the dining environment often require substantial fixed investments and may therefore be less flexible in the short run, while food taste is a fundamental attribute of a restaurant's offering and may leave limited scope for adjustment. For example, a restaurant specializing in spicy cuisine may find it difficult to respond to requests for lighter dishes.

6 Managerial Actions and Ripple Effects on Review Writing

6.1 Managerial Actions

Researchers on online reputation management have largely focused on actions that are directly visible to consumers, for example, how managers publicly respond to consumer reviews on digital platforms (Ananthakrishnan et al., 2023; Chen et al., 2019; Chevalier et al., 2018; Deng & Ravichandran, 2024; Proserpio & Zervas, 2017; Yang et al., 2019). Yet ARMS highlights a second, under-explored channel, through which restaurants process negative feedback internally, coordinate follow-up tasks, and implement operational fixes. Because these specific actions may never appear in public-facing communications, we refer to them as *back-end* managerial actions, in contrast to *front-end* responses posted on the platform. In practice, managers can allocate efforts to both front-end and back-end channels, but greater back-end efforts may reduce the resource for front-end responses.

Table 7. *Dimensional Improvements*

Dep. Var.	Proportion of Positive Reviews within the Dimension k					
	(1)	(2)	(3)	(4)	(5)	(6)
	Environment	Service	Food Taste	Waiting Time	Price	Food Hygiene
ARMS	0.142***	0.156***	0.054**	0.140***	0.132***	0.133***
	(0.037)	(0.038)	(0.021)	(0.037)	(0.037)	(0.041)
ARMS $\times PreNegProp^k$	0.043	0.082***	0.015	0.052**	0.087***	0.142***
	(0.027)	(0.024)	(0.021)	(0.020)	(0.021)	(0.020)
Dep. Var. Mean	0.753	0.724	0.855	0.265	0.550	0.503
Week FE	YES	YES	YES	YES	YES	YES
Restaurant FE	YES	YES	YES	YES	YES	YES
Treatment Linear Trend	YES	YES	YES	YES	YES	YES
Covariates $\times \lambda_t$	YES	YES	YES	YES	YES	YES
Observations	5723	5723	5723	5723	5723	5723
Clusters	122	122	122	122	122	122

Notes: This table depicts dimensional improvements by examining the heterogeneous effects of ARMS according to baseline negative review intensity within 6 dimensions. The dependent variable is weekly dimensional quality measured by the proportion of positive reviews within each dimension. Standard errors are in parentheses and clustered at the restaurant level. *, ** and *** indicate statistical significance at 10%, 5% and 1% level, respectively.

6.1.1 Effects of ARMS Adoption on Front-end Actions

We calculated weekly managerial response rate, and then evaluated the impact of ARMS adoption on them. Table 8 reports the results. Across all specifications consistent with baseline analysis, we find that ARMS partially crowds out public-facing replies, implying that structured back-end workflow introduced by ARMS can substitute for front-end review management in practice.

6.1.2 Back-End Attitudes and Degree of Improvement

We then turn to the work ticket data, which records restaurant staff's back-end actions and thus allows us to examine how back-end staff attitudes toward negative reviews shape the effectiveness of ARMS adoption. As discussed in Section 3.2, based on the discussion texts in work tickets, the attitudes to negative reviews can be broadly classified as *defensive* versus *reflective*. The former refers to a tendency to attribute complaints to consumer-related factors, such as excessive expectations, whereas the latter emphasizes internal service or operational shortcomings. Such differentiated attitudes are likely to affect the extent to which restaurants can translate ARMS adoption into tangible quality improvements.

To empirically investigate this channel, we compile all discussion texts at the restaurant level and classify each text as defensive or reflective using the DeepSeek-V3 large language model. We then compute, for each restaurant, the proportion of texts exhibiting a defensive attitude and define an indicator variable, $Defensive_i$, equal to one if this proportion exceeds the sample average. We interact

Table 8. *ARMS Impact on Managerial Response Rate*

Dep. Var.	Managerial Response Rate			
	(1)	(2)	(3)	(4)
ARMS	-0.209***	-0.340***	-0.339***	-0.384***
	(0.038)	(0.030)	(0.029)	(0.039)
Dep. Var. Mean	0.731	0.731	0.731	0.731
Week FE	YES	YES	YES	YES
Restaurant FE	YES	YES	YES	YES
Treatment Linear Trend		YES	YES	YES
Covariates $\times f(t)$			YES	
Covariates $\times \lambda_t$				YES
Observations	5723	5723	5723	5723
Clusters	122	122	122	122

Notes: This table presents the change in front-end managerial response rate following the adoption of ARMS. The specifications are consistent with baseline analysis in Table 2. Standard errors are in parentheses and clustered at the restaurant level. *, ** and *** indicate statistical significance at 10%, 5% and 1% level, respectively.

$Defensive_i$ with the treatment dummy and re-estimate the baseline specification.

Two caveats are worth noting. First, because only 81 restaurants activated the work ticket management tool, the analysis is conducted on a reduced sample. Second, work ticket data are only available after ARMS adoption, implying that $Defensive_i$ is constructed using post-treatment information. Including such a variable on the right-hand side of the regression introduces a classic "bad control" concern (Angrist & Pischke, 2009). Consequently, while the estimates are still informative about the role of back-end staff attitudes in shaping ARMS effectiveness, they should be interpreted as correlation rather than causality.

Table 9 reports the results. We find that ARMS delivers significantly smaller improvements for restaurants in which internal discussions of negative reviews are more defensive. Quantitatively, the quality gains for restaurants with a defensive attitude are less than half of those observed for their more reflective counterparts. These findings indicate that the returns to ARMS are not mechanical consequences of system installation or activation. Instead, they are related to whether frontline staff use the system as an input for shortcoming recognition and continuous improvement, rather than responding by deflecting blame. From a managerial perspective, an important accompanying task when deploying tools such as ARMS is therefore to pair technological rollout with organizational complements that shape how employees interpret and act upon the information surfaced by the technology. In short, effective technology adoption requires integration with organizational change, rather than standalone implementation.

Table 9. *Back-End Attitudes and Degree of Improvement*

Dep. Var.	Rating	% Negative	Sentiment
	(1)	(2)	(3)
ARMS	0.230***	-0.054***	0.117***
	(0.070)	(0.017)	(0.026)
ARMS × Defensive	-0.151**	0.031**	-0.055**
	(0.058)	(0.014)	(0.026)
Dep. Var. Mean	4.353	0.064	0.692
Week FE	YES	YES	YES
Restaurant FE	YES	YES	YES
Treatment Linear Trend	YES	YES	YES
Covariates $\times \lambda_t$	YES	YES	YES
Observations	4840	4840	4743
Clusters	99	99	99

Notes: This table shows heterogeneous effect of back-end attitudes. Restaurants are defined as defensive if the proportion of defensive discussion texts in work tickets exceeds the sample average. The regression sample is restricted to treatment restaurants that have retained records of internal discussions and control restaurants. Standard errors are in parentheses and clustered at the restaurant level. *, ** and *** indicate statistical significance at 10%, 5% and 1% level, respectively.

6.1.3 Back-End Attitudes and Front-End Actions

The preceding results motivate us to examine whether restaurants' back-end attitudes toward negative feedback are systematically linked to their public-facing response behavior. Much of the existing literature interprets front-end responses to consumer reviews as evidence of managerial attention or learning. However, such responses may not be an unbiased indicator of genuine managerial engagement if back-end attitudes shape how restaurants choose to respond publicly. In particular, front-end responses may capture not only efforts to learn from consumer feedback and improve operations, but also defensive reputation management aimed at mitigating public criticism without substantive internal change. As a result, relying solely on front-end responses risks mischaracterizing how restaurants actually process and act upon consumer feedback.

We proceed by classifying restaurants into defensive and reflective by the indicator $Defensive_i$ we defined earlier. For each group, we compute the overall average managerial response rate in the pre-adoption period and in the post-adoption period within our sample window, and then calculate the before-and-after change as the difference between the post- and pre-adoption averages. We then test whether the pre-adoption level, the post-adoption level, and the before-and-after change differ across the two groups.

The results in Table 10 show a clear difference. Defensive restaurants exhibit higher response rates than reflective restaurants both before and after adoption. At the same time, response rates decline after adoption for both groups, with the absolute decline being slightly smaller for the defensive group

than for the reflective group. These patterns provide summary evidence that reflective back-end attitude is negatively correlated with front-end response behavior. Because reflective restaurants achieve greater quality improvements than defensive ones, and because ARMS-adopting restaurants improve quality more than non-adopters despite a decline in front-end response rates, these findings suggest that ARMS-enabled reflective back-end actions are more effective than front-end responses.

Table 10. *Managerial Response Rate by Back-End Attitudes*

Variable	Reflective	Defensive	Difference
	(1)	(2)	(3)
Overall pre-adoption average response rate	0.689	0.958	-0.269***
	(0.310)	(0.113)	(0.052)
Overall post-adoption average response rate	0.488	0.846	-0.357***
	(0.306)	(0.195)	(0.057)
Before-after change	-0.201	-0.113	-0.088*
	(0.279)	(0.185)	(0.053)
Observations	41	40	81

Notes: This table reports the results of comparison of front-end managerial response rate by back-end attitudes. Columns (1) and (2) show the means and standard deviations (in parentheses) of the corresponding variables. Column (3) displays the differences between the reflective and defensive groups, with standard errors in parentheses. *, ** and *** indicate statistical significance at 10%, 5% and 1% level, respectively.

6.2 Ripple Effects on Consumer Engagement in Review Writing

Finally, we move beyond restaurant-level outcomes and examine whether the improvements induced by ARMS translate into changes in consumer review writing behavior. Because ARMS adoption is a back-end decision unobservable to consumers, the differences we observe in consumer review behavior between ARMS-adopting and non-ARMS restaurants would likely reflect consumer experience of products and services in the target restaurant, which could be a function of ARMS adoption.

We analyze two aspects of review behavior, focusing first on the informativeness of reviews. To this end, we construct two measures. The first is the average number of dimensions (out of the six) mentioned in a review. The second is the proportion of reviews that do not mention any specific dimension, which are largely uninformative from a content perspective. Columns (1) and (2) of Table 11 report regression results using these two measures as outcome variables. The estimates jointly indicate that consumer reviews became more multidimensional and detailed following ARMS adoption, reflecting an increase in the richness and informativeness of review content.

Another aspect of consumer review behavior is the effort exerted in writing reviews, which we measure by (i) whether consumers attach pictures to their reviews, and (ii) the length of the review text. Correspondingly, Columns (3) and (4) report results using the proportion of reviews with attached pictures and the logarithm of review length as outcome variables, respectively. The estimates show

Table 11. *Ripple Effect on Consumer Engagement in Review Writing*

Dep. Var.	Average Dimension Count	Topicless Proportion	Picture-attached Proportion	Log Average Length
	(1)	(2)	(3)	(4)
ARMS	0.209**	-0.054***	0.054*	0.128*
	(0.085)	(0.020)	(0.032)	(0.069)
Dep. Var. Mean	2.286	0.101	0.569	4.396
Week FE	YES	YES	YES	YES
Restaurant FE	YES	YES	YES	YES
Treatment Linear Trend	YES	YES	YES	YES
Covariates $\times \lambda_t$	YES	YES	YES	YES
Observations	5723	5723	5723	5612
Clusters	122	122	122	122

Notes: This table reports the ripple effect of ARMS on consumer engagement in review writing. The dependent variable in Column (1) is the number of dimensions mentioned in review texts, and Column (2) replaces it with the proportion of reviews that do not mention any dimension. In Column (3), the dependent variable is the proportion of reviews with attached pictures. Column (4) examines the weekly average log length of review texts. Standard errors are in parentheses and clustered at the restaurant level. *, ** and *** indicate statistical significance at 10%, 5% and 1% level, respectively.

positive effects on both dimensions, indicating that consumers exerted greater effort when composing and posting reviews on *Dianping*.

We acknowledge that informativeness and consumer effort are closely intertwined. Writing more informative reviews generally requires greater effort, and, conversely, longer and more effortful reviews are typically more informative. Therefore, the results in Table 11 should be interpreted as suggestive evidence of changes in consumer review behavior, observed from complementary angles, rather than as definitive interpretations of any single variable. That said, we carefully select measures that capture the informativeness and effort embedded in consumer reviews, while abstracting from the underlying quality of restaurant products and services being evaluated. This approach allows us to distinguish the effect of ARMS adoption on restaurant quality (documented in Sections 5 and 6) from its potential spillover effects on reviewers' effort.

Additionally, the observed impact on consumer review writing provides further evidence that our findings are unlikely to be driven purely by strategic manipulation. Prior research has documented that review solicitation or incentivization can degrade the information quality of consumer reviews (Gao et al., 2025; Khern-am-nuai et al., 2018). Consequently, if the post-ARMS improvements were primarily the result of strategic manipulation, such as bribing consumers to leave 5-star ratings, we would expect consumers to expend less effort and be more perfunctory in writing reviews. In contrast, as shown in Table 11, consumers appear to be more engaged in drafting reviews, which directly counters a manipulation-driven interpretation.

7 Managerial Implications

Our findings yield several managerial implications for restaurants and, more broadly, for firms seeking to leverage user-generated content for operational improvement.

First, the role of ARMS highlights that the central challenge in the digital age is not the lack of consumer information, but the ability to make publicly available information actionable. Online review platforms such as *Dianping* already generate vast volumes of consumer feedback, yet such information is often dispersed, noisy, and difficult to process in real time. ARMS adds value not by creating new information, but by transforming existing public information into timely, structured, and decision-relevant inputs for managers. This suggests that, under conditions of information overload, managerial bottlenecks increasingly lie in attention, prioritization, and execution rather than in information acquisition per se. For practitioners, investing in systems that reduce cognitive and organizational frictions in processing feedback may yield returns comparable to, or even exceeding, investments in traditional quality monitoring technologies.

Second, our evidence on back-end staff attitudes underscores that technology alone is insufficient to generate performance gains. Using work ticket data, we show that restaurants whose internal discussions of negative reviews are more reflective rather than defensive experience significantly larger quality improvements following ARMS adoption. This finding implies that the effectiveness of digital monitoring tools critically depends on how frontline employees and managers interpret and respond to the information these tools surface. Defensive reactions that attribute complaints to consumer bias or unreasonable expectations can blunt the learning potential of consumer feedback, even when such feedback is delivered in a timely and organized manner. From a managerial perspective, this highlights the importance of complementary organizational practices, such as training, incentive design, and leadership communication, that foster a learning-oriented mindset. Technology adoption should therefore be paired with deliberate efforts to shape employee attitudes toward feedback, rather than treated as a purely technical upgrade.

Third, the observed ripple effects on consumer review writing suggest the existence of a virtuous cycle between operational improvement and consumer engagement. We find that after ARMS adoption, consumer reviews become more informative, more detailed, and more effortful. This pattern is consistent with a feedback loop in which genuine improvements in service quality motivate consumers to invest more effort in articulating their experiences, which in turn generates richer information for firms to learn from and further improve. For managers, this implies that quality improvement initiatives may generate indirect benefits by enhancing the informational value of subsequent consumer feedback, thereby amplifying long-run returns. For platforms such as *Dianping*, these results point to a novel mechanism to stimulate high-quality user participation: by facilitating merchants' effective

utilization of consumer feedback, platforms may indirectly encourage consumers to contribute more substantive and informative reviews, improving the overall quality of platform content.

Taken together, our findings suggest that the managerial value of digital review systems lies not only in aggregating information like providing star ratings, but in enabling learning, shaping organizational responses, and fostering positive feedback loops between firms and consumers. Effective deployment of such systems requires aligning technology like ARMS, organizational culture, and platform design to fully realize their potential.

8 Conclusion

This paper studies how the automated review monitoring system transforms online feedback into operational improvements. Exploiting the staggered adoption of ARMS, we find that adoption increases average ratings, reduces the proportion of negative reviews, and improves textual sentiment. These gains are not driven by review manipulation: results are stable when excluding 5-star reviews, and we do not observe an increase in masked negativity. We further document heterogeneity: ARMS effects are stronger for initially lower-quality restaurants, and improvements are larger in dimensions where restaurants historically performed poorly.

Additional analysis provides implications on managerial actions and consumer engagement in review writing. First, as a tool to strengthen back-end management, the adoption of ARMS substitutes for front-end managerial response. Second, the benefits are substantially muted when back-end attitudes exhibit a more defensive style. Third, back-end attitudes are systematically related to front-end response behavior. Finally, consumers write more detailed and multidimensional reviews following the adoption of ARMS, which indicates higher engagement and suggests that reviews are less likely to be subject to manipulation-driven distortions.

Our findings highlight that digital technologies can turn eWOM from a passive reputation constraint into an active input for operational improvement. Realizing these gains, however, requires more than simply installing the tool. Managers need to embed it into routine operations so that feedback is systematically translated into informative insights and follow-up actions, without resorting to ineffective reputation management and manipulation.

References

Acemoglu, D., Makhdoumi, A., Malekian, A., & Ozdaglar, A. (2022). Learning from Reviews: The Selection Effect and the Speed of Learning. *Econometrica*, *90*(6), 2857–2899.

Ananthakrishnan, U., Proserpio, D., & Sharma, S. (2023). I Hear You: Does Quality Improve with Customer Voice? *Marketing Science*, *42*(6), 1143–1161.

Anderson, M., & Magruder, J. (2012). Learning from the Crowd: Regression Discontinuity Estimates of the Effects of An Online Review Database. *Economic Journal*, *122*(563), 957–989.

Anenberg, E., Kuang, C., & Kung, E. (2022). Social Learning and Local Consumption Amenities: Evidence from Yelp. *Journal of Industrial Economics*, *70*(2), 294–322.

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Archak, N., Ghose, A., & Ipeirotis, P. G. (2011). Deriving the Pricing Power of Product Features by Mining Consumer Reviews. *Management Science*, *57*(8), 1485–1509.

Aziz, A., Li, H., & Telang, R. (2023). The Consequences of Rating Inflation on Platforms: Evidence from a Quasi-experiment. *Information Systems Research*, *34*(2), 590–608.

Basuroy, S., Chatterjee, S., & Ravid, S. A. (2003). How Critical Are Critical Reviews? The Box Office Effects of Film Critics, Star Power, and Budgets. *Journal of Marketing*, *67*(4), 103–117.

Bolton, G. E., Kusterer, D. J., & Mans, J. (2019). Inflated Reputations: Uncertainty, Leniency, and Moral Wiggle Room in Trader Feedback Systems. *Management Science*, *65*(11), 5371–5391.

Borusyak, K., Jaravel, X., & Spiess, J. (2024). Revisiting Event-study Designs: Robust and Efficient Estimation. *Review of Economic Studies*, *91*(6), 3253–3285.

Burtch, G., Hong, Y., Bapna, R., & Griskevicius, V. (2018). Stimulating Online Reviews by Combining Financial Incentives and Social Norms. *Management Science*, *64*(5), 2065–2082.

Büschken, J., & Allenby, G. M. (2016). Sentence-based Text Analysis for Customer Reviews. *Marketing Science*, *35*(6), 953–975.

Cabral, L., & Li, L. (2015). A Dollar for Your Thoughts: Feedback-conditional Rebates on eBay. *Management Science*, *61*(9), 2052–2063.

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based Improvements for Inference with Clustered Errors. *Review of Economics and Statistics*, *90*(3), 414–427.

Cengiz, D., Dube, A., Lindner, A., & Zipperer, B. (2019). The Effect of Minimum Wages on Low-wage Jobs. *Quarterly Journal of Economics*, *134*(3), 1405–1454.

Chakraborty, I., Kim, M., & Sudhir, K. (2022). Attribute Sentiment Scoring with Online Text Reviews: Accounting for Language Structure and Missing Attributes. *Journal of Marketing Research*, *59*(3), 600–622.

Chen, W., Gu, B., Ye, Q., & Zhu, K. X. (2019). Measuring and Managing the Externality of Managerial Responses to Online Customer Reviews. *Information Systems Research*, *30*(1), 81–96.

Chen, Y., Harper, F. M., Konstan, J., & Li, S. X. (2010). Social Comparisons and Contributions to Online Communities: A Field Experiment on Movielens. *American Economic Review*, *100*(4), 1358–1398.

Chen, Z., Yu, B., Yang, C., Zhou, Y., Yao, S., Qian, X., Wang, C., Wu, B., & Wu, J. (2021). An Extended Time Series (2000–2018) of Global NPP-VIIRS-like Nighttime Light Data from a Cross-sensor Calibration. *Earth System Science Data*, *13*(3), 889–906.

Chevalier, J. A., Dover, Y., & Mayzlin, D. (2018). Channels of Impact: User Reviews when Quality is Dynamic and Managers Respond. *Marketing Science*, *37*(5), 688–709.

Chevalier, J. A., & Mayzlin, D. (2006). The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal of Marketing Research*, *43*(3), 345–354.

Cho, H. S., Sosa, M. E., & Hasija, S. (2022). Reading between the Stars: Understanding the Effects of Online Customer Reviews on Product Demand. *Manufacturing & Service Operations Management*, *24*(4), 1977–1996.

Deng, C., & Ravichandran, T. (2024). Managerial Response to Online Positive Reviews: Helpful or Harmful? *Information Systems Research*, *35*(4), 1802–1823.

Deng, Y., Zheng, J., Khern-Am-Nuai, W., & Kannan, K. (2022). More than the Quantity: The Value of Editorial Reviews for a User-generated Content Platform. *Management Science*, *68*(9), 6865–6888.

Fang, L. (2022). The Effects of Online Review Platforms on Restaurant Revenue, Consumer Learning, and Welfare. *Management Science*, *68*(11), 8116–8143.

Farronato, C., & Zervas, G. (2022). Consumer Reviews and Regulation: Evidence from NYC Restaurants. *Working Paper*.

Filippas, A., Horton, J. J., & Golden, J. M. (2022). Reputation Inflation. *Marketing Science*, *41*(4), 733–745.

Fradkin, A., Grewal, E., & Holtz, D. (2021). Reciprocity and Unveiling in Two-sided Reputation Systems: Evidence from an Experiment on Airbnb. *Marketing Science*, *40*(6), 1013–1029.

Gandhi, A., Hollenbeck, B., & Li, Z. (2024). Misinformation and Mistrust: The Equilibrium Effects of Fake Reviews on Amazon.com. *Working Paper*.

Gao, B., Wang, J., Ding, X., & Guo, Y. (2025). The Pitfalls of Review Solicitation: Evidence from a Natural Experiment on TripAdvisor. *Management Science*, *71*(2), 1671–1691.

Goes, P. B., Lin, M., & Au Yeung, C.-m. (2014). "Popularity Effect" in User-generated Content: Evidence from Online Product Reviews. *Information Systems Research*, *25*(2), 222–238.

Goodman-Bacon, A. (2021). Difference-in-differences with Variation in Treatment Timing. *Journal of Econometrics*, *225*(2), 254–277.

He, S., Hollenbeck, B., & Proserpio, D. (2022). The Market for Fake Reviews. *Marketing Science*, *41*(5), 896–921.

Hollenbeck, B. (2018). Online Reputation Mechanisms and the Decreasing Value of Chain Affiliation. *Journal of Marketing Research*, *55*(5), 636–654.

Horton, J., & Golden, J. (2015). Reputation Inflation in an Online Marketplace. *Working Paper*.

Huang, N., Burtch, G., Gu, B., Hong, Y., Liang, C., Wang, K., Fu, D., & Yang, B. (2019). Motivating User-generated Content with Performance Feedback: Evidence from Randomized Field Experiments. *Management Science*, *65*(1), 327–345.

Jiang, L. D., Ravichandran, T., & Kuruzovich, J. (2023). Review Moderation Transparency and Online Reviews: Evidence from A Natural Experiment. *MIS Quarterly*, *47*(4), 1693–1708.

Karanam, S. A., Agarwal, A., & Barua, A. (2025). Follow Your Heart or Listen to Users? The Case of Mobile App Design. *Information Systems Research*, *36*(3), 1846–1870.

Khern-am-nuai, W., Kannan, K., & Ghasemkhani, H. (2018). Extrinsic Versus Intrinsic Rewards for Contributing Reviews in an Online Platform. *Information Systems Research*, *29*(4), 871–892.

Lappas, T., Sabnis, G., & Valkanas, G. (2016). The Impact of Fake Reviews on Online Visibility: A Vulnerability Assessment of the Hotel Industry. *Information Systems Research*, *27*(4), 940–961.

Lewis, G., & Zervas, G. (2016). The Welfare Impact of Consumer Reviews: A Case Study of the Hotel Industry. *Working Paper*.

Li, L., & Xiao, E. (2014). Money Talks: Rebate Mechanisms in Reputation System Design. *Management Science*, *60*(8), 2054–2072.

Luca, M. (2016). Reviews, Reputation, and Revenue: The Case of Yelp.com. *Working Paper*.

Luca, M., & Zervas, G. (2016). Fake it Till You Make it: Reputation, Competition, and Yelp Review Fraud. *Management Science*, *62*(12), 3412–3427.

Mayzlin, D., Dover, Y., & Chevalier, J. (2014). Promotional Reviews: An Empirical Investigation of Online Review Manipulation. *American Economic Review*, *104*(8), 2421–2455.

Mejia, J., Mankad, S., & Gopal, A. (2019). A for effort? Using the Crowd to Identify Moral Hazard in New York City Restaurant Hygiene Inspections. *Information Systems Research*, *30*(4), 1363–1386.

Moretti, E. (2011). Social Learning and Peer Effects in Consumption: Evidence from Movie Sales. *Review of Economic Studies*, *78*(1), 356–393.

Pocchiari, M., Schoenmueller, V., & Dover, Y. (2024). Online Review Updating: Prevalence and Implications for Platforms and Businesses. *Working Paper*.

Proserpio, D., & Zervas, G. (2017). Online Reputation Management: Estimating the Impact of Management Responses on Consumer Reviews. *Marketing Science*, *36*(5), 645–665.

Shin, D., Vaccari, S., & Zeevi, A. (2023). Dynamic Pricing with Online Reviews. *Management Science*, *69*(2), 824–845.

Sun, L., & Abraham, S. (2021). Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects. *Journal of Econometrics*, *225*(2), 175–199.

Tadelis, S. (2016). Reputation and Feedback Systems in Online Platform Markets. *Annual Review of Economics*, *8*(1), 321–340.

Wang, L., Luo, X., Qiu, L., Xu, F., & Cui, X. (2025). Win by Hook or Crook? Self-injecting Favorable Online Reviews to Fight Adjacent Rivals. *Information Systems Research*, *36*(3), 1333–1353.

Wu, C., Che, H., Chan, T. Y., & Lu, X. (2015). The Economic Value of Online Reviews. *Marketing Science*, *34*(5), 739–754.

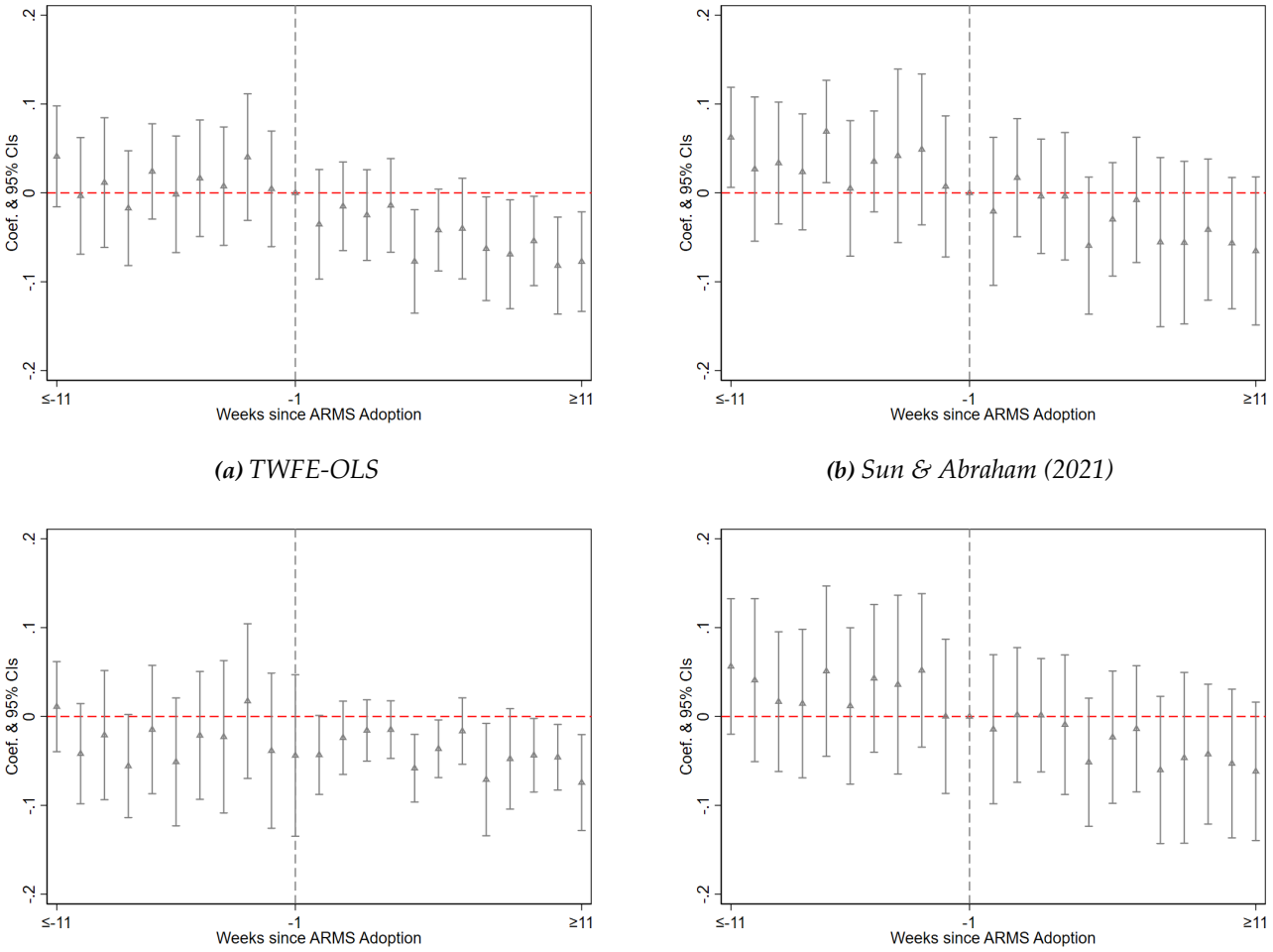
Yang, M., Zheng, Z., & Mookerjee, V. (2019). Prescribing Response Strategies to Manage Customer Opinions: A Stochastic Differential Equation Approach. *Information Systems Research*, *30*(2), 351–374.

Yu, Y., Khern-am-nuai, W., & Pinsonneault, A. (2022). When Paying for Reviews Pays Off: The Case of Performance-contingent Monetary Rewards. *MIS Quarterly*, *46*(1), 609–626.

Zhao, Y., Yang, S., Narayan, V., & Zhao, Y. (2013). Modeling Consumer Learning from Online Product Reviews. *Marketing Science*, *32*(1), 153–169.

Zhu, F., & Zhang, X. (2010). Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics. *Journal of Marketing*, *74*(2), 133–148.

A Additional Figures and Tables



(a) *TWFE-OLS*

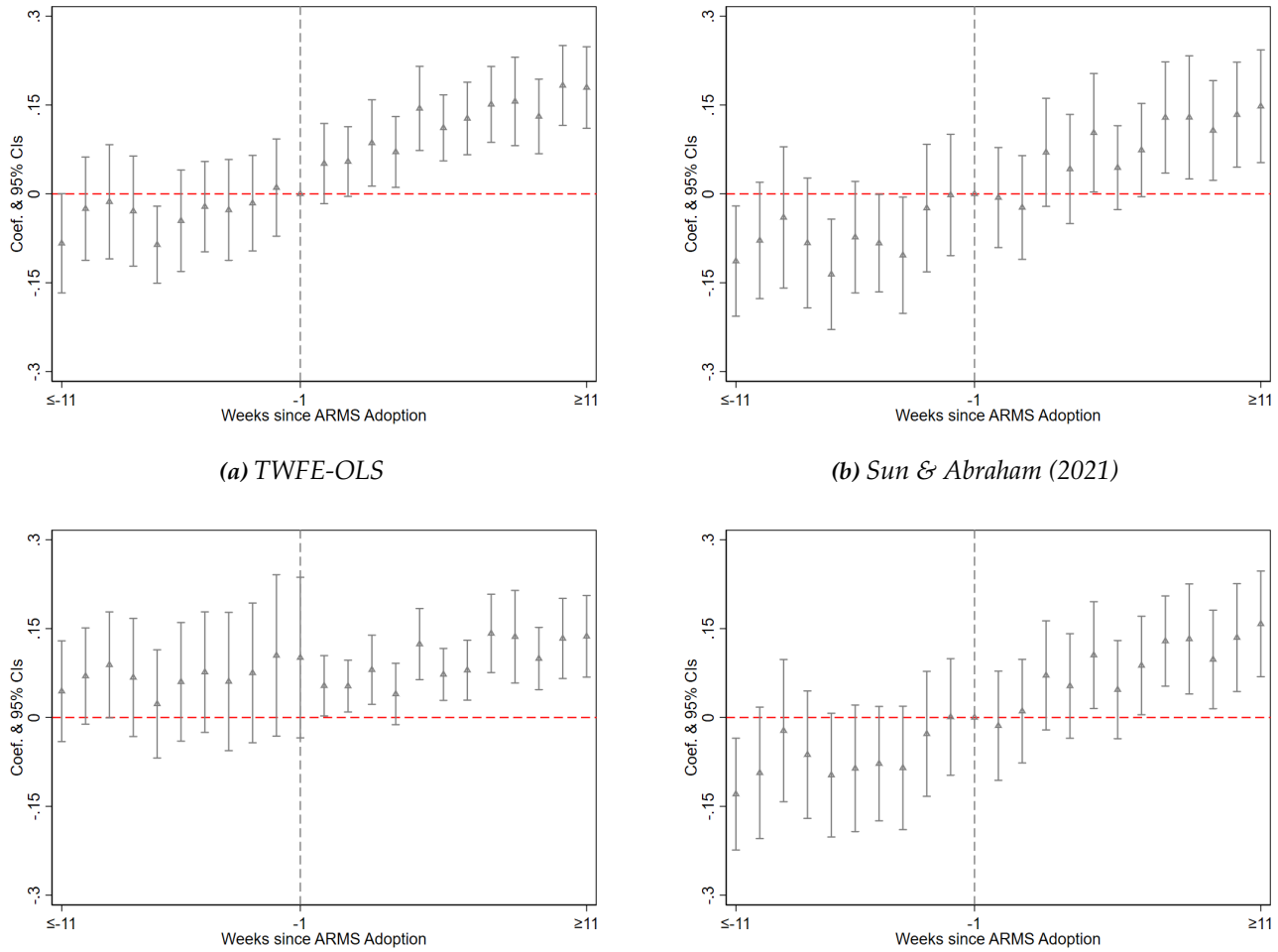
(b) *Sun & Abraham (2021)*

(c) *Borusyak et al. (2024)*

(d) *Cengiz et al. (2019)*

Figure A1. *Dynamic Effect of ARMS on Proportion of Negative Reviews*

Notes: This figure portrays estimates from the relative time model as specified in Equation (2). The dependent variable is the proportion of negative reviews. The x-axis represents the time period relative to ARMS activation. Points represent coefficient estimates and vertical bars show 95% confidence intervals.



(a) *TWFE-OLS*

(b) *Sun & Abraham (2021)*

(c) *Borusyak et al. (2024)*

(d) *Cengiz et al. (2019)*

Figure A2. *Dynamic Effect of ARMS on Proportion of Negative Reviews*

Notes: This figure portrays estimates from the relative time model as specified in Equation (2). The dependent variable is the weekly average sentiment score. The *x*-axis represents the time period relative to ARMS activation. Points represent coefficient estimates and vertical bars show 95% confidence intervals.

Table A1. *Robustness Check: Alternative Specifications*

Dep. Var.	Average Rating Score		
	(1)	(2)	(3)
ARMS	0.131**	0.142**	0.098*
	[0.040]	(0.066)	(0.051)
Dep. Var. Mean	4.366	4.396	4.366
Week FE	YES	YES	YES
Restaurant FE	YES	YES	YES
Treatment Linear Trend	YES		YES
Covariates $\times \lambda_t$	YES	YES	YES
Lag # of Reviews			YES
Only Treatment Restaurants		YES	
Observations	5723	4361	5723
Clusters	10	104	122

Notes: This table reports baseline results using alternative specifications to show robustness. All regressions control for week and restaurant fixed effects. Columns (1) and (3) control for treatment-specific linear time trend. Column (2) includes only treatment restaurants. Column (3) further controls one-period lag of the number of consumer reviews. Column (1) reports the p-value of wild cluster bootstrap (Cameron et al., 2008) in the bracket. Standard errors in parentheses are clustered at the restaurant level for Columns (2) and (3). *, ** and *** indicate statistical significance at 10%, 5% and 1% level, respectively.