https://pubsonline.informs.org/journal/isre

INFORMATION SYSTEMS RESEARCH
Articles in Advance, pp. 1–11
ISSN 1047-7047 (print), ISSN 1526-5536 (online)

Can Regulation Restore Authenticity? Evidence from Policies Against Online Review Suppression

Aida Sanatizadeh,[a] Gordon Burtch,[b] Yili Hong,[c,*] Yuheng Hu[d]

[a] College of Business, Northern Illinois University, DeKalb, Illinois 60115; [b] Questrom School of Business, Boston University, Boston, Massachusetts 02215; [c] Miami Herbert Business School, University of Miami, Coral Gables, Florida 33146; [d] Fisher College of Business, Ohio State University, Columbus, Ohio 43210

*Corresponding author

Contact: asanatizadeh@niu.edu (AS); gburtch@bu.edu, https://orcid.org/0000-0001-9798-1113 (GB); khong@miami.edu, https://orcid.org/0000-0002-0577-7877 (YiH); hu.3331@osu.edu, https://orcid.org/0000-0003-3665-1238 (YuH)

Received: July 20, 2023
Revised: July 23, 2024; April 1, 2025; August 22, 2025
Accepted: September 25, 2025
Published Online in Articles in Advance: December 22, 2025

https://doi.org/10.1287/isre.2023.0436

Abstract. Anecdotal evidence suggests that businesses threaten consumers with legal action to suppress negative online reviews, creating a chilling effect that undermines transparency and trust online. Although policymakers increasingly recognize this issue and have taken steps to curb it, academic research exploring the effectiveness of regulatory interventions remains surprisingly scarce. This paper addresses that gap by providing a systematic evaluation of the Consumer Review Fairness Act (CRFA), a federal law enacted in the United States in December 2016 to curb review suppression. Leveraging detailed data from TripAdvisor, we demonstrate that the CRFA has had measurable impacts; following the law's enactment, hotel reviews in the United States have become systematically more negative, as indicated by both star ratings and textual sentiment. Furthermore, reviews have grown systematically longer. These results demonstrate that, following passage of the CFRA, reviews have become generally more informative but systematically more negative. Examining heterogeneity, we find that these declines in ratings have been most concentrated among less reputable hotels and hotels facing stronger competition, indicating that such businesses had been more heavily engaged in review suppression before the CFRA's passage. Furthermore, we see that these effects have accrued most heavily to reviews involving American consumers and long-tenured users, highlighting the role and influence of legal jurisdictional boundaries and platform experience. Finally, supplemental analyses speak to the prevalence of review censorship on TripAdvisor and the chilling effects of legal threats on consumers' subsequent reviews. We discuss the implications for policy, practice, and scholarship related to consumer protection and online transparency.

History: Juan Feng, Senior Editor; Yixin Lu, Associate Editor.
Supplemental Material: The online appendix is available at https://doi.org/10.1287/isre.2023.0436.

Keywords: Consumer Review Fairness Act (CRFA) • negative feedback • online reviews • opinion suppression

Introduction

Online reviews play a crucial role in consumer decision-making, with even small shifts in review valence having profound economic consequences. For example, a one-star increase in a restaurant's average rating can boost sales by nearly 18% (Luca 2016). Given these high stakes, it is unsurprising that businesses often go to great lengths to obtain positive feedback. Existing research highlights a widespread practice across industries in which firms purchase fake positive reviews to artificially boost their reputation (see, e.g., Luca and Zervas 2016, Liu and Feng 2021, He et al. 2022). Yet despite the attention given to fake reviews, there is another troubling practice that remains largely overlooked: businesses' suppression of truthful negative feedback through legal threats.

Media reports offer striking examples of businesses intimidating reviewers into removing critical feedback. These include a New York gynecologist who filed a $1 million defamation lawsuit against a patient over an unfavorable review (Moutos et al. 2020), a Missouri tourist attraction that sued a Kansas man and his daughter for a moderate three-star review,[1] and even an American guest at a Thai hotel who was arrested and jailed for refusing to delete a negative TripAdvisor review.[2] Recognizing the seriousness of such abuses, the Federal Trade Commission clarified as recently as August 2024 that businesses are prohibited from using unfounded legal threats or intimidation to remove negative consumer reviews.[3]

Suppressing negative feedback not only biases consumer perceptions of business quality but can have

even more damaging consequences than fake positive reviews. Negative reviews are especially influential because consumers often rely disproportionately on critical feedback when evaluating products and services (Sen and Lerman 2007, Mayzlin et al. 2014, Yin et al. 2014). Although fake positive reviews artificially inflate a firm's rating, negative feedback still remains visible. By contrast, removing truthful negative reviews completely erases vital information. Moreover, legal threats can trigger chilling effects, leading consumers who experience intimidation to self-censor by refraining from future negative reviews—or altering their tone to be more positive—to avoid legal risks. This chilling effect potentially distorts the entire review ecosystem, undermining trust and consumer welfare broadly.

In response to these challenges, policymakers and review platforms have begun to act. For example, Yelp introduced a "Questionable Legal Threats Alert" in 2020, highlighting businesses that attempt to silence reviewers. Most significantly, the U.S. government enacted the Consumer Review Fairness Act (CRFA) in December 2016, explicitly prohibiting businesses from using legal threats to suppress negative reviews and establishing penalties to enforce compliance (Goldman 2017).[4] Yet despite these regulatory interventions, little systematic evidence exists on whether they effectively curb review suppression. Our study addresses this important gap. Leveraging extensive data from TripAdvisor, we systematically evaluate whether and how the CRFA has influenced consumer reviewing behavior for U.S. hotels. Specifically, we explore two primary research questions: (a) How effective have recent policy responses been in reducing the suppression of negative online reviews, and (b) which types of hotels and consumers have been most impacted by these policies?

We employ a difference-in-differences design, comparing reviews posted about U.S.-based hotels with those for hotels located in jurisdictions unaffected by the CRFA. Our findings provide clear evidence that the CRFA has meaningfully reduced review suppression. After the CRFA's enactment, hotel reviews in the U.S. became systematically longer and less positive in textual sentiment and exhibited lower average ratings, indicating that previously suppressed negative feedback has emerged. Further analyses reveal that these effects are most pronounced among less reputable hotels—those with historically lower ratings—and among hotels operating in highly competitive markets. This suggests that businesses with weaker reputations or facing intense competition were previously more inclined to silence criticism. Additionally, the CRFA's effects are strongest among American reviewers, underscoring the critical influence of geographic and jurisdictional context on consumers' willingness to speak candidly online. The effects are also more pronounced among users with longer tenure on the platform, suggesting that experienced reviewers, who are more likely to have encountered or recognized legal threats, are exceptionally responsive to the CRFA policy change. To assess the robustness and generalizability of our findings, we replicate the analysis using Google Places reviews and examine the impact of related state-level legal changes, particularly the repeal of criminal defamation laws. These robustness checks consistently support our primary results.

This research contributes significantly to several literatures. First, we provide what is, to our knowledge, the first systematic empirical analysis of negative review suppression, enriching the broader discussion on consumer feedback and online ratings (Chevalier and Mayzlin 2006, He et al. 2022). Second, our results offer novel evidence of the chilling effects associated with businesses' threats against consumers, highlighting spillover effects that extend beyond specific interactions. When consumers fear legal repercussions, their reviewing behavior across multiple businesses and platforms can be distorted, reducing the reliability and authenticity of online reviews generally. Third, our study underscores how governmental regulations and legal frameworks critically shape online consumer behavior, complementing prior research on how social contexts and cultural environments influence product evaluations (Danescu-Niculescu-Mizil et al. 2009, Hong et al. 2016, Nakayama and Wan 2019).

Our work also has important managerial implications. Prior studies have underscored the economic importance of online reviews, emphasizing their substantial influence on business revenue and consumer well-being (Wu et al. 2015, Liu et al. 2017, Feng et al. 2019, Fang 2022). Therefore, actively monitoring and flagging businesses that attempt review suppression is vital for digital platforms. Policymakers, meanwhile, can draw confidence from the apparent effectiveness of regulatory interventions like the CRFA. By safeguarding consumers' rights to honest feedback, these regulations ultimately promote greater transparency, trust, and informed decision-making in the online marketplace.

Theoretical Predictions

In this section, we develop our theoretical predictions concerning the effects of policies designed to mitigate opinion suppression in online review platforms. We seek to conceptualize two outcomes associated with a reduction of suppression that are commonly associated with authenticity and informational quality: negative evaluation and review length.

Online reviews play a central role in consumer decision-making (Aral 2014). Reviews enable consumers

to share their opinions of products or services with other potential buyers (Dellarocas 2003, Lee and Bradlow 2011, Lu et al. 2013), providing a vital source of information, particularly in product categories that are overlooked by professional critics (Reimers and Waldfogel 2021). As a result, online reviews have a significant impact on sales (Chevalier and Mayzlin 2006). Recognizing this, many businesses will go to great lengths to manage and prevent negative reviews, with some going so far as to purchase fake positive reviews for themselves (Mayzlin et al. 2014, He et al. 2022) or fake negative reviews for competitors (Luca and Zervas 2016). Although several prior works have examined the issue of fake reviews, no empirical work has systematically considered businesses' use of an alternative strategy for rating manipulation, namely the suppression of honest negative feedback, for example, via legal threat.

Many businesses attempt to prevent or remove truthful negative reviews, instituting "gag clauses" in contracts (Ponte 2016) or threatening lawsuits. The problem has been sufficiently prevalent that in late 2016, President Barack Obama signed legislation directly intended to curb the behavior. This legislation, the CRFA, was otherwise known as the "Right to Yelp!" bill. The CRFA prohibits businesses' attempts to censor negative online reviews via contractual gag clauses (Calvert 2018) and has instituted fines and penalties for businesses that continue to engage in such behavior (Goldman 2017, Calvert 2018). Because the law explicitly bars businesses' attempts to silence negative feedback using the legal system, we expect that its passage will have led to an increase in the prevalence of negative ratings and text sentiment in business reviews. This leads us to our first prediction, namely, that *following the enactment of the CRFA, reviews of businesses in the United States will become more negative relative to businesses located in other countries.*

By removing legal barriers and empowering consumers to share honest feedback without fear of suppression, the Consumer Review Fairness Act (CRFA) creates an environment where reviewers can express themselves more freely. Before the CRFA, businesses could use non-disparagement clauses and legal threats to deter customers from expressing negative information in reviews, whether through low ratings or negative text, leading to self-censorship. The fear of legal consequences can thus lead not only to a reduction in the number of negative reviews businesses receive but also to a reduction in the volume of information expressed within any given review, as consumers may omit certain pieces of (negative) information. If consumers stick only to provable facts, for example, to minimize risk, reviews will become relatively shorter. Once the CRFA is in place and such chilling effects are lifted, consumers can thus be expected to provide more comprehensive accounts of their experiences with businesses. Freed from the concern of retaliation, they may feel more comfortable articulating negative aspects, and in greater detail. As a result, reviews are likely to increase in average length. This leads us to our second prediction, namely, that *following the enactment of the CRFA, reviews of businesses in the United States will become longer, on average, relative to businesses located in other countries.*

Methods

Data Set and Measures

Our primary data set is from TripAdvisor, one of the most popular online review platforms. TripAdvisor provides travel-related reviews for hotels, restaurants, airlines, and experiences. Statistics from 2019 indicate that the platform hosts more than 1 billion reviews and opinions of nearly 8 million businesses.[5] In June of 2021, we collected all English-language hotel reviews posted to TripAdvisor, prior to 2020, that were associated with 170 major worldwide hotel chains.[6]

For each hotel, we collected the hotel and chain name, hotel star level (e.g., 3-star hotel), address, and textual description. For each review associated with each hotel, we collected the review's URL, rating valence (an integer between 1 and 5), review text, the author's profile name, the date the review was posted, and the reported date of the hotel stay. Using the review text, we constructed a measure of textual sentiment employing the VADER (Valence Aware Dictionary and Sentiment Reasoner) library, a lexicon- and rule-based sentiment analysis tool specifically tuned to sentiment expressed in social media. Because VADER is an English-language package, we focus our analyses on English-language reviews.

To facilitate our analysis, we also created a pair of binary variables reflecting whether a review was posted after the CRFA's passage, *PostCRFA*, assigning a value of 1 to all reviews published during or after 2017, and whether the reviewed hotel is located in the United States, *TreatUS*.

Finally, we define several pretreatment variables for consideration as eventual moderators. These include an indicator of whether the average historical rating for the hotel was greater than or equal to 4 out of 5 stars (*HotelReputable*), the number of unique hotels within a focal hotel's zip code (*HotelCompetition*), a binary indicator of whether a reviewer resides in the United States (*AmericanReviewer*), and days elapsed since a reviewer's first review on TripAdvisor (*ReviewerTenure*).

Using this sample, we next implemented a matching procedure employing coarsened exact matching (CEM), matching affected hotels in the United States to comparable hotels in other countries. We match on

several pretreatment hotel-level characteristics, including the hotel's prior average rating valence, the year a hotel was established, the proportion of reviews written prior by individuals residing in English-speaking countries (i.e., the average of the country language variable across prior reviews), whether the hotel chain's headquarters is located in the United States, and the total number of properties that the parent hotel company owns. We plot covariate balance in terms of standardized mean differences (Cohen's *d*) in Figure 1, demonstrating the efficacy of our matching procedure. We can see that all covariates exhibit a standardized mean difference of less than 0.2, a typical threshold for a "small" effect size (Cohen 2013). We also report *t*-tests evaluating mean differences in each covariate, obtained via a linear regression of the treatment indicator onto our matching covariates (see Table 1). Our estimation sample comprises just over 2 million review-level observations. Table 2 presents the descriptive statistics for the variables used in our analyses.

Research Design

We employ a difference-in-differences design (Angrist and Pischke 2008). To evaluate our theoretical predictions regarding the impact of the CRFA on reviews of U.S. hotels, we employ the specification reflected by Equation 1, where h indexes hotels, t indexes years, and j indexes reviewers. $TreatUS_h$ is an indicator of whether the hotel being reviewed is affected by the CRFA, and $PostCFRA_t$ is an indicator of whether a review was written following the enactment of the CRFA policy, that is, whether the observation takes place in 2017 or later. We incorporate fixed effects for the hotel (φ_h), year (τ_t), and reviewer (δ_j). Finally, ε_{htj} represents our idiosyncratic error term. Here, Y reflects review-level outcomes, including valence, sentiment, and length. This specification is an extension of the standard two-way fixed effect estimator, given that we have many observations per hotel, reviewer, and period.

$$Y_{htj} = \beta_1 * (TreatUS_h \times PostCRFA_t) + \varphi_h + \tau_t + \delta_j + \varepsilon_{htj} \quad (1)$$

After estimating the baseline specification, we consider an event study specification, implementing a dynamic difference-in-differences (DID) regression to assess the temporal aspects of the treatment's influence and to test the parallel trends assumption. We implement this by expanding the PostCRFA dummy into a vector of year-quarter dummies. We interact these quarter dummies with our *TreatUS* indicator, omitting the dummy associated with the first quarter of 2017, which serves as the reference period. Lastly, we explore heterogeneity, incorporating additional moderators to the treatment, namely measures of hotel popularity, zip-code level, hotel competition, and the review author's location (American vs. not).

Figure 1. (Color online) Covariate Balance Before and After Matching

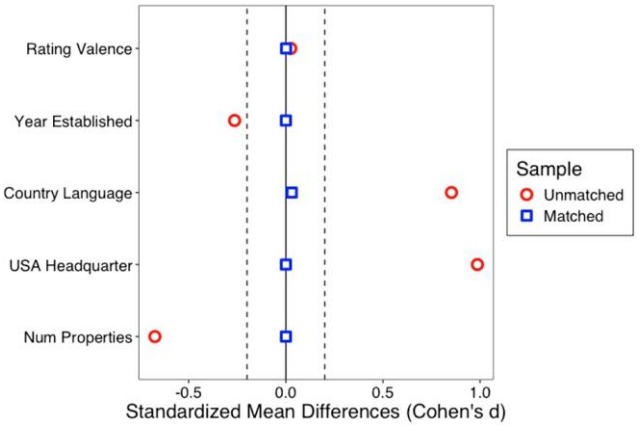


Table 1. Association Between Covariates and Treatment

Variables	Coefficient	SE	*t*-value	*p*-value
(Avg) rating valence	−0.0023	0.0115	−0.2027	0.8394
Year established	0.0000	0.0005	−0.0549	0.9562
Country language	0.0001	0.0001	1.2005	0.2300
U.S. headquarter	−0.0064	0.0266	−0.2410	0.8096
(Avg) number of properties	0.0000	0.0000	−0.0714	0.9431
Constant	−0.0023	0.0115	−0.2027	0.8394
Hotels		25,048		
RMSE		0.4957		

Results

Descriptive Analyses and Evidence of Review Suppression

Before conducting our analyses of the CRFA's impact on reviewing activity, we sought to understand the prevalence of review suppression. We considered the ~115,000 reviews posted between January and June 2021, the period leading up to our initial data collection. We revisited the URL for each online review six months later in December 2021 to check whether the review had been deleted. TripAdvisor does not allow business owners to delete the reviews they receive; only a review's author or TripAdvisor itself can delete a review. The vast majority of reviews that TripAdvisor deletes are removed because of violations of platform policies, for example, foul language, or reviews the platform believes to be fake. As a result, the vast majority of reviews included in the initial six-month snapshot would already have been filtered by the platform. In turn, any of these reviews that were removed over the subsequent six-month period would reflect a review author's choice to delete the content.

Table 2. Descriptive Statistics

Variables	Obs	Mean	SD	Min	Max
PostCRFA	2,209,786	0.373	0.484	0.000	1.000
TreatUS	2,209,786	0.768	0.422	0.000	1.000
Rating valence	2,209,786	3.960	1.081	1.000	5.000
Textual sentiment	2,209,786	0.714	0.472	−0.999	1.000
Review length	2,209,786	483.322	329.890	25.000	21,094.000
HotelReputable	2,209,786	0.514	0.500	0.000	1.000
HotelCompetition[a]	2,209,786	5.735	9.923	1.000	145.000
AmericanReviewer	2,209,786	0.701	0.458	0.000	1.000
ReviewerTenure	2,209,786	469.944	561.620	0.000	2,525.000
Year	2,209,786	2,015.870	1.763	2,013.000	2,019.000
Year established	2,209,786	1,958.785	25.779	1,742.000	2,013.000
Country language	2,209,786	0.914	0.280	0.000	1.000
U.S. headquarters	2,209,786	0.822	0.382	0.000	1.000
Number of Properties	2,209,786	7,042.952	2,424.412	24.000	20,000.000

[a]Defined based on a hotel's zip code or postal code.

Among these reviews, we observed that approximately 1% were eventually deleted over the subsequent six months. Moreover, consistent with deletion due to review suppression efforts on the part of businesses, we observed a clear, statistically significant relationship between the negativity of a review's valence and its probability of eventually being deleted (see Online Appendix A for additional details). Specifically, we estimate that a 1-star decrease in rating valence is associated with an approximate 26% (0.26 pp) increase in the probability that a review is deleted by its author within six months. Although suggestive, these observations indicate that, indeed, review suppression remains prevalent across TripAdvisor.

Beyond understanding the scale of the issue, we also explored the degree to which businesses' attempts at review suppression via legal threats might chill threatened consumers' subsequent reviewing activity around other businesses. We compiled data on the reviewing activity for two groups of consumers, namely (i) consumers who had posted to the TripAdvisor discussion forums stating that they had received a legal threat from a business about an online review they had posted and (ii) a matched set of active reviewers who had not appeared in the discussion forum posts or replies. Leveraging these two samples, we contrasted the reviewing behavior of threatened consumers against that of the matched sample before and after the threatened users' forum posts. We observed a statistically significant, systematic increase in the average valence and textual sentiment of reviews authored by threatened users (refer to Online Appendix B for more details of this analysis). These results indicate that legal threats from businesses induce users to withhold subsequent negative feedback.

These initial findings suggest that online review suppression remains prevalent and has material consequences on the health of online review platforms and the informedness of consumer decision-making, underscoring the need for intervention. We now focus on one such intervention, the passage of the CRFA by U.S. Congress, to better understand its efficacy and impacts.

Main Results

We report the results of our DID estimations in Table 3. In columns 1 and 2, we observe that the average rating valence and textual sentiment of reviews decline to a statistically significant degree following the passage of the CRFA. The average rating valence among affected (U.S.) hotels declined by approximately 0.05 stars ($p < 0.001$), whereas the average textual sentiment declined by an estimated 0.01 points ($p < 0.001$). Together, these results support our theoretical prediction on review negativity.

By removing legal barriers and empowering consumers to share honest feedback without fear of suppression, the CRFA creates an environment where reviewers can express themselves more freely. Before the CRFA, businesses could use non-disparagement

Table 3. Effect of CRFA on Review Characteristics

Variables	(1) Rating valence	(2) Textual sentiment	(3) Review length
PostCRFA × TreatUS	−0.0435***	−0.0118***	5.6262**
	(0.0049)	(0.0020)	(2.0595)
Hotel FE	Yes	Yes	Yes
Reviewer FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	2,209,786	2,209,786	2,209,786
Adj. R^2	0.5701	0.4543	0.5650

Note. Robust standard errors in parentheses, clustered by country (144 clusters); effective sample size is reduced because of the inclusion of reviewer fixed effects because a large fraction of reviewers contribute just one review within our sample.

**$p < 0.01$; ***$p < 0.001$.

clauses and legal threats to deter customers from leaving negative or detailed reviews, leading to self-censorship. The fear of legal consequences can lead not only to a reduction in the number of negative reviews of businesses but also to a reduction in the volume of information expressed within any given review because consumers omit specific details. If consumers stick only to provable facts, for example, to minimize risk, we might also expect reviews to be shorter. Once the CRFA is in place and such chilling effects are lifted, consumers can be expected to provide more comprehensive accounts of their experiences with the businesses. Freed from the concern of retaliation, they may feel encouraged to articulate both positive and negative aspects in greater detail. As a result, we conduct a secondary analysis testing for an increase in total review text. This is precisely what we see in column 3 of Table 3; we estimate an approximate five-character increase in review length, on average ($p < 0.01$).

Heterogeneous Effects

Having established average effects, we will now consider heterogeneous treatment effects. Specifically, within the same DID framework, we extend Equation 1 to incorporate moderators related to hotel characteristics (pretreatment popularity and competition) and review author characteristics (home country and account tenure).

First, we consider *HotelReputable*, a static feature that captures whether a hotel had an average rating of at least 4.0 before the CRFA's passage. The estimation results are reported in Table 4. We observe that the CRFA had a weaker impact on the average rating valence of reviews received by hotels that held a positive reputation before the policy change. Conversely, this result implies that treated hotels that lacked a positive reputation began to receive relatively more negative feedback after the CRFA was passed. This result is in line with the expectation that less reputable hotels are likely to be of lower quality and were thus more likely to have engaged in review suppression before the CRFA's passage to improve their ratings. Therefore, CRFA is likely to have a stronger impact on less reputable hotels.

Next, we consider the moderating influence of competition that a hotel faces. We measure hotel competition level as the number of unique hotels within a given zip or postal code. A higher number of unique

Table 4. Heterogeneous Effects of Hotel and Reviewers' Characteristics on Rating Valence Post CRFA

Models	(1) Hotel reputability	(2) Hotel competition	(3) Author home country	(4) Author account tenure
PostCRFA × TreatUS	−0.0562***	−0.0348***	−0.0315***	0.0443^{+}
	(0.0083)	(0.0044)	(0.0050)	(0.0236)
PostCRFA × HotelReputable	0.0150			
	(0.0103)			
PostCRFA × TreatUS × HotelReputable	0.0254*			
	(0.0102)			
PostCRFA × HotelCompetition		−0.0003		
		(0.0002)		
PostCRFA × TreatUS × HotelCompetition		−0.0015***		
		(0.0002)		
TreatUS × American			−0.0598***	
			(0.0053)	
PostCRFA × American			0.0198^{+}	
			(0.0106)	
PostCRFA × TreatUS × American			−0.0282**	
			(0.0104)	
TreatUS × UserTenure				0.0127***
				(0.0034)
PostCRFA × UserTenure				0.0129***
				(0.0037)
PostCRFA × TreatUS × UserTenure				−0.0113***
				(0.0032)
Hotel FE	Yes	Yes	Yes	Yes
Reviewer FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	2,209,786	2,209,786	2,172,153[a]	2,209,786
R^2	0.5702	0.5702	0.5698	0.5702

Notes. Robust standard errors in parentheses, clustered by country. Main effect of American is identified because a very small number of reviewers change their home location over time.

[a]Observation count declines slightly because we geocode the locations that reviewers state in their profile, employing the Google Maps API (some of these locations cannot be successfully mapped to a country). Furthermore, the main effect of UserTenure is identified despite the inclusion of reviewer and year fixed effects because UserTenure is measured in days since the first review.

$^{+}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

hotels within the same location indicates higher competition. The results in Table 4 suggest that hotels in more competitive markets were more affected by CRFA, particularly regarding their average star rating, likely because such hotels were more likely to engage in review suppression before the CRFA's passage. This is generally consistent with the prior literature on fake reviews and the idea that maintaining a positive reputation is critical to standing out among peers.

We now turn our attention to moderators related to reviewer characteristics (models 3 and 4 in Table 4). We observe that American consumers are relatively more likely to exhibit shifts in their reviews of American hotels following the policy change compared with foreign reviewers. Furthermoer, we find that the effects are relatively more pronounced among users of longer tenure on the platform. This aligns with our expectation that reviewers with lengthier platform tenure are likely to have experienced or observed legal threats from businesses and thus should be more aware of and likely to respond to the CRFA policy change.

Robustness Checks

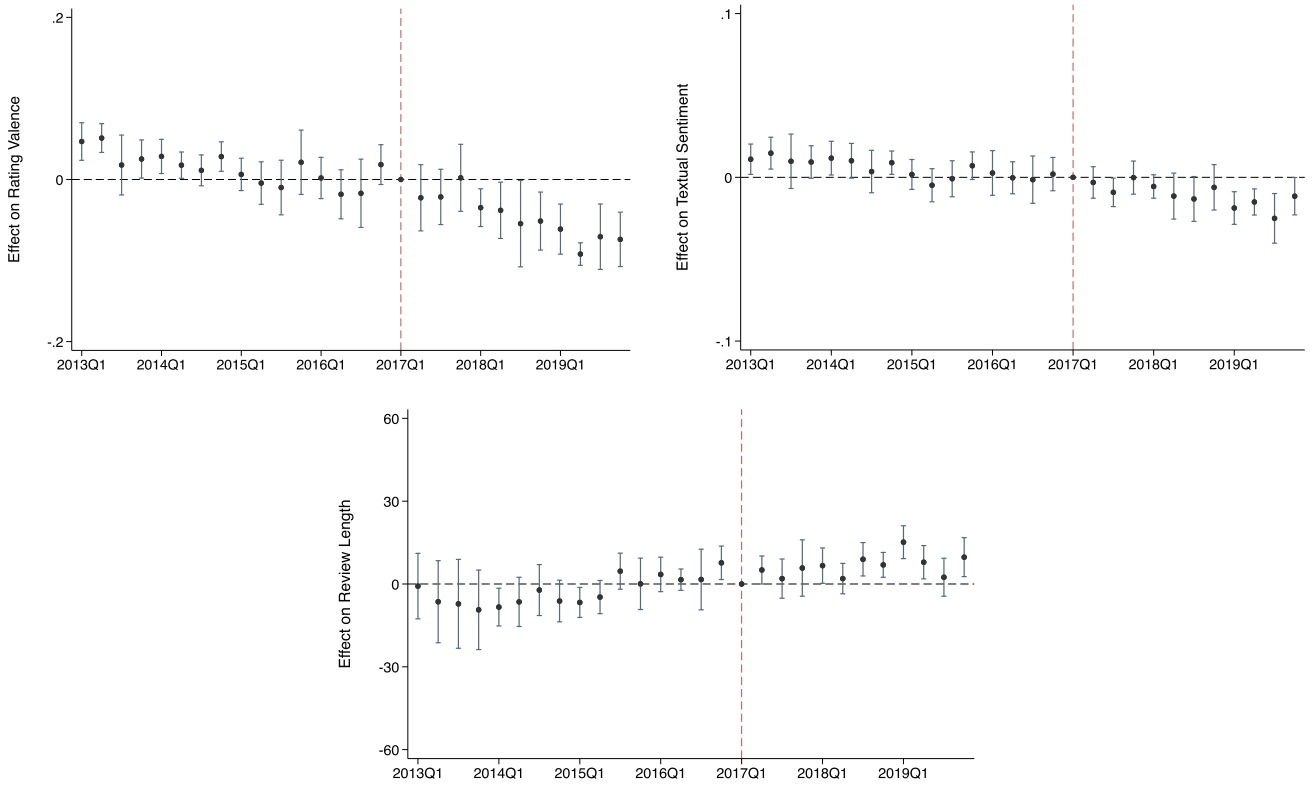
Event Study Specification

We next report the results of our dynamic difference-in-differences estimations, assessing the temporal dynamics of the treatment and evaluating the validity of the parallel trends assumption. We report our coefficients graphically, in Figure 2, along with 90% confidence intervals. We see statistically significant effects emerge shortly after the CRFA is signed into law, and the effects grow stronger over time. Furthermore, we observe what appears to be a general absence of effects in the periods leading up to the policy change, which suggests that the control group is a reasonably reliable counterfactual for the treated group. To further enhance the robustness of the causal interpretations, we also report next an alternative estimation approach using Matrix Completion counterfactual estimators, which further shows clear posttreatment effects and a general lack of pretreatment trends.

Matrix Completion

We next consider an alternative estimator, namely the Matrix Completion (MC) counterfactual estimator (Athey et al. 2021). We construct a hotel half-year panel that averages the characteristics of all reviews posted about a hotel in each six-month period. The MC estimator is particularly helpful when missing values are present in panel data. This is because the estimator employs nuclear norm matrix completion via singular value decomposition (SVD), a common technique for imputing unobserved user-item ratings in the recommender systems literature. The estimator works by first inferring a latent factor structure from the baseline panel,

Figure 2. (Color online) Dynamic Effects of CRFA on Review Characteristics



specifically a low-rank approximation of the original hotel-time panel matrix that yields the most accurate value relative to the true panel when the low-rank approximation is inverted. That learned factor structure is subsequently used to impute all unobserved values in the original (high-rank) panel, including values that were absent from the original panel because of missingness as well as values that were not observed because they reflect counterfactual outcomes for each unit under a potential outcomes framework. Finally, the difference between factual and counterfactual outcomes is used to infer a treatment effect.

This approach is useful here because in any hotel-time pair where no reviews were authored, our aggregated review characteristics will be missing in the baseline panel. The MC estimator imputes those missing values, enabling us to ensure the robustness of our conclusions to issues of missingness. We follow the best practices (Liu et al. 2024) and implement this estimation employing the "fect" package in R. Figure 3 presents graphical depictions of the event study estimates from each estimation, which are notably similar to our earlier two-way fixed effect-based event-study estimates. Specifically, we observe a general lack of pretreatment trends and increasing effects post-CRFA.

Generalizability

We also explore the generalizability of our findings in two respects. First, we replicate our findings using review data from a different review platform, Google Places. Second, we consider alternative policy shocks, namely state repeals of criminal defamation laws.

Google Places

We consider the generalizability of our findings to an alternative context (reviews from Google Places). We collected review data for a subset of our hotels (305) from Google Places. We repeat our baseline estimations using this sample, obtaining the results reported in Table 5, consistent with our earlier findings based on TripAdvisor data.

State Repeals of Criminal Defamation Laws

We also consider the generalizability of our findings to alternative policy changes. From 2002 to 2015, a series of states repealed their criminal defamation laws, beginning with Maryland in 2002, Arkansas in 2005, Washington in 2009, Colorado in 2012, Rhode Island in 2014, and Georgia in 2015. Based on this schedule, we create a binary variable, *Repeal*, equal to one for reviews posted to hotels located in a state where repeal has previously taken place and zero otherwise. We limit our sample only to reviews of hotels in the United States for this estimation to facilitate comparability and the plausibility of parallel trends. Our estimations again include a combination of hotel, reviewer, and time fixed effects. We observe the results reported in Table 6, which are once again consistent with our main findings related to the CRFA. Specifically, we again see that the repeal of a state's criminal defamation laws leads to a decline in average rating valence ($\beta = -0.0359$) and textual sentiment ($\beta = -0.012$) as well as an increase in average review length ($\beta = 6.423$). The effect magnitudes are also comparable to those found with the CRFA policy change. The findings show that consumers are significantly more inclined to provide negative feedback when they are assured protection from legal actions by businesses.

Discussion

Our study delves into a relatively underexplored area of online content dynamics: the impact of businesses' legal actions on suppressing negative opinions in the context of online reviews. Online reviews are fundamental to consumer decision-making processes; however, the censorship or suppression of these reviews can markedly influence market dynamics and obscure authentic consumer feedback. Based on observations from TripAdvisor.com, we show that the threat of

Figure 3. Matrix Completion: Effects of CRFA on Review Characteristics

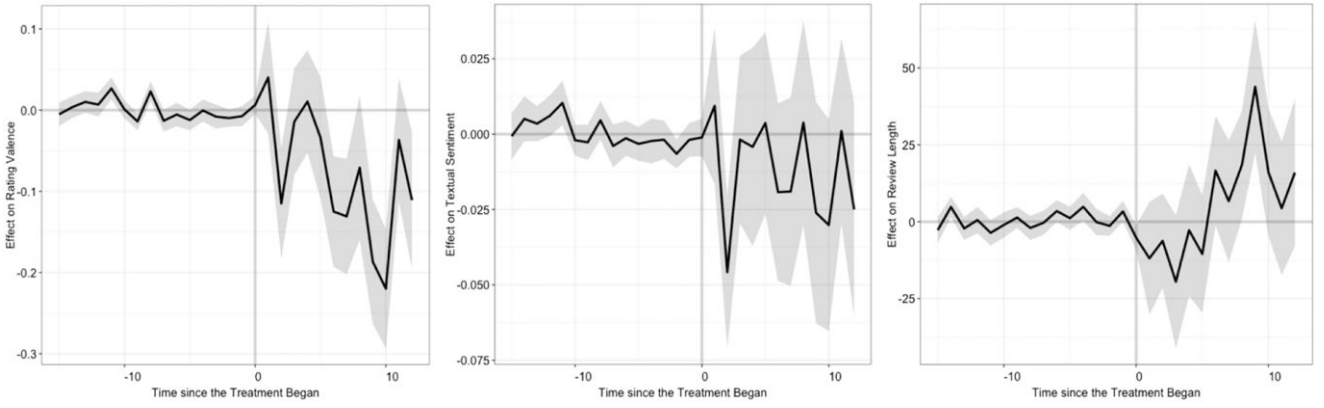


Table 5. Effect of CRFA on Review Characteristics (Google Places)

Variables	(1) Rating valence	(2) Textual sentiment	(3) Review length
PostCRFA × TreatUS	−0.2648***	−0.0813***	1.3230**
	(0.0254)	(0.0101)	(0.4358)
Hotel FE	Yes	Yes	Yes
Reviewer FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	38,584	38,584	38,584
R^2	0.2532	0.2353	0.2908

Note. Robust standard errors in parentheses, clustered by country.
**$p < 0.01$; ***$p < 0.001$.

legal repercussions from businesses toward negative reviews significantly discourages the sharing of lengthier, negative, and likely honest feedback that contains useful information. This fear of facing legal repercussions deters consumers from posting their negative experiences to avoid potential legal fallout. As a result, the landscape of online reviews has evolved into one marked by caution, with negative feedback often being downplayed, suppressed, or entirely excluded. This tendency fosters an imbalanced portrayal of consumer satisfaction and product quality because it systematically excludes critical feedback crucial for a comprehensive understanding of a business's products or services.

However, the enactment of the CRFA in the United States has had a material impact, protecting consumers' freedom of expression on online review platforms. Reducing threats of litigation from businesses has fostered a more genuine and expressive environment for consumer feedback. Notably, the results about CRFA on TripAdvisor reviews are further corroborated in an alternative review setting (Google Photos) and a different policy change (state repeals of criminal defamation laws). The policy impact not only aids consumers, who rely on reviews to make informed decisions, but also benefits businesses by providing deeper insights into consumers' experiences, thereby encouraging an open and constructive exchange of information between consumers and businesses.

Table 6. Effect of State Repeal of Criminal Defamation Law on Reviews

Variables	(1) Rating valence	(2) Textual sentiment	(3) Review length
Repeal	−0.0359***	−0.0120**	6.4230***
	(0.0080)	(0.0031)	(1.3342)
Hotel FE	Yes	Yes	Yes
Reviewer FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	1,657,227	1,657,227	1,657,227
R^2	0.5714	0.4556	0.5623

Note. Robust standard errors in parentheses, clustered by state.
**$p < 0.01$; ***$p < 0.001$.

Our research makes several notable contributions to the literature. First, we contribute to the literature on consumer evaluation and online reviews (Ho et al. 2017, Feng et al. 2019), online word-of-mouth (Huang et al. 2019, Wang et al. 2020, Sun et al. 2025), and more broadly that on user-generated content (Liu et al. 2017, Pu et al. 2020, Greenstein et al. 2021, Liu and Feng 2021) by presenting an initial examination of review suppression and policy responses aimed at mitigating it based on large-scale data from a leading online review platform. Furthermore, we present descriptive analyses in our appendices that speak to the prevalence of review suppression and its chilling effects on consumers. As the body of literature on online reviews and user-generated content continues to expand and evolve, it is imperative that researchers begin to explore the legal dimensions of this phenomenon.

Second, we provide empirical evidence about the effect that policy responses have had on review suppression. Our findings demonstrate that policy efforts can have meaningfully beneficial effects, particularly among reviewing consumers who reside in the policy jurisdiction and those who have lengthier experience on reviewing platforms. Our findings thus bear valuable insights and implications for policymakers and review platforms who might seek to lobby for policy intervention. More generally, these results add to prior research that has demonstrated the significant influence of geography and the variation in social context and culture that comes with it on product evaluations (Danescu-Niculescu-Mizil et al. 2009, Hong et al. 2016, Nakayama and Wan 2019). Our work highlights the additional importance of local legal and regulatory frameworks as crucial factors affecting online review generation and suppression.

Third, and last, we contribute to the existing work on consumer trust and trust elicitation in digital platforms (Wu et al. 2015, Duffy 2017, Stevens et al. 2018). By highlighting the prevalence and adverse consequences of review suppression, we identify the need for further research on this topic. Ensuring transparency and authenticity in consumer reviews is necessary to ensure trust and confidence in reviewing platforms and to ensure consumer protection. Beyond the primary insights we offer here, it is worth noting that we have also observed an association between the prevalence of review suppression and a lack of freedom of expression in a country (see Online Appendix C for more details). Together, these results emphasize the global significance of the problem and the role that government and policymakers can play in ensuring transparency in business practices and service quality to support consumer decision-making.

Limitations and Future Work

Of course, this study is subject to several limitations. First, our data are primarily sourced from TripAdvisor.com, which is unique in several respects that may limit generalizability. For example, TripAdvisor allows any individual to post reviews without verification of consumption (i.e., staying at the hotel), a policy that differs from other platforms that host hotel ratings (e.g., Expedia.com). Although we supplemented our results with review data from Google Places, future work might thus benefit from examining patterns of review suppression on other review platforms, such as Yelp.

Second, our study relies exclusively on archival, observational data, which presents difficulties in exploring the nuances of suppression as an underlying mechanism and its specific mode of occurrence. For example, although the focus of this study is review suppression deriving from a legal threat, suppression arising from a retailer-provided incentive is also possible, such as a post-consumption discount or payment (i.e., a bribe) as a means of getting users to remove negative reviews. Our initial exploration of user forum posts on TripAdvisor, Yelp, and Google Places suggests that both mechanisms are likely to be at play because consumers actively post questions indicating they are experiencing both interventions. Accordingly, more work is needed to understand the prevalence of each phenomenon because the proper policy responses to addressing each will likely differ. Because some people might fail to report legal threats, and most will not report receipt of incentives, addressing this issue adds ample complexity. Future work might thus draw on alternative methodologies and data, for example, surveying consumers to ask more detailed questions about experiences with suppression. Third, and last, our study is presently limited to the immediate act of suppression and deletion on TripAdvisor without consideration of potential spillover effects on suppressed consumers' subsequent reviewing activity on other platforms. Going forward, it would be interesting to examine individual-level reviewing behavior over time to evaluate whether receiving threatening emails or phone calls from a business owner may depress negative feedback that consumers provide elsewhere.

Conclusion

We report on a series of studies that address an underresearched issue within the literature on online reviews, namely the suppression of negative reviews and policy responses. We provide what is, to our knowledge, a first estimate of the prevalence of review suppression via legal threats in the context of online hotel reviews. We document systematic evidence that review suppression is more likely to occur as reviews grow more negative. We also show an analysis that demonstrates that businesses' attempts at suppressing critical reviews do not have a contained effect on the negative consumer feedback they hope to eliminate; rather, threatened consumers demonstrate significant shifts in their reviewing activity after experiencing legal threats, producing systematically more positive reviews afterward, suggesting a chilling effect. Most notably, we demonstrate the role that government policies and legislatures may have on this issue, showing the impact of legal frameworks like the CRFA and the repeal of defamation law on mitigating review suppression. Policies and legislation that curb review suppression improve the informational quality of reviews. With fewer constraints, reviewers are more likely to provide candid and comprehensive feedback, including honest negative evaluations and offering lengthier, more detailed accounts. We hope that this work can bring attention to understudied issues within the lengthy literature dealing with online reviews, particularly those related to the policy impacts.

Endnotes

[1] https://www.news-leader.com/story/news/local/ozarks/2018/06/04/man-branson-sued-fun-park-tripadvisor-review/655714002/.

[2] https://nytimes.com/2020/11/11/world/asia/thailand-hotel-tripadvisor-jail.html.

[3] https://www.ftc.gov/news-events/news/press-releases/2024/08/federal-trade-commission-announces-final-rule-banning-fake-reviews-testimonials.

[4] https://www.congress.gov/bill/114th-congress/house-bill/5111; the Consumer Review Fairness Act became Public Law No: 114-258, on December 14, 2016.

[5] See https://tripadvisor.mediaroom.com/US-about-us.

[6] Non-English reviews represent just 0.06% of all reviews in our initial sample; we omit reviews posted in 2020 or later to avoid the influence of COVID-19, which affected travel heterogeneously across countries and began in early 2020.

References

Angrist JD, Pischke JS (2008) *Mostly Harmless Econometrics: An Empiricist's Companion* (Princeton University Press, Princeton, NJ).

Aral S (2014) The problem with online ratings. *MIT Sloan Manag. Rev.* 55(2):47–52.

Athey S, Bayati M, Doudchenko N, Imbens G, Khosravi K (2021) Matrix completion methods for causal panel data models. *J. Am. Stat. Assoc.* 116(536):1716–1730.

Calvert C (2018) Gag clauses and the right to gripe: The Consumer Fairness Act of 2016 & state efforts to protect online reviews from contractual censorship. *Widener Law Rev.* 24(1):203–234.

Chevalier JA, Mayzlin D (2006) The effect of word of mouth on sales: Online book reviews. *Nat. Bureau Econom. Res.* 43(3):345–354.

Cohen J (2013) *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. (Routledge, New York).

Danescu-Niculescu-Mizil C, Kossinets G, Kleinberg J, Lee L (2009) How opinions are received by online communities: A case study on Amazon.Com helpfulness votes. *WWW – Proc. 18th Internat. World Wide Web Conf.* (Association for Computing Machinery, New York), 141–150.

Dellarocas C (2003) The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management Sci.* 49(10):1407–1424.

Duffy A (2017) Trusting me, trusting you: Evaluating three forms of trust on an information-rich consumer review website. *J. Consumer Behav.* 16(3):212–220.

Fang L (2022) The effects of online review platforms on restaurant revenue, consumer learning, and welfare. *Management Sci.* 68(11): 8116–8143.

Feng J, Li X, Zhang XM (2019) Online product reviews-triggered dynamic pricing: Theory and evidence. *Inform. Systems Res.* 30(4):1107–1123.

Greenstein S, Gu G, Zhu F (2021) Ideology and composition among an online crowd: Evidence from Wikipedians. *Manage. Sci.* 67(5):3067–3086.

Goldman E (2017) Understanding the consumer review fairness act of 2016. *Mich. Telecomm. Tech. Law Rev.* 24(1):1–15.

He S, Hollenbeck B, Proserpio D (2022) The market for fake reviews. *Marketing Sci.* 41(5):896–921.

Ho YC, Wu J, Tan Y (2017) Disconfirmation effect on online rating behavior: A structural model. *Inform. Systems Res.* 28(3):626–642.

Hong Y, Huang N, Burtch G, Li C (2016) Culture, conformity, and emotional suppression in online reviews. *J. Assoc. Inform. Systems* 17(11):737–758.

Huang N, Sun T, Chen P, Golden JM (2019) Word-of-mouth system implementation and customer conversion: A randomized field experiment. *Inform. Systems Res.* 30(3):805–818.

Lee TY, Bradlow ET (2011) Automated marketing research using online customer reviews. *J. Marketing Res.* 48(5):881–894.

Liu Y, Feng J (2021) does money talk? The impact of monetary incentives on user-generated content contributions. *Inform. Systems Res.* 32(2):394–409.

Liu Y, Feng J, Liao X (2017) When online reviews meet sales volume information: Is more or accurate information always better? *Inform. Systems Res.* 28(4):723–743.

Liu L, Wang Y, Xu Y (2024) A practical guide to counterfactual estimators for causal inference with time-series cross-sectional data. *Amer. J. Political Sci.* 68(1):160–176.

Lu X, Ba S, Huang L, Feng Y (2013) Promotional marketing or word-of-mouth? Evidence from online restaurant reviews. *Inform. Systems Res.* 24(3):596–612.

Luca M (2016) Reviews, reputation, and revenue: The case of Yelp.-Com, Harvard Business School NOM Unit Working Paper No. 12-016, Boston.

Luca M, Zervas G (2016) Fake it till you make it: Reputation, competition, and Yelp review fraud. *Management Sci.* 62(12):3412–3427.

Mayzlin D, Dover Y, Chevalier J (2014) Promotional reviews: An empirical investigation of online review manipulation. *Amer. Econom. Rev.* 104(8):2421–2455.

Moutos CP, Verma K, Phelps JY (2020) Negative patient reviews and online defamation: A guide for the obstetrician-gynecologist. *Obstet. Gynecol.* 136(6):1221–1226.

Nakayama M, Wan Y (2019) The cultural impact on social commerce: A sentiment analysis on Yelp ethnic restaurant reviews. *Inform. Management* 56(2):271–279.

Ponte LM (2016) Protecting brand image or gaming the system? Consumer 'gag' contracts in an age of crowdsourced ratings and reviews. *Willlim Mary Bus. Law Rev.* 7(1):59–149.

Pu J, Chen Y, Qiu L, Cheng HK (2020) Does identity disclosure help or hurt user content generation? Social presence, inhibition, and displacement effects. *Inform. Systems Res.* 31(2):297–322.

Reimers I, Waldfogel J (2021) Digitization and pre-purchase information: The causal and welfare impacts of reviews and crowd ratings. *Amer. Econom. Rev.* 111(6):944–1971.

Sen S, Lerman D (2007) Why are you telling me this? An examination into negative consumer reviews on the Web. *J. Interactive Marketing* 21(4):76–94.

Stevens JL, Spaid BI, Breazeale M, Esmark Jones CL (2018) Timeliness, transparency, and trust: A framework for managing online customer complaints. *Bus. Horiz.* 61(3):375–384.

Sun T, Wei YM, Golden J (2025) Geographical pattern of online word of mouth: How offline environment influences online sharing. *Inform. Systems Res.*, ePub ahead of print August 19, https://doi.org/10.1287/isre.2019.9532.

Wang L, Gunasti K, Shankar R, Pancras J, Gopal R (2020) Impact of gamification on perceptions of word-of-mouth contributors and actions of word-of-mouth consumers. *MIS Quart.* 44(4):1987–2011.

Wu C, Che H, Chan TY, Lu X (2015) The economic value of online reviews. *Marketing Sci.* 34(5):739–754.

Yin D, Bond SD, Zhang H (2014) Anxious or angry? Effects of discrete emotions on the perceived helpfulness of online reviews. *MIS Quart.* 38(2):539–560.